Gravothermal collapse of self-interacting dark matter halos as the origin of intermediate mass black holes in Milky Way satellites

Tamar Meshveliani[*], Jesús Zavala, and Mark R. Lovell

*Centre for Astrophysics and Cosmology, Science Institute, University of Iceland,
Dunhagi 5, 107 Reykjavik, Iceland*

Milky Way (MW) satellites exhibit a diverse range of internal kinematics, reflecting in turn a diverse set of subhalo density profiles. These profiles include large cores and dense cusps, which any successful dark matter model must explain simultaneously. A plausible driver of such diversity is self-interactions between dark matter particles (SIDM) if the cross section passes the threshold for the gravothermal collapse phase at the characteristic velocities of the MW satellites. In this case, some of the satellites are expected to be hosted by subhalos that are still in the classical SIDM core phase, while those in the collapse phase would have cuspy inner profiles, with a SIDM-driven intermediate mass black hole (IMBH) in the center as a consequence of the runaway collapse. We develop an analytical framework that takes into account the cosmological assembly of halos and is calibrated to previous simulations; we then predict the timescales and mass scales (M_{BH}) for the formation of IMBHs in velocity-dependent SIDM (vdSIDM) models as a function of the present-day halo mass, M_0. Finally, we estimate the region in the parameter space of the effective cross section and M_0 for a subclass of vdSIDM models that result in a diverse MW satellite population as well as their corresponding fraction of SIDM-collapsed halos and those halos' inferred IMBH masses. We predict the latter to be in the range 0.1–$1000 M_\odot$ with a $M_{\mathrm{BH}} - M_0$ relation that has a similar slope, but lower normalization, than the extrapolated empirical relation of supermassive black holes found in massive galaxies.

I. INTRODUCTION

The cold dark matter (CDM) model is highly successful at explaining observations of the large-scale structure of the Universe (e.g., Ref. [1]). However, it has challenges in matching observations on small scales, such as in the regime of dwarf galaxies (for a recent review, see, e.g., Ref. [2]). Observationally, these challenges have been established prominently for dwarf galaxies in the Local Group and particularly within the Milky Way (MW) satellites. For instance, the dynamical mass is dominated by dark matter (DM) in the inner region of several bright MW satellites yet is low compared to the inner densities of the plausible subhalo hosts of MW analogs found in collisionless CDM simulations; this is the *classical* too-big-to-fail (TBTF) problem [3,4]. Another recurrent challenge is that several of the MW satellites are best explained by density profiles of constant density, known as "cored" density profiles, rather than the steep inner density slope found in CDM simulations, referred to as "cuspy" profiles; satellites that are reported to have cored profiles include Fornax, Sculptor, Crater II, and Antlia II [5–9]. Overall, it is now well established that the MW satellites have a diverse range of internal kinematics, which is likely associated with a subhalo population that exhibits a considerable diversity of inner density profiles, from large cores to dense cuspy systems [10–14]. This diversity is analogous to that of rotation curves observed in higher mass, gas-rich dwarf galaxies [15,16].

It is important to emphasize that such dwarf-scale challenges are only insurmountable within CDM if no other physical mechanisms related to gas/stellar (*baryonic*) physics are considered. There are in fact several baryonic processes that are known to exist that can alleviate these challenges. For instance, supernova feedback can inject energy into the inner DM halo, reducing its density [17]. If impulsive enough, this is an efficient and irreversible cusp-core transformation mechanism in dwarf galaxies [18–20]. In addition, tidal forces on the satellite by the MW DM halo and the MW disk can effectively lower the densities of MW subhalos if their orbits pass sufficiently close to the disk [21]. The diverse orbits of the MW satellites combined with this effect enhance the diversity of inner DM densities relative to the CDM-only expectations [22]. However, how efficient these processes are in creating the observed diversity of the MW satellite population remains uncertain since, for example, the impact of supernova feedback is expected to be small in very faint DM-dominated systems

[*]tam15@hi.is

with low stellar-mass ratios [23]. On the other hand, it has been argued that the tidal field of the MW system might not be strong enough to explain the extremely low densities of bright satellites such as Crater II and Antlia II [24,25].

One exciting possibility is that the properties of the MW satellites provide clues about the DM nature beyond CDM. In particular, if DM particles have strong self-interactions, they can impact the nonlinear evolution of halos, significantly reducing their inner densities [26]. Modified N-body simulations that incorporate self-interacting DM (SIDM) have shown that the collisions experienced by DM particles with each other lead to significant momentum exchange. This process effectively transfers heat from the dynamically hot outer regions of the halo to the colder central regions, thus lowering the central density of halos and creating constant (isothermal) density cores [27–34]. SIDM models can create sizeable DM cores and alleviate the classical TBTF problem if the transfer cross section per unit mass, σ_T/m_χ, is $\gtrsim 1$ cm^2 g^{-1} at the characteristic scales/velocities of MW satellites $\lesssim 50$ km/s [35]. It is also possible to alleviate significantly the diversity of rotation curves in higher mass dwarf galaxies (characteristic velocities >50 km/s) if $\sigma_T/m_\chi \gtrsim 2$–3 cm^2 g^{-1} [36,37].

Constant cross section SIDM models have been constrained more strongly at large scales/velocities. Particularly, σ_T/m_χ is required to be $\lesssim 0.1$–1 cm^2 g^{-1} at the scale of clusters based on gravitational lensing, x-ray morphology, and dynamical analysis in cluster mergers [38–43]. At scales corresponding to massive elliptical galaxies, previous constraints based on x-ray morphology have been shown to be weaker than anticipated by DM-only simulations [44] once baryonic effects are included. Current simulations including baryons have shown that $\sigma_T m_\chi \sim 1$ cm^2 g^{-1} is consistent with the morphologies of elliptical galaxies [45]. With such constraints at larger scales, a constant cross section SIDM model is already only narrowly viable as an alternative to CDM to explain the properties of dwarf galaxies. Recent developments regarding the diversity in the inner densities of the MW satellites virtually rule out this possibility since, with such a low cross section, it is not possible to generate very-high-density satellites such as Willman I [13,46–48].

Remarkably, what is needed for SIDM models to remain an interesting, viable alternative to CDM is to have even larger cross sections (> 10 cm^2 g^{-1}) at the scale of the MW satellites in order to trigger the gravothermal collapse phase (see below). Such large cross sections can be naturally accommodated by particle models with a velocity-dependent cross section (e.g., through Yukawa-like interactions; see e.g., Refs. [49–51]), where DM behaves as a collisional fluid on small scales and is essentially collisionless at cluster scales. Long after the core-formation phase, further DM particle collisions lead to heat outflow from the hotter inner region to the colder outskirts of the halo. Since gravitaionally bound systems have negative specific heat, mass/energy is continuously lost from the inner region, while the density and temperature continue to grow in a runaway instability that drives the collapse of the inner core. This phenomenon is known as the gravothermal catastrophe [52] and is observed in globular clusters, where the collapse is mainly halted by the formation of binary stars, which act as energy sinks [53]. For SIDM halos, the physical mechanism is the same, but without the formation of bound DM states to act as energy sinks, the collapse continues, eventually reaching a relativistic instability that results in the formation of a black hole [28,54–57]. If the core-collapse phase has been reached at the scales of the MW satellites, then the SIDM predictions become radically different with some of the satellites expected to be hosted by (sub)halos with SIDM cores, while those in the collapse phase would have cuspy (collapsed) inner DM regions [13].

Given the problems with constant cross section SIDM models mentioned above, it has been argued recently that such models could be reconciled with the MW satellite population by suggesting that the collapse phase might be accelerated in the host (sub)halos of MW satellites by mass loss via tidal stripping [58], since mass loss enhances the negative temperature gradient in the outskirts of the (sub)halo and makes the heat outflow more efficient. Accelerated core collapse has been invoked to explain the diversity of the MW's dwarf spheroidal galaxies in constant cross section models with $\sigma_T/m_\chi \gtrsim 2$–3 cm^2 g^{-1} [14,58–60]. However, Ref. [61] recently simulated SIDM subhalo satellites as they orbit the MW system and found that energy gain due to collisions between particles in the subhalo and the host instead inhibits core collapse in subhalos.

Another study, Ref. [62], showed that subhalos in models with constant cross sections between 1 and 5 cm^2 g^{-1} are not dense enough to match the densest ultrafaint and classical dwarf spheroidal galaxies in the MW, and 5 cm^2 g^{-1} is not sufficient to enforce collapse even with the tidal effect of a MW disk and bulge. This seemingly closes the last possibility for velocity-independent SIDM models (see also discussion in Sec. II G). On the other hand, this result motivates the exploration of velocity-dependent SIDM models, where recent full cosmological simulations with a specific benchmark model [13,63] have shown that cross sections $\gtrsim 50$ cm^2/g at velocities $\lesssim 30$ km/s naturally result in a diverse bimodal population of MW satellites, predicting both cuspy, high-velocity dispersion subhalos, consistent with dense systems (particularly ultrafaint satellites), and cored, low-velocity dispersion subhalos, consistent with brighter low-density satellites. These results have been confirmed and expanded to generic velocity-dependent SIDM models by the recent cosmological simulation suite TangoSIDM [64].

In this work, we adopt the benchmark SIDM model presented in Refs. [13,63] to explore the consequences of gravothermal collapse for the formation of intermediate mass

black holes (IMBHs) in the MW satellite population. Our goal is twofold: (i) to compile a simple analytical framework (calibrated to the simulations in Refs. [13,63]) that provides predictions for the formation timescales and mass scales of IMBHs in SIDM halos under arbitrary velocity-dependent cross sections, and (ii) to provide the range of IMBH masses that is expected given the plausible range of cross sections that produce a diverse MW satellite population, i.e., a bimodal—core-cusp—satellite distribution.

This paper is organised as follows. In Sec. II, we describe our model for the evolution of SIDM halos. We start with our adopted primordial halo density profile and the concentration-mass relation, describe our computation of the threshold time for the cusp-core transformation, and finally estimate the timescales and masses of IMBHs expected in the SIDM model due to gravothermal collapse. We also include the impact of tidal stripping. In Sec. III, we present our results, discuss how they are impacted by the various properties of the model, and put our work in the context of other related studies. Finally, we draw conclusions in Sec. IV.

II. GRAVOTHERMAL COLLAPSE IN SIDM HALOS

Our goal in this section is to follow the relevant stages in the evolution of an SIDM halo: (i) formation of the progenitor cuspy (i.e., CDM-like) halo, (ii) development of the central core, and (iii) gravothermal collapse of the core and formation of the black hole. In addition, we discuss how tidal stripping might affect the gravothermal collapse timescale.

A. Cosmic evolution of SIDM halos

In an SIDM halo where thermalization occurs due to close, rare interactions with large momentum transfer, a relaxation time can be defined due to self-scattering at the characteristic radius[1] r_{-2}, which is given by

$$t_{\rm r} = \frac{\lambda}{a\sigma_{\rm vel}}, \qquad (1)$$

where $\sigma_{\rm vel}$ is the characteristic velocity dispersion, $a = \sqrt{16/\pi}$ for hard-sphere scattering of particles with a Maxwell-Boltzmann velocity distribution [54] and $\lambda^{-1} = \rho(r_{-2})\sigma_{\rm T}/m_\chi$ is the mean free path, which is inversely proportional to the local density $\rho(r_{-2})$ and the cross section per unit mass $\sigma_{\rm T}/m_\chi$ (evaluated at the characteristic velocity $\sigma_{\rm vel}$ in the case of velocity-dependent SIDM models). Therefore, the scattering rate (mean free path)

is higher (shorter) in denser regions. Within the region where the age of the inner halo is comparable to the relaxation time, self-scattering has a significant impact on the inner DM structure turning the cusp into a core.

In CDM, where DM is collisionless, the velocity dispersion peaks near the scale radius, r_{-2}. By contrast, in SIDMs, elastic scattering leads to momentum exchange between DM particles, which, given the positive gradient of the velocity dispersion profile within r_{-2}, effectively results in heat transfer from the outside in, up to the radius where the velocity dispersion peaks. As a result, a central isothermal core is formed, which continues to grow until it is roughly the size of the scale radius and thus reaches a quasiequilibrium state. After core formation, subsequent collisions lead to momentum/energy flow from the center to the outskirts of the halo, where the velocity dispersion profile has a negative slope. Heat loss in the core results in the infall of DM particles to more tightly bound orbits, where they experience more interactions and are heated further due to the negative heat capacity of the self-gravitating system; a similar phenomenon occurs in globular clusters, [65]. Without energy sinks, the core suffers a runaway instability, transforming the core into an ever denser cusp, which ultimately results in the formation of a black hole [54].

An SIDM halo undergoes gravothermal collapse in a timescale $t_{\rm coll} \approx 382 t_{\rm r}$, as described in Sec. II F. The relaxation time depends on the halo mass and time of assembly/formation (described in Sec. II D).

B. Primordial density profile

We assume that in the SIDM cosmology DM assembles into spherical self-gravitating halos in virial equilibrium, with a *primordial* structure that is the same as that of CDM halos. This is a reasonable assumption at sufficiently high redshift when the average number of collisions in the center of halos is still well below one per Hubble time, and thus the structure of the halo has been affected only minimally. Cosmological simulations have shown that DM core sizes are only a small fraction of their value at $z = 0$ when the Universe is around 1 Gyr old ($z \sim 5$), e.g., Ref. [66].

The spherically averaged density profiles of equilibrium collisionless CDM halos are well approximated by a two-parameter formula known as the Navarro, Frenk, and White (NFW) profile [67,68],

$$\rho_{\rm NFW}(r) = \rho_{\rm crit}\frac{\delta_{\rm char}}{r/r_{-2}(1 + r/r_{-2})^2}, \qquad (2)$$

where r_{-2} is the radius at which the logarithmic slope of the profile is -2, $\rho_{\rm crit}$ is the critical density of the Universe, and the characteristic overdensity $\delta_{\rm char}$ is given by

$$\delta_{\rm char} = \frac{200}{3}\frac{c^3}{k(c)}, \qquad (3)$$

---

[1]From here on in, we assign the characteristic radius to the scale radius of the halo, which for the NFW profile is equal to r_{-2}, the radius at which the logarithm slope of the profile is -2; see Sec. II B.

where $k(c) = \ln(1 + c) - c/(1 + c)$ and the concentration c is defined as $c = r_{200}/r_{-2}$ with r_{200} being the virial radius, which is defined in this work as the radius where the mean density of the halo is 200 times ρ_{crit}.

C. Concentration-mass relation and formation redshift

The NFW profile is to first order a one free parameter profile since the virial mass of the halo and its concentration are strongly correlated, with a 1σ scatter in log c of order 0.1 [68,69]. We use the concentration-mass relation modeled in Refs. [70,71], where the authors link the enclosed mass profile of a halo at a given time with the prior mass aggregation history of the halo. In particular, following Ref. [71], we can define an *assembly/formation* redshift of a halo of mass M_0 at a redshift z_0 as the redshift z_{-2} when the enclosed mass within r_{-2} at z_0, M_{-2}, was first assembled into progenitors more massive than a certain fraction f of M_0. M_0 is defined as the mass within the virial radius $M_0 = (4\pi/3)r_{200}^3 200\rho_{crit}$. The virial mass of the halo at z_{-2} is equal to M_{-2} and can be computed from the assembly history:

$$M_{-2} = M_0 \times \mathrm{erfc}\left(\frac{\delta_c(z_{-2}) - \delta_c(z_0)}{\sqrt{2(\sigma^2(f \times M) - \sigma^2(M))}}\right). \quad (4)$$

The expression in parentheses on the right-hand side corresponds to the collapsed mass fraction in Extended Press-Schechter theory [72], where $\delta_c(z_{-2}) = \delta_c/D(z)$ is the redshift-dependent critical density for collapse with the linear growth factor $D(z)$ and $\sigma(M)$ is the rms mass variance. For the NFW profile, the mass is connected to the concentration by

$$\frac{M_{-2}}{M_0} = \frac{k(1)}{k(c)}, \quad (5)$$

$$\frac{\langle \rho_{-2} \rangle}{\rho_{crit}(z_{-2})} = 200c^3 \frac{k(1)}{k(c)}. \quad (6)$$

The key assumption in the model is that the mean density inside r_{-2} is directly proportional to the critical density of the Universe at an assembly redshift z_{-2},

$$\frac{\langle \rho_{-2} \rangle}{\rho_{crit}(z_{-2})} = C\left(\frac{H(z_{-2})}{H(z_0)}\right)^2, \quad (7)$$

where C is a free parameter. Throughout this paper, we use $f = 0.02$ and $C = 575$ [73]. Inserting Eqs. (5) and (6) into Eq. (4), we have a transcendental equation for the formation redshift $z_{form} = z_{-2}$ as a function of M_0, which can then be used to obtain the concentration c.

D. Threshold time for the cusp-core transformation

As a benchmark case, we set the halo formation time z_{-2} of an SIDM halo extant at the present day to be the threshold epoch at which the cusp-core transformation begins, $z_{cc} = z_{form} = z_{-2}$. At this epoch, we assume that the SIDM halo has a NFW profile with a virial mass equal to the enclosed mass within r_{-2} at $z_0 = 0$, $M(z_{-2}) = M_{-2}|_{z_0}$. The concentration of this primordial SIDM halo is calculated by repeating the method described in Sec. II C, but this time setting $z_0 = z_{-2}$. The range of z_{-2} values for the range of present-day halo masses that we are interested in, $10^8 \le M_0 \le 10^{12}M_\odot$, is given by $6.1 \ge z_{-2} \ge 3$. As we noticed earlier, given this relatively high redshift range, our choice of setting $z_{cc} = z_{-2}$ is reasonable because the effect of collisions in the inner halo is minimal at early times.

The next step is to develop a method to calculate the relevant timescale for gravothermal collapse (Sec. II F), for which we build a simplified model in which the evolutionary stages of the SIDM halo occur in isolation. This approach is somewhat different than the full cosmological setting, where halo mergers are an active mechanism of halo growth with transitory stages that affect the inner centre of the halo. Although the cuspy NFW profile of CDM halos is resilient to merger activity, e.g., Ref. [74], the situation might in principle be more complex in an SIDM scenario with gravothermal collapse for the following reason. In the standard SIDM model without core collapse, the merger between a small halo with a larger one is that of two shallow (corelike) profiles with the smaller one having progressed further in its core development since it forms earlier; the result of this merger is a DM profile that is also cored [75]. Thus, we would naively expect that halo mergers will not delay the cusp-core transformation. However, cosmological mass infall in general might delay the core-collapse phase by pumping energy into the central region to stabilize the core [76]. Moreover, in a velocity-dependent SIDM halo with a sharp difference between the cross section of low-mass halos to that of large mass halos, the former are expected to go through the cusp-core-collapse stages much faster than the latter, resulting in a scenario in which mergers between low-mass core-collapsed (cuspy) halos and high-mass cored halos are possible. This has the potential to delay the core-collapse phase.

Since our goal is to provide a simple, first-order estimate for the black hole formation time, rather than a comprehensive calculation, we assume that in a relatively extreme scenario, a significant merger would reset the clock for the cusp-core-collapse stage. For this event, we adopt the last major merger (LMM), which we define as a mass ratio of $10:1$ or higher between the two merging halos, and we label the corresponding redshift as z_{LMM}. In order to calculate z_{LMM}, we use the fitting formula for the mean merger rate $dN_m/d\xi/dz$—in units of mergers per halo per unit redshift per unit of mass ratio ξ—for a halo of mass

$M(z)$ at redshift z obtained from the combined Millennium and Millennium II datasets in Ref. [77]:

$$\frac{dN_{\rm m}}{d\xi dz}(M(z), \xi, z)$$
$$= A\left(\frac{M(z)}{10^{12}M_\odot}\right)^\alpha \xi^\beta \exp\left[\left(\frac{\xi}{\tilde{\xi}}\right)^\gamma\right] \times (1+z)^\eta, \quad (8)$$

where the best-fit parameters are $(\alpha, \beta, \gamma, \eta) = (0.133, -1.995, 0.263, 0.0993)$ and $(A, \tilde{\xi}) = (0.0104, 9.72 \times 10^{-3})$. The mass $M(z)$ is given by integrating the mean mass growth rate of halos, taken from Ref. [77]:

$$\frac{dM}{dt} = \langle \dot{M} \rangle_{\rm mean} = 46.1 M_\odot \ {\rm yr}^{-1} \left(\frac{\rm M}{10^{12} M_\odot}\right)^{1.1}$$
$$\times (1 + 1.11z)\sqrt{\Omega_{\rm m}(1+z)^3 + \Omega_\Lambda}, \quad (9)$$

where Ω_m and Ω_Λ are, respectively, the DM and dark energy density parameters evaluated at the present day.

The cumulative number of mergers $N_{\rm m}(\xi_{\rm min}, M_0, z_0, z)$ for a halo of mass M_0 at redshift z_0 is then given by

$$N_{\rm m}(\xi_{\rm min}, M_0, z_0, z) = \int_{z_0}^z dz \int_{\xi_{\rm min}}^1 d\xi \frac{dN_{\rm m}}{d\xi dz}[M(z), \xi, z], \quad (10)$$

where we use the minimum mass ratio for a major merger to be $\xi_{\rm min} = 0.1$. When the above equation equals 1, meaning that the halo experienced one major merger event, we find the corresponding $z_{\rm LMM}$ for given halo of mass M_0; we therefore only consider the properties of halos extant at $z_0 = 0$.

Having adopted all these considerations, we assume that a viable range for the threshold epoch of the cusp-core transformation is given by $z_{-2} < z_{\rm cc} < z_{\rm LMM}$. The corresponding cosmic time for this epoch is given by

$$t(z) = t_0 \int_0^{1/(1+z)} \frac{da}{\dot{a}} = \frac{2 \sinh^{-1}\left(\sqrt{\frac{\Omega_\Lambda}{\Omega_{\rm m}}}(1+z)^{-3/2}\right)}{3H_0\sqrt{\Omega_\Lambda}}. \quad (11)$$

where t_0 is the age of the Universe and H_0 is the Hubble parameter. For reference, Fig. 1 shows the range of plausible threshold times as a function of halo mass M_0.

E. Velocity-dependent SIDM cross section

The cross section per unit mass, σ/m_χ, is the key physical property that characterizes a specific SIDM model. We consider a class of models where self-scattering between DM particles are mediated by a massive force carrier of mass m_ϕ through an attractive Yukawa potential with coupling strength α_c [29,51,78]. Furthermore, we assume that the elastic scattering between SIDM particles
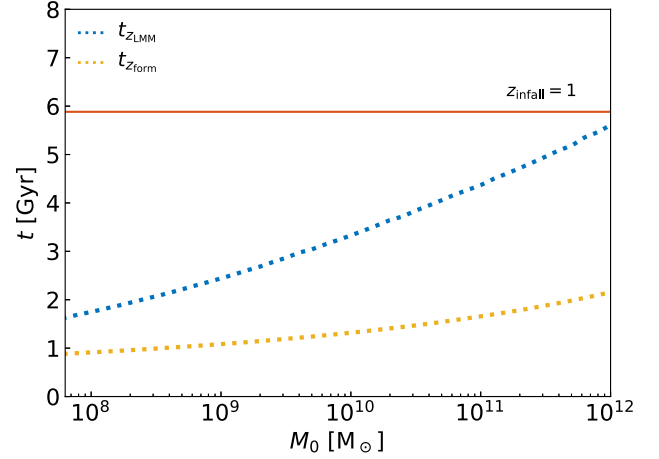

FIG. 1. Cosmic time for the cusp-core transformation threshold of SIDM halos as a function of the present-day halo mass M_0. The blue dotted line is the function of $z_{\rm LMM}$, and the yellow dotted line is the function of $z_{\rm form}$. The horizontal line indicates the *infall* time of subhaloes, which we fix to $z_{\rm infall} = 1$ (see Sec. II G).

can be modeled by the screened Coulomb scattering interaction observed in a plasma, which in the classical regime is well fitted by the transfer cross section,

$$\frac{\sigma_{\rm T}}{\sigma_{\rm T}^{\rm max}} \approx \begin{cases} \frac{4\pi}{22.7}\beta^2 \ln(1+\beta^{-1}), & \beta < 0.1 \\ \frac{8\pi}{22.7}\beta^2(1 + 1.5\beta^{1.65})^{-1}, & 0.1 < \beta < 10^3 \\ \frac{\pi}{22.7}(\ln \beta + 1 - 0.5\ln^{-1}\beta)^2, & \beta > 10^3, \end{cases}$$
$$(12)$$

where $\beta = \pi v_{\rm max}^2/v^2 = 2\alpha_c m_\phi/(m_\chi v^2)$ and $\sigma_{\rm T}^{\rm max} = 22.7/m_\phi^2$ and v is the relative velocity of the DM particles. Here, $v_{\rm max}$ is the velocity at which $(\sigma_{\rm T} v)$ peaks at a transfer cross section equal to $\sigma_{\rm T}^{\rm max}$.

In this work, we use the benchmark velocity-dependent SIDM model introduced in Ref. [13] with $v_{\rm max} = 25$ km/s and $\sigma_{\rm T}^{\rm max} = 60$ cm^2 g^{-1}. Reference [13] performed cosmological zoom simulations of an MW-size halo in which self-interactions were extremely frequent, and showed that the interaction frequency of DM particles in the subhalo centers was sufficient to trigger the gravothermal catastrophe phase in a fraction of the subhalo population. They subsequently demonstrated that frequent self-interactions constitute an alternative explanation for the diverse distribution of inner DM densities found in the MW satellite population (see Sec. I).

In our idealized model, we are interested in the characteristic scales of a given halo that are relevant to set a single characteristic cusp-core-collapse timescale. In particular, we assign a single relaxation timescale for a halo due to self-scattering using Eq. (1). We set a single characteristic velocity dispersion $\sigma_{\rm vel}$, which is given by the maximum of the velocity dispersion profile $\sigma_{\rm vel} = \sigma_{\rm r}(r_{\rm max})$ of the primordial NFW halo at the beginning ($z_{\rm cc}$) of the cusp-core

transformation. The radius at which this maximum occurs is of $\mathcal{O}(1)$ of the maximum size of the SIDM core that eventually develops, and the value of $\sigma_r(r_{\max})$ sets the *temperature* of the fully developed isothermal core. We now describe in detail how we calculate $\sigma_r(r_{\max})$.

We start by referring to the local radial velocity dispersion $\sigma_{\mathrm{r}}(r)$, which can in principle be obtained self-consistently by solving the Jeans equation,

$$\frac{1}{\rho}\frac{\mathrm{d}}{\mathrm{d}r}(\rho\sigma_{\mathrm{r}}^2) + 2\beta\frac{\sigma_{\mathrm{r}}^2}{r} = -\frac{\mathrm{d}\Phi}{\mathrm{d}r}, \qquad (13)$$

where $\beta = 1 - \sigma_\theta^2/\sigma_{\mathrm{r}}^2$ is the velocity anisotropy parameter and Φ is the gravitational potential, which for the NFW profile is given by

$$\frac{\Phi(s)}{V_{200}^2} = -\frac{1}{k(c)}\frac{\ln(1+cs)}{s}, \qquad (14)$$

where $s = r/r_{200}$ and V_{200} is the circular velocity at $r = r_{200}$:

$$V_{200}^2 = G\left(\mathrm{M_0}^2 \times \frac{4}{3}\pi 200\rho_{\mathrm{crit}}\right)^{1/3}. \qquad (15)$$

Here, we assume the simplest case of isotropic orbits, where $\sigma_\theta(r) = \sigma_{\mathrm{r}}(r)$ and $\beta = 0$. In this case, the solution to the Jeans equation can be computed analytically [79], giving the one-dimensional velocity dispersion,

$$\frac{\sigma_{\mathrm{r}}^2}{V_{200}^2}(s, \beta = 0) = \frac{1}{2k(c)}c^2 s(1+cs)^2\left[\pi^2 - \ln(cs) - \frac{1}{cs} - \frac{1}{(1+cs)^2} - \frac{6}{1+cs} + \left(1 + \frac{1}{c^2s^2} - \frac{4}{cs} - \frac{2}{1+cs}\right)\right.$$
$$\left. \times \ln(1+cs) + 3\ln^2(1+cs) + 6\mathrm{Li}_2(-cs)\right], \qquad (16)$$

where $\mathrm{Li}_2(x)$ is the dilogarithm. Using Eq. (16), we compute $\sigma_{\mathrm{vel}} = \sigma_r(r_{\max})$ for a given value of M_0 and c.

Finally, we compute a characteristic value for the transfer cross section $\langle\sigma_{\max}\rangle$ by computing the thermal average of the transfer cross section at $r_{\max}$, i.e., within the SIDM core. We assume that the velocity distribution of DM particles can be approximated by a Maxwell-Boltzmann distribution. Although such a distribution is not self-consistent with the NFW profile, e.g., Ref. [80], it is a reasonable approximation for our purposes because the distribution within the SIDM core will eventually become Maxwellian [81]. Therefore,

$$\langle\sigma_{\max}\rangle = \frac{1}{2\sigma_{\mathrm{vel}}^3\sqrt{\pi}}\int(\sigma_{\mathrm{T}})v^2 e^{-v^2/4\sigma_{\mathrm{vel}}^2}\mathrm{d}v. \qquad (17)$$

For reference, Fig. 2 shows the characteristic value for the transfer cross section per unit mass as a function of halo mass M_0 today for the benchmark SIDM model (vd100) from Ref. [13] that we use for calibration in our work.

F. Mass scales and timescales for black hole formation

We estimate the timescale for the formation of a black hole in the center of an SIDM halo following the procedure laid out by Ref. [82]. This timescale applies after the threshold time for the cusp-core transformation z_{cc} of the primordial NFW halo as discussed in Sec. II D. The formula from Ref. [82] is based on the (spherical) gravothermal fluid model, which has been used in the past to study the gravothermal catastrophe in SIDM halos, e.g., Refs. [54,56,57,83–86]. The recent study by Ref. [82] is

particularly suitable for our work since, contrary to previous papers, it considers velocity-dependent cross sections. In Appendix A, we make a comparison between that reference's estimate of the collapse time and that made in Ref. [57]. Before describing the key equation, we present a short summary of the physical process based on Ref. [54].

In the gravothermal evolution of an SIDM halo, we can distinguish between a long mean free path (LMFP) regime, where the typical distance a particle travels is much longer than the gravitational Jeans scale and the short mean free path (SMFP) regime, where the situation is reversed. In the LMFP regime, particles orbit the inner halo many times
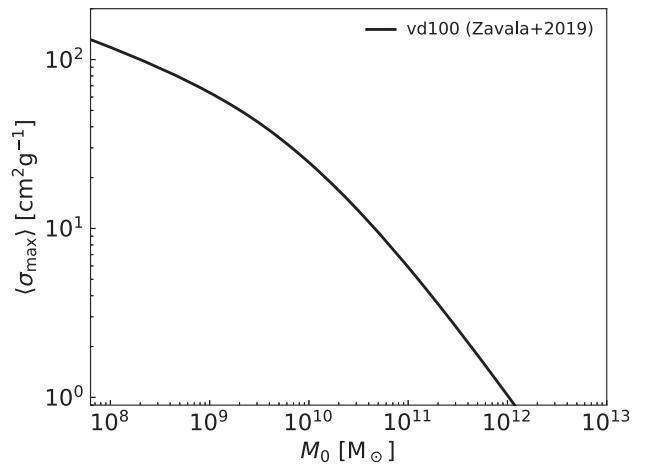

FIG. 2.   Characteristic transfer cross section per unit mass as a function of a halo mass at $z = 0$ [Eq. (17)] for the benchmark SIDM model (vd100) from Ref. [13].

unperturbed before being scattered, while in the SMFP regime, particle motions in the core are constrained by multiple collisions. In these two regimes, the heat conduction and the mass transfer between the core and the extended halo are different.

Initially, as the core is growing in size, both the core and extended halo are firmly in the LMFP limit; during this phase, the inner core is approaching the transitional regime, in between the LMFP and the SMFP where the Knudsen number is of $\mathcal{O}(1)$. While the extended halo remains nearly collisionless, with nearly the same primordial NFW density profile, the core evolves, and as the gravothermal collapse is triggered, it transitions into the SMFP regime where it essentially behaves as a fluid. Although the velocity dispersion and the density of the core both increase, the latter increases much faster with time, $d\log(\sigma_r^2)/d\log(\rho_{\rm core}) \approx 0.1$ [54], and it drives the core into two components: a dense SMFP inner core that continues to evolve, and a more dilute LMFP outer core, with a nearly constant density, which connects to the extended nearly unperturbed NFW halo. Eventually, the temperature in the inner core is so high that it enters the relativistic regime and dynamical instability occurs that leads to the formation of the black hole [54] (see also Ref. [87]). The classical gravothermal fluid formalism cannot be used once the system becomes relativistic. However, the classical formalism in the LMFP regime allows us to follow the evolution of the core to high central densities, and in fact, since most of the evolution occurs in the LMFP regime, with only the last instants prior to collapse being in the SMFP regime, the timescale for collapse is dominated by the LMFP evolution of the core, and thus the classical approach can be used to estimate the timescale of interest.

According to Ref. [82], the amount of time required for the primordial NFW to evolve from a central cusp to a SIDM core and then to a fully collapsed cored when a black hole forms—i.e., the cusp-core-collapse timescale—is given by

$$t_{\rm coll} \approx 382 t_{\rm r}(z_{\rm cc}), \tag{18}$$

where $t_{\rm r}(z_{\rm cc})$ is the relaxation time, defined as the mean time between individual collisions, which we introduced briefly at the beginning of Sec. II A [see Eq. (1)]. Reference [82] gives a formula for $t_{\rm r}(z_{\rm cc})$ based on the properties of the primordial NFW halo,

$$t_r(z_{\rm cc}) \simeq 1.47~{\rm Gyr} \times \left(\frac{0.6}{C}\right)$$
$$\times \left(\frac{{\rm cm}^2/{\rm g}}{\sigma_{\rm c,0}(z_{\rm cc})}\right)\left(\frac{100~{\rm km/s}}{\sigma_{\rm vel}(z_{\rm cc})}\right)\left(\frac{10^7 M_\odot~{\rm kpc}^{-3}}{\rho_{\rm s}(z_{\rm cc})}\right), \tag{19}$$

where C is a fitting parameter, which we set to 0.57 following Ref. [82]. The parameter $\sigma_{c,0}$ is a type of cross section average given by

$$\sigma_{c,0}(z_{\rm cc}) = \frac{3}{2}\frac{\langle \sigma_{\rm visc} v^3 \rangle}{\langle v^3 \rangle}, \tag{20}$$

where $\sigma_{\rm visc} = \int d\sigma \sin^2\theta$ is the viscosity cross section with the scattering deflecting angle θ. The difference between $\sigma_{\rm visc}$ and $\sigma_{\rm T}$ is small and depends on the SIDM particle physics model, which gives the specific angular dependence for the differential cross section [33]. We assume that $\sigma_{\rm visc} \approx \sigma_{\rm T}$ for isotropic scattering, since our results are calibrated on the SIDM simulations by Ref. [13], which use elastic isotropic scattering using $\sigma_{\rm T}$. The cosmic time from the big bang until the formation of the black hole is finally given by

$$t_{\rm BH} = t(z_{\rm cc}) + 382 t_r(z_{\rm cc}), \tag{21}$$

where $t(z_{\rm cc})$ is the time at which the core-collapse transition begins, and is computed with Eq. (11).

The black hole is expected to form from material in the SMFP region. Reference [55] estimates the mass of this seed black hole $M_{\rm BH}$ based on the mass in the core, $M_{\rm core}$, that is in the SMFP regime. In the late stages of the core evolution, the $M_{\rm core}$-$\sigma_{\rm vel}$ relation determines the mass of the core at the relativistic instability, which occurs when $\sigma_{\rm vel} \approx c/3$. When the inner core is sufficiently dense, mass is continuously lost from its surface as outer layers cool and expand to join the outer core, with Ref. [54] predicting that the $M_{\rm core}$-$\sigma_{\rm vel}$ relation settles to $d\log M_{\rm core}/d\log(\sigma_{\rm vel}^2) \approx -0.85$. Therefore, the seed black hole mass is predicted to be

$$M_{\rm BH}(z^*) = M_{\rm core}(z^*)\left(\frac{\sigma_{\rm vel}^2(z^*)}{(c/3)^2~{\rm km}^2~{\rm s}^{-2}}\right)^{0.85}. \tag{22}$$

The region outside the collapsed core relaxes to a dynamically stable equilibrium system of particles that continue to orbit the central black hole, and subsequent interactions in this region will feed the black hole.

We should note that the behavior above has been developed for a system in isolation. In an evolving halo growing in a cosmological scenario, mass accretion might modify this behavior as we described in Sec. II D. A detailed treatment of the impact of cosmological accretion in the gravothermal fluid equations, and in particular on the scale of the SMFP region, goes beyond the scope of this work. Instead, we consider a simple approach in which we establish a range of plausible black hole masses by considering the epoch z^* at which Eq. (22) should be evaluated. A lower estimate for the seed black hole mass would be to set $z^* = z_{\rm cc}$, that is to assume that the scale of the core (LMFP region), and thus the scale of the collapsing SMFP region, is set by the properties of the halo, essentially

its mass $M(z_{cc})$, at the threshold time for the cusp-core transformation. This, however, ignores the fact that the halo mass grows during the cusp-core transformation and up to the point of collapse [z_{BH} given by Eq. (21)]; the size of the core should thus grow as well, affecting the scale of the collapsing region. An upper estimate for the seed black hole mass can then be given by setting $z^* = z_{BH}$, that is to assume that the size of the SMFP region is set by the last stages of the cusp-core-collapse evolution when the LMFP region (core) has grown to a size set by $M(z_{BH})$. We then bracket the plausible range of black hole masses by

$$M_{BH}(z^* = z_{cc}) < M_{BH} < M_{BH}(z^* = z_{BH}). \quad (23)$$

To estimate the core mass M_{core} in Eq. (22), we use the results of the gravothermal fluid approach in Ref. [82] [the same reference we use in Eq. (19)] where it is estimated that the maximal core size of the halo is

$$r_{core} \simeq 0.45 r_{-2} \quad (24)$$

before the collapse regime begins. This core radius is defined as the radius at which the local density is half that of the central density. The core size estimate in Eq. (24) is roughly consistent with simulation results where the core size is found to be $\lesssim r_{-2}$ [66,88]. Assuming that the region beyond the core remains essentially collisionless and with a profile that matches the NFW distribution, we then have

$$M_{core}(z^*) = M_{NFW}(r_{core}; z^*), \quad (25)$$

where $M_{NFW}(r)$ is given by the NFW radial density profile [Eq. (2)] for a halo with a mass $M_0(z^*)$ given by the mass accretion history formula [Eq. (9)] evaluated at z^* and a concentration $c(z^*)$ obtained from this mass using the model described in Sec. II C.

G. Impact of tidal stripping in the core-collapse phase

Tidal stripping is the process by which DM in the outskirts of a smaller halo is removed by tidal forces as it orbits within a larger host. References [14,58–60] argued that the tidal interactions with the halo and central galaxy of the host accelerate the core-collapse timescale by increasing the temperature, thus making heat outflow more efficient. These previous works have invoked tidal acceleration of core collapse as an explanation for the diversity of the MW's dwarf spheroidal galaxies (dSphs), based on constant cross section SIDM models with relatively low cross sections $\sigma_T/m_\chi \sim 1$–5 cm^2 g^{-1}, while the velocity-dependent model presented in Ref. [13] (see also Refs. [63,89]) relies on large cross sections $\sigma_T/m_\chi > 10$ cm^2 g^{-1} at the characteristic velocities of the dSphs to ensure core collapse.

In order to take into account the impact of tidal stripping in accelerating the core-collapse phase, we use the results

presented in Ref. [58], where tidal stripping is assumed to modify the NFW profile for $r > r_t$ in the following way: $\rho_{NFW}(r_t) \times (r_t/r)^{p_t}$ where $p_t = 5$ (based on [90]) and r_t is a truncation radius. This modified profile is a simple way to incorporate the impact of mass loss from the outer region in the timescale for collapse. We use the case of $r_t = r_{-2}$ and estimate the acceleration of the gravothermal catastrophe due to tidal stripping as

$$t_{coll,t} \approx \frac{1}{10} t_{coll}, \quad (26)$$

where t_{coll} is the timescale for core-collapse without the tidal effects [see Eq. (18)]. We only apply Eq. (26) for (sub) halos after the infall redshift into the MW halo, which for simplicity we fix to be $z_{infall} = 1$ for the entire (sub)halo mass range we consider in this work.[2]

III. RESULTS AND DISCUSSION

Our main goal is to investigate the consequences for velocity-dependent SIDM models, which invoke core collapse to explain the diversity of the MW satellite population (such as that in Ref. [13]), on the formation timescales for IMBHs and their masses.

Figure 3 shows the gravothermal collapse timescale—which is approximately the BH formation timescale—as a function of the halo's present day mass M_0 [Eq. (21)]. The solid lines—together with their shaded regions, which are given by the scatter in the cosmological concentration-mass relation—bracket the range of possible threshold epochs for the cusp-core-collapse evolution to begin (see Sec. II D), with the yellow corresponding to the assembly redshift of the primordial NFW halo $z_{cc} = z_{form} = z_{-2}$, while the blue corresponds to the epoch of the last major merger $z_{cc} = z_{LMM}$. The points where the right edge of the blue region, $t_{BH}(z_{LMM})$, and the right edge of the yellow region, $t_{BH}(z_{form})$, cross the $z = 0$ horizontal line roughly indicate the maximum mass of an isolated SIDM halo that could undergo core collapse by the present day, for both of these cases. For the case where $z_{cc} = z_{LMM}$, the maximum mass is $\sim 3 \times 10^9 M_\odot$, and for $z_{cc} = z_{form}$, it is $\sim 5 \times 10^{10} M_\odot$.

Since we are interested in the (sub)halos that could host the MW satellites, we have considered the impact of tidal effects in the timescale for black hole formation. As discussed in Sec. II G, for simplicity, we assume that all (sub)halos in the mass range considered in Fig. 3 become satellites at an infall redshift $z_{infall} = 1$ and that tidal forces by the host halo strip the material from the subhalo making the heat outflow from the center to the outskirts of the

---

[2]We note that, although the subhalo infall redshift distribution is broad, depending on the mass and orbit of individual subhalos, it roughly has a median value of $z \sim 1$ for the subhalo population of MW-size halos [88,91]. Since this is the population we are interested in, and since we are not considering detailed orbital properties, we fix $z_{infall} = 1$.
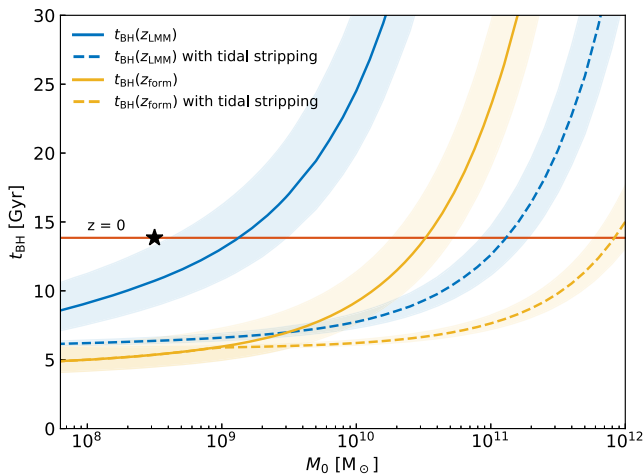
FIG. 3. The timescale to form a black hole [Eq. (21)] in SIDM halos as a function of the present-day halo mass M_0 (in isolation) for the velocity-dependent model vd100 [13,63] shown in Fig. 2. The yellow solid line shows the case where the starting time for the cusp-core-collapse evolution is set to the assembly/formation time of the primordial (pre-SIDM) CDM NFW halo, $z_{cc} = z_{form}$ (see Sec. II C and Fig. 1 where the yellow dotted line marks the cosmic time corresponding to z_{form}), while the blue solid line brackets the impact of cosmological accretion by setting $z_{cc} = z_{LMM}$, which is the epoch of the last major merger for given halo of mass M_0 (see Sec. II D and Fig. 1 where the blue dotted line marks the cosmic time corresponding to z_{LMM}). Shaded regions indicate a scatter of $\pm 10\%$ in the concentration-mass relation, e.g., Ref. [92]. Dashed lines are the corresponding cases including the acceleration of the collapse time driven by tidal stripping [58] assuming a mass-independent infall epoch of $z_{infall} = 1$ (see Sec. II G). The horizontal orange line indicates the age of the Universe. The black star symbol marks the transition mass where 50% of the subhalo population is estimated to be in the core-collapse regime according to Ref. [25]; see also Refs. [13,63].

subhalo much more efficient and thus reduce the timescale for collapse by a factor of 10 [see Eq. (26)]. This significant acceleration of the core-collapse phase is shown as dashed lines in Fig. 3, which shifts the upper limit of the mean mass of a halo that could undergo the core collapse to $\sim 2 \times 10^{11} M_\odot$ and $\sim 10^{12} M_\odot$ for the $t_{BH}(z_{LMM})$ and $t_{BH}(z_{form})$ cases, respectively.

The black star symbol in Fig. 3 indicates the halo mass at which 50% of (sub)halos are estimated to undergo core collapse by $z = 0$ according to the results in Ref. [25], which is based on the simulation of the vd100 model presented in Ref. [13]. The simulation results are not compatible with our modeling of an early cusp-core transformation (yellow lines), and therefore are also inconsistent with the acceleration effect due to tidal stripping (dashed lines). In the following, we discuss the effects that are likely behind this result:

(i) *Tidal acceleration of core-collapse?*—The impact of additional environmental effects taking place
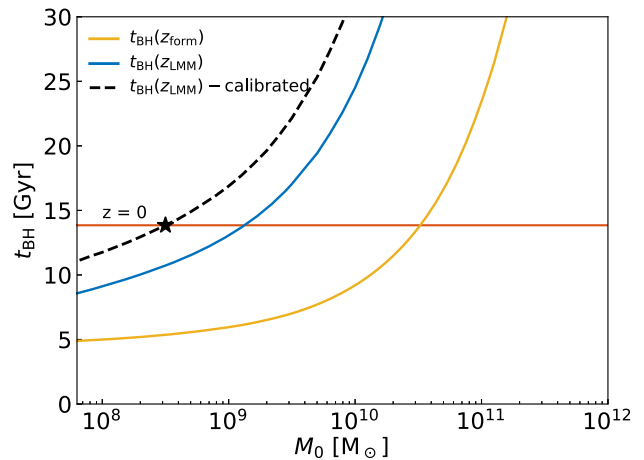

FIG. 4. Timescale for black hole formation in SIDM halos as a function of the present-day SIDM halo mass M_0 in isolation. The solid blue and yellow lines are the same as those in Fig. 3, which bracket the range of the possible threshold epochs for the cusp-core-collapse evolution to begin (for mean values of the concentration-mass relation). The dashed black line is $t_{BH}(z_{LMM})$ (i.e., the blue line) recalibrated to the simulation-based result of Ref. [25] (black star) with the calibration factor $C = 0.42$ [see Eq. (19)].

between a (sub)halo and the host halo during mergers, such as the evaporation of subhalo particles due to interactions with particles in the host, have been found to counteract the tidal stripping effect, delaying—or even disrupting—the core-collapse phase in models with low cross sections $\sigma_T/m_\chi \lesssim 10 \text{ cm}^2 \text{ g}^{-1}$ [93]. Recent N-body cosmological SIDM simulations of a MW-size halo and its subhalos with a cross section in the range $\sigma_T/m_\chi \sim 1$–$5 \text{ cm}^2 \text{ g}^{-1}$ confirm that subhalos do not experience core collapse; thus, larger values are required [62]. These recent results essentially rule out the constant cross section SIDM model as a viable possibility to explain the diversity of the MW satellite population and therefore strengthen the case for a velocity-dependent SIDM model with core collapse such as the one explored here based in Ref. [13] (see also Ref. [89]). Moreover, these results indicate that core collapse is not accelerated in the manner anticipated by the tidal stripping model in Ref. [58], and thus the dashed lines in Fig. 3 are likely overestimating its effectiveness.

(ii) *Cosmological accretion.*—Based on the previous discussion, the likely range of validity for our modeling is shown in Fig. 4, where we have omitted the concentration-mass relation scatter for clarity. It is clear that the simulation result (black star symbol) is closer to the model where the cusp-core transformation begins later, at $z_{cc} = z_{LMM}$. This finding supports the case for cosmological accretion increasing the time of the evolution of a SIDM halo

spent in the cored, quasistable regime, possibly due to energy injection from infall material into the central core as discussed in Sec. II D. A good match to the simulation results can then be achieved by setting $z_{cc} = z_{LMM}$ in our model and adjusting the C parameter in Eq. (19) to $C \approx 0.42$, which is shown as a dashed line in Fig. 4. We note that C is a calibration factor which, following Ref. [82], we had set to $C = 0.57$ in Fig. 3. This agrees with Ref. [83], which calibrated this parameter to a very similar value using the *isolated* SIDM simulations in Ref. [56]. We can then model the impact of cosmological accretion by either setting z_{cc} to z_{LMM} and makin a small modification to C (which is the case we adopt) or by modifying the value of C significantly (starting from $z_{cc} = z_{form}$) and invoking a needed recalibration of the parameter based on cosmological simulations. For our model, the latter case can be achieved by setting $z_{cc} = z_{-2}$ and fixing $C \approx 0.19$.

A. IMBHs in the ultrafaint galaxies

One consequence of invoking core collapse of SIDM halos to explain the diversity of inner DM densities in the MW satellite population is that those collapsed satellites will host central black holes. In particular, the vd100 model explored in Ref. [13] predicts that the dense ultrafaint galaxies, specifically, Segue I, Segue II, Willman I, and Boötes II, will be hosted by gravothermally collapsed subhalos (see Fig. 3 in Ref. [13]). We can compute the expected seed black hole mass for a collapsed SIDM halo in the vd100 model using our framework [see Eq. (22)]. This is shown in Fig. 5 as a function of M_0, the halo mass in isolation. We use our calibrated model with $z_{cc} = z_{LMM}$ and $C = 0.42$ and consider two cases to illustrate the impact of halo concentration in our results: (i) the solid violet line where halos have a mean concentration at M_0 (c_{mean}), and (ii) the dashed red line for halos with a concentration in excess[3] to the mean by 2σ ($c_{mean} + 2\sigma$). We first notice that, although the two lines representing these cases almost overlap with each other in Fig. 5, they in fact have a different slope and normalization since Eq. (22) depends (weakly) on concentration.[4] The net impact of concentration in the value of M_{BH} for a given halo mass is up to 5%.

Second, these predicted $M_{BH} - M_0$ relations are truncated at different halo masses, represented by the vertical downward arrows of the respective color. In the first case (using c_{mean}), this cutoff mass occurs at $M_0 \sim 3 \times 10^8 M_\odot$

---

[3]Recall we are using a mass-independent spread of the distribution of halos in the concentration-mass relation equivalent to 0.1 dex for 1σ of the distribution.

[4]This dependence is not strong because z_{BH} decreases with M_0; thus, the halo mass at this redshift, $M(z_{BH})$, increases with M_0, which makes the relevant concentration, $c(M(z_{BH}))$, almost independent of present-day halo mass M_0.
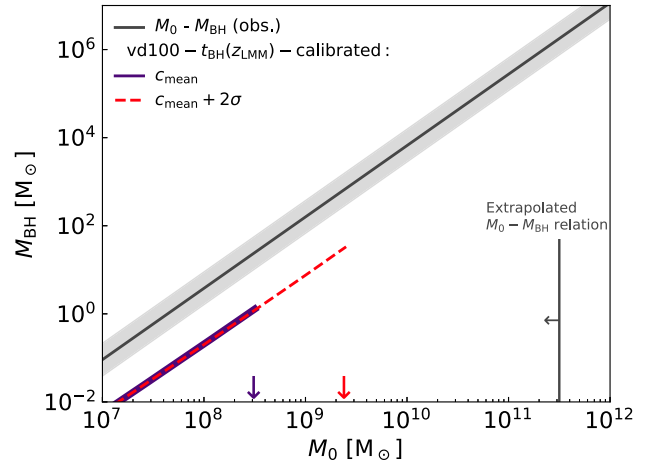

FIG. 5. Black hole mass–halo mass relation. The violet and red lines follow the estimate of the SIDM-core-collapse formula [Eq. (22)] with the former using the mean concentration-mass relation, while the latter uses a $+2\sigma$ value over the mean. For these cases, the arrows indicate the corresponding halo mass at which $t_{BH} = 0$, i.e., the mass at which 50% and 2.5% of the halo population at that mass is in the core-collapse regime, violet and red, respectively. The gray line is the extrapolation toward lower masses of the empirical relation for supermassiv black holes (SMBHs) in galaxies with halo masses $> 10^{11.5} M_\odot$, while the gray band represents the intrinsic scatter on this relation. adopted from Ref. [94].

(violet arrow) and can be interpreted as the mass at which 50% of the halos have core collapsed and 50% of the halos are still in the core phase. At higher (lower) masses, the fraction of core-collapsed halos is lower (higher) depending on halo concentration. For example, SIDM halos with $M_0 \sim 2 \times 10^9 M_\odot$ are only expected to have collapsed by the present day if they have a concentration larger than the mean by 2σ (red arrow), which represents only a small fraction (2.5%) of the halos at this mass. Therefore, in the vd100 model, only a small fraction of the massive (sub) halos in a MW-size system, which are expected to host the MW satellites, would have undergone core collapse.

As we discussed in Sec. II F, the range of plausible BH masses depends on the epoch (z^*) at which one should evaluate the relevant properties of the core that determine the SMFP region that collapses [Eq. (23)]. We established $z^* = z_{BH}$ as a reasonable choice, and it is the one that appears in Fig. 5 and in subsequent figures. Modifying this choice to the earliest plausible epoch $z^* = z_{cc}$ results instead in a smaller BH mass. For our default choice of $z_{cc} = z_{lmm}$, the difference is up to 30% over the relevant mass range with a weak dependence on halo mass.

The gray line in Fig. 5 represents the observed SMBH-mass–halo-mass relation for massive galaxies extrapolated to low masses [94], while the shaded gray band represents the intrinsic observational scatter. This relation and its spread have only been measured in halos larger than $10^{11.5} M_\odot$, and therefore we plot the extrapolated values

down to $10^7 M_\odot$ in Fig. 5. Such an extrapolation of the empirical $M_{BH} - M_0$ relation to the regime of dSphs ($M_0 \leq 10^{10} M_\odot$) would imply IMBHs in the range between $4-7 \times 10^3 M_\odot$. Remarkably, the slope of the predicted SIDM-driven $M_{BH} - M_0$ relation by Eq. (22) is very similar to that of the SMBH–halo mass relation, while the normalization is approximately 2 orders of magnitude lower.

An important aspect of our model is that it aims to explain the diversity of the MW satellite population, by invoking a velocity-dependent SIDM model. This model predicts that only a fraction of the satellites has undergone collapse and hosts IMBHs, specifically the least massive (in a model like vd100, those with $M_0 \lesssim 3 \times 10^9 M_\odot$). For instance, the massive central black hole inferred recently in the dSph Leo I with $M_{BH} \sim 3 \times 10^6 M_\odot$ [95] is too massive to lie in the extrapolated $M_{BH} - M_0$ relation, and the properties of the halo associated with Leo I, being one of the brightest satellites of the MW, would likely put it in a range of M_0 values close to, but nevertheless above, the threshold for collapse in the vd100 model. The significance of this issue depends on the specific velocity-dependent SIDM model assumed. For instance, in Ref. [89], Leo I is associated with a halo of an initial mass of $M_{200} \sim 3 \times 10^9 M_\odot$ that has gravothermally collapsed, according to the inferred velocity-dependent SIDM model tuned to explain the diversity of the dSph satellite population in that work. Nevertheless, the central black hole mass inferred for Leo I in Ref. [95] is several orders of magnitude larger than can be explained by gravothermal collapse alone and a significant growth of the seed black hole by other means would be required. For other bright dSphs, there might be a different type of conflict; for instance, Fornax and Ursa Minor have upper limits to the presence of a central black hole of around $M_{BH} \sim 3 \times 10^4 M_\odot$ [96,97]. These systems have M_0 values likely in the range around the threshold for core collapse. The precise value of M_0 inferred from the observed kinematics of the dSph depends on several quantities, such as the modeled DM profile and the orbital parameters, but with both being associated to cored systems in Ref. [13] for the vd100 model, they are not expected to be associated with collapsed subhalos.

We can classify SIDM models in the cross section–halo mass parameter space as to whether they generate a combination of cored and gravothermally collapsed halos in the dwarf galaxy regime. More specifically, we determine the normalization boundaries of the self-interacting cross section according to whether or not the gravothermal collapse regime is expected to occur in a fraction of the MW satellites' host subhalos. In practice, and for simplicity, we only consider SIDM models described by the classical velocity-dependent formula for a Yukawa-like interaction model σ_T [i.e., Eq. (12)], and, at first, we fix the relative velocity v_{max} at which $(\sigma_T v)$ peaks; in this way, we only vary the normalization σ_T^{max}.
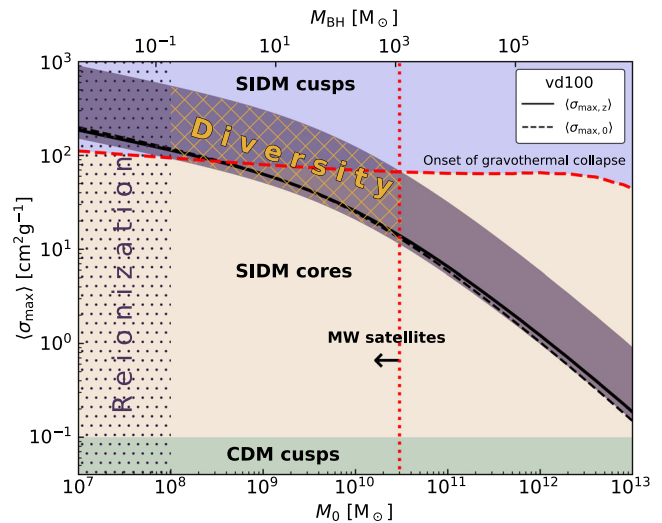

FIG. 6. The effective cross section as a function of a present-day SIDM halo mass M_0 in isolation. The black solid and dashed lines represent the effective cross section for the vd100 model evaluated at z_{LMM} and $z = 0$, respectively; we use our default model with $z_{cc} = z_{LMM}$ and calibrated to the simulation analysis in Ref. [25]. The x axis on the top shows the corresponding (SIDM-driven) black hole mass for a given M_0. The red dashed line (nearly horizontal) indicates the required cross section value for the onset of gravothermal collapse: SIDM-driven cuspy halos lie above (light violet), while SIDM cores lie below (beige) down to the point where the cross section is so low that halos are essentially CDM-like (light green). The red dotted line marks the upper mass for the dSph MW satellites to reside, e.g., Ref. [24], while the hashed/dotted region to the left starts at the mass where reionization significantly suppresses galaxy formation. The dark violet band indicates the region where vdSIDM models like vd100 but with different normalization produce a diverse MW satellite population hosted by halos that could either be cored or cuspy.

Figure 6 shows the effective cross section $\langle \sigma_{max} \rangle$ [thermal average of σ_T / m_χ at r_{max}; see Eq. (17)] as a function of a halo mass at the present day, M_0. The corresponding BH masses for a given M_0 are plotted on the top x axis using Eq. (22). The vd100 model results for the cases in which $\langle \sigma_{max} \rangle$ is computed at $z = z_{cc} = z_{LMM}(M_0)$ and at $z = 0$ are shown as the black solid and dashed lines, respectively. The former is the relevant cross section to set the core-collapse timescale—notice that the relevant epoch is a function of mass—while the latter is shown as reference to make the connection with Fig. 2, where the thermal average is evaluated at $z = 0$ for all halos.

The hashed/dotted region at $M_0 \leq 10^8 M_\odot$ marks an approximate lower limit on the mass of halos where galaxy formation is efficient; below this mass, heating during the epoch of reionization severely reduces the efficiency of cooling and subsequent star formation, e.g., Ref. [98]. The vertical dotted red line is an approximate upper limit on the mass of possible halo hosts for the population of dSph MW

satellites, e.g., Ref. [11]. Thus, the mass range 10^8–$3 \times 10^{10} M_\odot$ represents the region inhabited by the MW satellites. Different colored regions indicate the range of cross section values where halos with different inner density profiles reside. The light-green region at < 0.1 cm^2 g^{-1} is where DM is effectively collisionless, and thus the structure of all halos is indistinguishable from CDM (i.e., cuspy).[5] The light brown region in the middle is where SIDM models deviate significantly from CDM and predict quasiequilibrium cored halos. The red dashed line marks the effective cross section for the onset of gravothermal collapse to occur by $z = 0$; it demarcates the transition from cores to SIDM-driven core-collapsed (cuspy) halos (light violet region on the top of Fig. 6).

The violet band in Fig. 6 denotes the set of cross section normalization values [σ_T^{max} in Eq. (12)] in vd100-like models that generate a diverse population of MW satellites, i.e., where halos with cored and cuspy profiles coexist. This section of parameter space is highlighted with a yellow hatched region within the range of halo masses expected for the MW satellites. Given that the timescale for core-collapse depends on halo mass and concentration, we can estimate the fraction of halos of a given mass M_0 that have undergone gravothermal collapse by considering the probability distribution of concentrations for halos at a fixed mass, which according to simulations follows a log-normal distribution, e.g., Ref. [100],

$$P(\log_{10} c) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{\log_{10} c - \langle \log_{10} c \rangle}{\sigma} \right)^2 \right],$$

(27)

where $\langle \log_{10} c \rangle$ is the median value of the concentration (in logarithm) and σ is its standard deviation. In our framework, the former is given by the concentration-mass relation model described in Sec. II C, while the latter is taken to be mass independent and fixed to 0.1 dex based on simulations.

Figure 7 shows the fraction of core-collapsed halos as a function of M_0 for the vd100 model (black solid line) and for its variations with different cross section normalizations that result in a diverse MW satellite population (dark violet), i.e., for the corresponding models shown in dark violet in Fig. 6. For a model with the normalization of vd100, most of the massive (sub)halos ($> 10^9 M_\odot$) that are expected to host the MW satellites are predicted to be cored, with only $\sim 30\%$ having undergone core collapse. However, the breadth of the potential collapsed fraction values indicates that even a small renormalization of the

---

[5]The upper boundary here is approximate since small cores are expected even at such low cross sections; however, simulation results (e.g., Ref. [35]) have shown that these cores are too small at the scale of MW satellites to constitute a significant deviation from the CDM case (see also Ref. [99]).
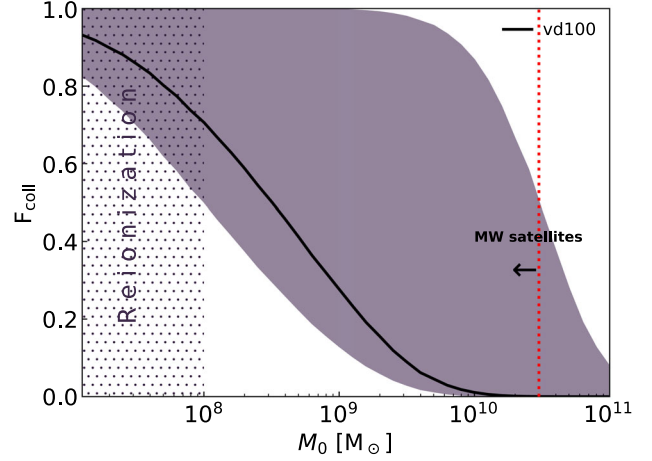

FIG. 7.   The fraction of core-collapsed halos as a function of M_0 for the vdSIDM models shown in Fig. 6, which have a diverse range of halo profiles (dark violet). The black solid line represents the benchmark vd100 model. As in Fig. 6, we use our default model with $z_{cc} = z_{LMM}$ and calibrated to the simulation analysis in Ref. [25].

model will significantly increase this fraction. We note that the prediction in Fig. 7 needs to be tested with MW-size simulations with a wide range of cross sections and with enough massive subhalos to sample the high-mass end of the subhalo population and subsequently recalibrated to match the full range of vd100-style models.

In principle, a complete exploration of the Yukawa-like interaction SIDM model in the classical regime requires variations of two parameters, v_{max} and the normalization σ_T^{max}. In addition, the family of SIDM models that generate gravothermally collapsed halos as well as cored halos is restricted by a couple of additional factors we have yet to consider. First, the impact of SIDM in DM structures at larger scales can be compared with observational probes of morphology based on lensing and x rays. These observations have constrained the transfer cross section $\sigma_T/m_\chi \lesssim 0.1$–1 cm^2 g^{-1} on massive galaxies and galaxy cluster scales (see Sec. I for references). In particular, we take the approximate constraint on the cross section in massive ellipticals set in Ref. [44] and more recently in Ref. [45] of $\langle \sigma_{max} \rangle \leq 1$ cm^2 g^{-1} at halo mass $M_0 = 10^{13} M_\odot$. Second, the SIDM cross section is subject to an upper limit set by the quantum zero-energy resonance, known as the unitarity bound, e.g., Ref. [101]. When the cross section saturates the unitarity bound, it is parametrized solely by the DM mass and is given by

$$\sigma_u = \frac{4\pi}{k^2},$$

(28)

where the relative momentum of the scattering particles k is defined as $k = (m/2)v$. We can then use σ_u to set an upper bound for the cross section σ_T^{max} that peaks at v_{max}:
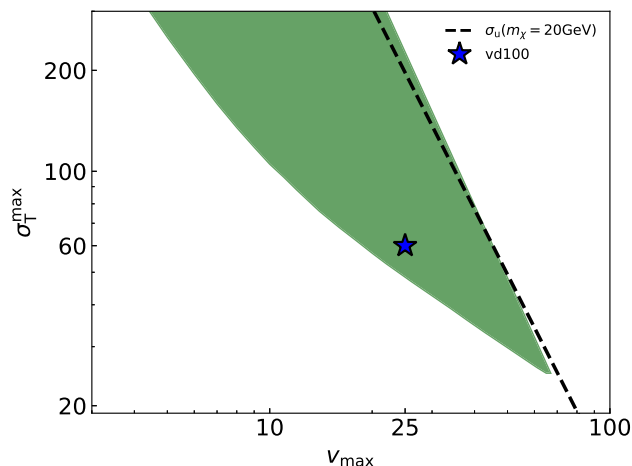
FIG. 8. The viable values of the parameter space that define the vdSIDM models studied here [Eq. (12)]: the relative velocity v_{max} where the transfer cross section σ_T^{max} peaks. The highlighted region is that of models that produce a bimodal distribution of halos (green area) and satisfies the constraint from elliptical galaxies $\langle \sigma_{max} \rangle \leq 1 \text{ cm}^2 \text{ g}^{-1}$ for a halo mass of $M_0 = 10^{13} M_\odot$. The black dashed line is the unitarity bound for the SIDM cross section for a particle mass $m_\chi = 20$ GeV. The blue star marks the vd100 model.

$$(\sigma_T v) \leq \sigma_T^{max} v_{max} \leq \sigma_u v_{max}. \qquad (29)$$

In Fig. 8, we show the range of viable parameter space, in peak relative velocity v_{max} and normalization σ_T^{max}, of the Yukawa-like SIDM models [Eq. (12)] that satisfy two conditions: having a diverse halo population (i.e., cored and gravothermally collapsed halos) in the range of masses suitable to host the MW satellite population and satisfying the constraint from elliptical galaxies. The unitarity bound (for $m_\chi = 20$ GeV as a reference) is marked as a dashed line, and the pair of σ_T^{max} and v_{max} corresponding to vd100 is marked with the blue star.

IV. CONCLUSIONS

If dark matter is made of particles that can strongly self-interact and therefore can be described as self-interacting DM, the nonlinear evolution of DM halos consists of two phases: a cusp-core transformation in which the originally cuspy halo develops a central isothermal core that is in quasiequilibrium and remains in cored configuration for several Gyr followed by a rapid gravothermal core-collapse phase. The ultimate consequence of this SIDM-driven collapse is the formation of a black hole with a mass that is a fraction of the central DM core mass [54]. If the cross section is strongly velocity dependent—and thus halo mass dependent—then the population of halos today is expected to exhibit a wide diversity of central density profiles. This behavior has been invoked as a viable way to explain the diversity of inner DM densities in the Milky Way satellites [13,64,89].

Velocity-dependent SIDM models are the natural result of several particle physics models, e.g., those with new light mediators that produce an effective Yukawa-like interaction between DM particles (for a review, see, e.g., Ref. [33]). These models also have the advantage of avoiding the stringent constraints on the cross section from gravitational lensing, x-ray morphology, and dynamical analysis in cluster mergers and elliptical galaxies, which limits $\sigma_T/m_\chi \lesssim 1 \text{ cm}^2 \text{ g}^{-1}$ at these scales [38–45], well below the threshold for gravothermal collapse. In this SIDM scenario, the structure of DM today is indistinguishable from the cold dark matter scenario at cluster scales. The deviation from CDM grows at smaller scales, starting with the development of spherical isothermal cores in the center of 10^{11}–$10^{12} M_\odot$ halos and followed by the onset of gravothermal collapse for the halos of dwarf spheroidal galaxies, a fraction of which should host SIDM-generated intermediate mass black holes.

In this work, we develop an analytical framework to predict the timescales and mass scales for the formation of IMBHs in SIDM halos, which includes the different stages in the cusp-core-collapse evolution of SIDM halos in a cosmological setting (Sec. II). This framework is calibrated to a high-resolution simulation of a benchmark velocity-dependent SIDM model (vd100; see Fig. 2), which has been tuned to produce a large diversity in the MW satellite population [13,25,63]. Our main results are summarized as follows:

(i) The black hole formation (gravothermal collapse) timescale as a function of present day halo mass (in isolation) is shown in Fig. 3. We consider two starting redshifts that bracket the range of the possible threshold epochs (z_{cc}) when the cusp-core-collapse evolution starts (see Sec. II D). The assembly/formation redshift of the primordial CDM NFW halo $z_{cc} = z_{form} = z_{-2}$ (yellow lines; defined as the time of assembly of the central region of the halo within its scale radius according to the model by Ref. [71]), and the redshift of the last major merger of the halo $z_{cc} = z_{LMM}$ (blue lines; takes into account cosmological accretion and it is computed using the simulation results of Ref. [77]). We also consider the possible acceleration of the collapse timescale driven by tidal stripping following [58] (dashed lines in Fig. 3; see Sec. II G).

(ii) We compare our results with the mass threshold ($\sim 3 \times 10^8 M_\odot$), where most of the MW (sub)halos are observed to have undergone gravothermal collapse according to the vd100 SIDM simulation analyses made in Ref. [63] and more recently in Ref. [25]. We find that our modeling can be fitted to this result by a small recalibration of the free parameter C in the gravothermal fluid formalism [Eq. (19)] by choosing $z_{cc} = z_{LMM}$ and by assuming that tidal stripping has no impact on the collapse

time (see Fig. 4). This choice is seemingly consistent with previous expectations that the core phase is delayed by cosmological infall [76] and supports the recent detailed simulation work by Ref. [61] that suggests the impact of tidal stripping in accelerating core collapse is likely overestimated in Ref. [58]— from which we developed our incorporate tidal stripping model—due to the competing environmental effect of subhalo heating through collisions with host halo particles.

(iii) We show the black hole mass M_{BH} as a function of the present day halo mass M_0 in Fig. 5 (violet and red lines). This estimated seed black hole mass is obtained with Eq. (22), which is derived by following the evolution of the part of the core that collapses to high central densities within the gravothermal fluid formalism. The development of the relativistic instability ultimately leads to the formation of the black hole [55]. Remarkably, the slope of the $M_{BH} - M_0$ SIDM-core-collapse relation is similar to that of the extrapolated SMBH–halo mass empirical relation found in massive galaxies [94]. This mechanism could then potentially constitute a continuation of the empirical relation towards the regime of dSphs, although the normalization is 2 orders of magnitude below the expectation, and thus the seed SIDM-driven black hole would need to grow substantially to satisfy such a scenario and would have to be rapidly accelerated compared to our predictions.

(iv) We also consider the impact of the cosmological scatter in the concentration-mass relation in our results by first considering a 10% scatter on the concentration at a fixed mass (today) (e.g., Ref. [92]). We find that more concentrated halos have their core collapse several Gyrs earlier (shaded areas in Fig. 3). The fraction of halos that is expected to collapse strongly depends on the range of concentrations available to the halo population at a given mass (see Fig. 7). The predicted black hole mass is, however, mainly set by halo mass and is only weakly affected by concentration (at the percent level).

(v) Finally, we estimated the range of self-scattering cross sections that result in a diverse MW satellite population, i.e., that generate both cored and core-collapsed host halos for MW satellites. We first consider SIDM models with the same velocity dependence as vd100 but with different normalization [see Eq. (12)]. The results are shown in Figs. 6 and 7. We found that the vd100 model has a normalization that is close to the lower limit to exhibit diversity with most of the subhalos in MW-size systems being cored, especially for halos $M_0 > 10^{10} M_\odot$, of which ∼70% are cored. The latter is expected since our default choice of parameters is calibrated to the recent results in Ref. [25], which

agrees qualitatively on the scarcity of massive core-collapsed subhalos in vd100. As noted by Ref. [25], this represents a potential issue with models such as vd100 since it would be more natural to expect the bright MW dSphs to be assigned to the most massive subhalos, which in this case are likely to be cored, and thus inconsistent with the properties of dense bright dSphs. We show that such a potential issue can be alleviated by a relatively small increase in the cross section normalization of the vd100 model, since the fraction of massive core-collapsed halos increases rapidly with larger cross sections (see Fig. 7). To expand upon this result, we explore in Fig. 8 the parameter space of the Yukawa-like SIDM model [Eq. (12)] to find the viable values of the peak velocity v_{max} and the normalization σ_T^{max} that satisfy: (i) having a diverse MW satellite population (as described before), (ii) satisfying constraints at larger scales, in particular those set by elliptical galaxies [44,45], and (iii) satisfying the constraint set on the cross section by the unitarity bound. These conditions set an upper bound for v_{max} to be $v_{max} < 70~\text{km s}^{-1}$ with a very narrow range of possible σ_T^{max} values, while an ever wider range of σ_T^{max} is suitable for smaller v_{max} values down to $v_{max} < 5~\text{km s}^{-1}$. To ascertain whether a single SIDM model can fit all MW satellites simultaneously requires much more precise estimates for the fraction of collapsed-to-total MW satellites.

In Appendix B, we discuss different values of calibration parameter C in the gravothermal fluid model and their impact on our results. Based on previous studies, this parameter has a plausible range between 0.4–0.75, with our calibrated value being at the low end. However, such a value seems to be appropriate/favored by analyses based on cosmological simulation [102] (see also Ref. [83]). Nevertheless, we explore in Appendix B how choosing a value at the high end $C = 0.75$ would impact our key results. We find that, although the timescale for black hole formation significantly changes with such a value of C (see Fig. 10), the impact on our key result is modest; that is, the region in the plane $\sigma_{max} - M_0$ that would result in a diverse (i.e., cuspy and cored) MW satellite population is shifted downward overall by about a factor of ∼1.5 in the normalization of the cross section (see Fig. 11).

If SIDM is to be invoked to explain the diversity of the dSph MW satellite population, then the presence of IMBHs in the center of cored-collapsed (cuspy) halos is unavoidable. Based on our work, the range of seed IMBH masses for the dSph MW satellites that can potentially be hosted by cored-collapsed SIDM halos is in the range 0.1–$1000 M_\odot$ (Fig. 6). The existence of such IMBHs may be verified by detailed observations of dSph kinematics with the upcoming generations of extremely large telescopes [103]. Due to their high spatial resolution, they should be able to find

$< 10^5 M_\odot$ black holes in $< 10^9 M_\odot$ in nearby galaxies through high-precision proper motion measurements [103]. For example, to reach the range of 10^3–$10^4 M_\odot$, Ref. [104] suggests looking at the disruption of stars passing close to IMBHs. Under certain conditions, IMBHs should typically acquire companions with orbital periods of years, corresponding to semimajor axes of $\sim$5–10 mas for $\sim 10^3 M_\odot$ IMBHs.

More precise observational kinematic data for the dSphs in the future will constrain further the inner DM content of these systems, possibly establishing their inner DM profile. This will conclusively determine the significance of the diversity problem and constrain different classes of solutions. On the theoretical side, the predictions of vdSIDM models such as those considered in this work need to be complemented with more dedicated simulations that explore the relevant parameter space of cross sections that contain the gravothermal collapse regime [13,64] and can be complemented with a semianalytical revision of the predicted seed black hole mass in the seminal work by Ref. [54] within a full cosmological setting. Finally, a key aspect to explore is the interplay between the formation and evolution of the visible baryonic galaxy with the collapsing SIDM core and its central black hole.

ACKNOWLEDGMENTS

APPENDIX A

In this work, we computed the relaxation timescales for self-scattering and core-collapse using Eqs. (18) and (19), respectively. These equations are adapted from Ref. [82], which incorporates a velocity-dependent SIDM cross
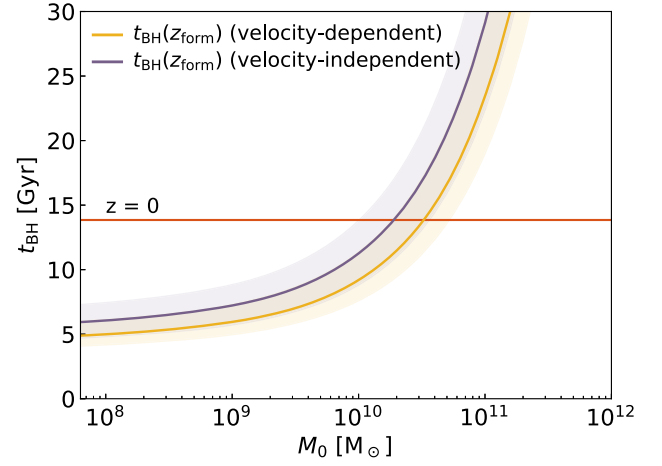

FIG. 9. Timescale to form a black hole in SIDM halos as a function of the present-day halo mass M_0 (in isolation). The yellow line [Eq. (21)] represents the cusp-core-collapse evolution timescale chosen for this work (equal to the curve of the same color in Fig. 3), and the violet line [Eq. (A3)] is based on Ref. [57]. As in Fig. 3, shaded bands indicate a concentration scatter of $\pm 10\%$ in the concentration-mass relation (e.g., Ref. [92]). The horizontal line indicates the age of the Universe at $z = 0$.

section into the gravothermal fluid formalism. In this appendix, we consider the impact on our results of using an alternative set of formulas developed by Ref. [57], where the SIDM cross section is assumed to be constant. According to this reference, the cusp-core-collapse timescale is given by

$$t_{\text{coll}} = 455.65 t_r(z_{\text{cc}}), \qquad (A1)$$

where $t_r(z_{\text{cc}})$ is the relaxation time derived for velocity-independent cross sections:

$$
\begin{aligned}
t_r(z_{\text{cc}}) &= \frac{1}{a\sigma_{\max}} \left(\frac{k(c)^2}{4\pi G^3}\right)^{1/6} \delta_{\text{char}}^{-7/6} \rho_{\text{crit}}(z_{\text{cc}})^{-7/6} M_0^{-1/3} \\
&= 0.310 \text{ Myr} \times \left(\frac{M_0}{10^{12} M_\odot}\right)^{-1/3} \left(\frac{k(c(z_{\text{cc}}))}{k(9)}\right)^{3/2} \left(\frac{c(z_{\text{cc}})}{9}\right)^{-7/2} \left(\frac{\rho_{\text{crit}}(z_{\text{cc}})}{\rho_{\text{crit}}(z=15)}\right)^{-7/6} \left(\frac{\langle \sigma_{\max} \rangle}{1 \text{ cm}^2/\text{g}}\right)^{-1}. \quad (A2)
\end{aligned}
$$

Then, the time to form a black hole is given by

$$t_{\text{BH}}(M_0, \sigma_{\max}) = t(z_{\text{cc}}) + 455.65 t_r(z_{\text{cc}}). \qquad (A3)$$

In Fig. 9, we compare the black hole formation timescale derived self-consistently for velocity-dependent cross sections [Eq. (21)] with the case derived assuming the velocity-independent cross section [Eq. (A3)]. For this comparison, we have used the case with $z_{\text{cc}} = z_{\text{form}} = z_{-2}$ uncalibrated and without tidal stripping (i.e., equivalent to

the solid yellow line in Fig. 3). The $t_{\text{BH}} - M_0$ curves have a similar shape in both cases across all the explored mass range, but with a $\sim$1 Gyr difference in normalization. For the purposes of this work, such a difference could be absorbed almost completely in the calibration factor C in Eq. (18).

APPENDIX B

In the gravothermal fluid formalism, the heat conductivity κ is a key quantity in the time evolution of the SIDM

halo. In the LMFP regime, it can be derived by dimensional analysis, but it carries an unknown parameter C that is the order of unity, which cannot be derived from first principles, e.g., Refs. [54,56–58,82,83]. To determine C, all these studies compared the evolution of the halo density profile given by the gravothermal fluid model to that obtained from different types of N-body simulations, where hard-sphere elastic scattering interactions were implemented. Most of these studies used isolated idealized simulations with only Refs. [58,82,83] using the cosmological constant–cross section SIDM simulations presented in Refs. [56,105]. For large cross sections that have entered the regime of core collapse, the latter studies suggest that a value of $C = 0.45$ is a better fit to cosmological simulations, while previous analyses based on isolated simulations preferred values around 0.6–0.75.

In this work, we used the formula for the relaxation time [Eq. (19)] and, initially, the fitting parameter $C = 0.57$ calibrated for the velocity-dependent SIDM model in Ref. [82] using isolated simulations. This value was roughly in between the values explored in the literature as discussed above. After comparing the timescale for gravothermal collapse (black hole formation) with the results from Ref. [25] based on the high-resolution velocity-dependent SIDM simulation performed in Ref. [13], we found that a value of $C = 0.42$ is a more accurate fit (see Fig. 4). This is in qualitative agreement with the analysis in Ref. [83], which found $C = 0.45$ consistent with the cosmological simulation they analyzed. However, we note that there are different methods used for calibration. Our calibration is based on comparing the mass threshold where $\geq 50\%$ subhalos of a MW-size host have collapsed, while Ref. [83] uses the density profile of the dwarf-size main host halo. More recently, Ref. [102] performed a new velocity-dependent SIDM cosmological simulation with a resolution similar to that of Ref. [13]. Using a comparison similar to ours, albeit with a different methodology, between the gravothermal fluid model and the simulation, Ref. [102] found find that a value of $C = 0.75$ overestimates the number of subhalos that should collapse by $z = 0$ by around a factor of 2. Thus, this result also suggests that a lower value of C fits cosmological simulations better, which is qualitatively in agreement with our finding.

In order to obtain a precise value of the parameter C that is appropriate for subhalos in a cosmological setting, it would be necessary to perform a detailed analysis across multiple high-resolution simulations that considers variations across different velocity-dependent cross sections and a large exploration across different host halo masses. This hypothetical study requires many more simulations than have been performed to date; currently, only two exhibit the required mass resolution (Refs. [13,102]). Such an analysis is beyond the scope of this paper. Nevertheless, we can take into account the uncertainty in the value of C by considering its impact on the key results of our work.
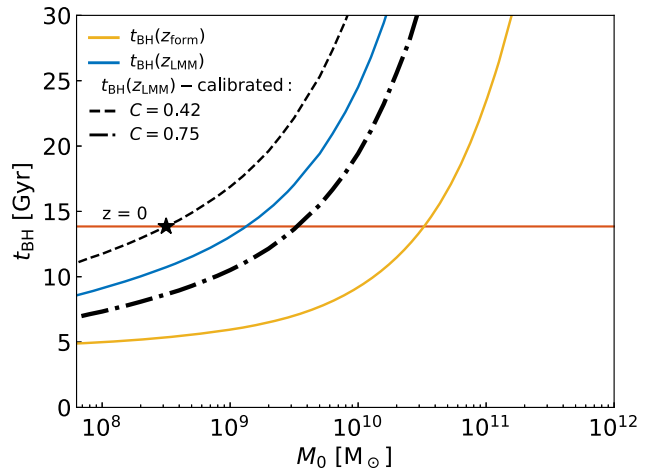

FIG. 10. The same as Fig. 4, but with the addition of the thick dot-dashed black line, which is the same as the calibrated model used for our key results (thin dashed black line) but with $C = 0.75$ instead.

We do this in this appendix by comparing the predictions of our model with the calibrated $C = 0.42$ value we have used in our key results, and with a value of $C = 0.75$. The latter is the highest value for this parameter explored in Refs. [54,58] and as we argued before is at the high end of plausible values as found in Ref. [102].

The impact on the timescale for black hole formation due to variations in C across this plausible range of values is shown in Fig. 10, specifically as the range between the dashed and thick dot-dashed black lines. We note that the resulting range is similar to the spread observed due to the 10% cosmological scatter in the concentration-mass relation (see Fig. 3) with an increase in the threshold mass that divides subhalos that have collapsed from those that have not by about an order of magnitude. Naturally, setting a higher value for the C parameter lowers the cross section values required to reach the threshold for the onset of gravothermal collapse by about a factor of 1.5; this is shown in Fig. 11 through the difference between the thin and thick red dashed lines. The overall change would then be a shift downward of the region that is expected to have a diverse MW subhalo satellite population of cores and cusps. We conclude that the overall impact of the C parameter uncertainty on our key plot Fig. 6 is a factor of ~ 1.5 on the normalization of the transfer cross section that results in a diverse MW satellite population. Finally, we show in Fig. 12 the fraction of core-collapsed halos as a function of M_0 when setting $C = 0.75$ instead of our default value of $C = 0.42$ as a thick dot-dashed black line. With this highest value of C, 80% of the (sub)halos with $> 10^9 M_\odot$ are expected to have undergone core collapse. This is around 50% higher than in the case of $C = 0.42$. Setting other values of C in the plausible range would result in different core-collapsed halo fractions in between the two black solid and dot-dashed lines.
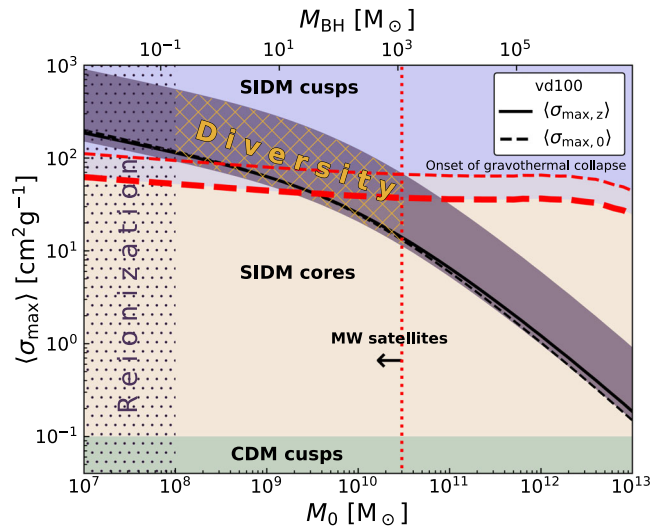
FIG. 11. The same as Fig. 6, but with the addition of the nearly horizontal thick red dashed line—the required cross section value for the onset of gravothermal collapse when $C = 0.75$ is used instead of $C = 0.42$, which is the calibrated value used for our key results (thin red dashed line).
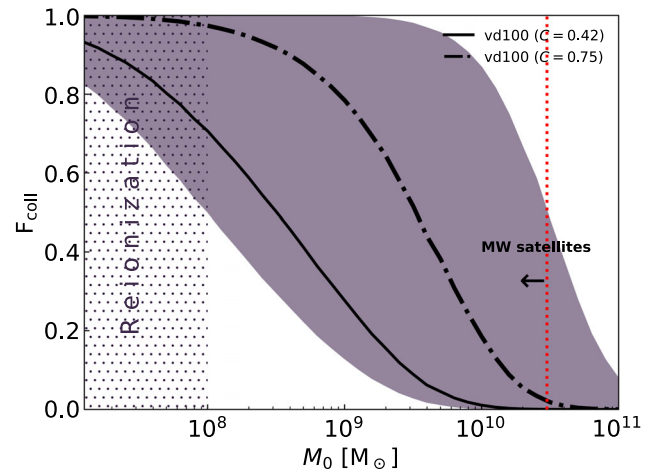

FIG. 12. The same as Fig. 7, but with the addition of the thick black dot-dashed line—the fraction of core-collapsed halos as a function of M_0 when $C = 0.75$ is used instead of $C = 42$, as in our key result (solid black line).

[1] V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly *et al.*, Nature (London) **435**, 629 (2005).

[2] J. S. Bullock and M. Boylan-Kolchin, Annu. Rev. Astron. Astrophys. **55**, 343 (2017).

[3] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, Mon. Not. R. Astron. Soc. **415**, L40 (2011).

[4] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, Mon. Not. R. Astron. Soc. **422**, 1203 (2012).

[5] M. G. Walker and J. Peñarrubia, Astrophys. J. **742**, 20 (2011).

[6] A. Agnello and N. W. Evans, Astrophys. J. Lett. **754**, L39 (2012).

[7] M. A. Breddels, A. Helmi, R. C. E. van den Bosch, G. van de Ven, and G. Battaglia, Mon. Not. R. Astron. Soc. **433**, 3173 (2013).

[8] G. Torrealba, S. E. Koposov, V. Belokurov, and M. Irwin, Mon. Not. R. Astron. Soc. **459**, 2370 (2016).

[9] G. Torrealba, V. Belokurov, S. E. Koposov, T. S. Li, M. G. Walker, J. L. Sanders, A. Geringer-Sameth, D. B. Zucker, K. Kuehn, N. W. Evans *et al.*, Mon. Not. R. Astron. Soc. **488**, 2743 (2019).

[10] A. Fattahi, J. F. Navarro, C. S. Frenk, K. A. Oman, T. Sawala, and M. Schaller, Mon. Not. R. Astron. Soc. **476**, 3816 (2018).

[11] R. Errani, J. Peñarrubia, and M. G. Walker, Mon. Not. R. Astron. Soc. **481**, 5073 (2018).

[12] J. I. Read, M. G. Walker, and P. Steger, Mon. Not. R. Astron. Soc. **484**, 1401 (2019).

[13] J. Zavala, M. R. Lovell, M. Vogelsberger, and J. D. Burger, Phys. Rev. D **100**, 063007 (2019).

[14] M. Kaplinghat, M. Valli, and H.-B. Yu, Mon. Not. R. Astron. Soc. **490**, 231 (2019).

[15] K. A. Oman, J. F. Navarro, A. Fattahi, C. S. Frenk, T. Sawala, S. D. M. White, R. Bower, R. A. Crain, M. Furlong, M. Schaller *et al.*, Mon. Not. R. Astron. Soc. **452**, 3650 (2015).

[16] I. M. E. Santos-Santos, J. F. Navarro, A. Robertson, A. Benítez-Llambay, K. A. Oman, M. R. Lovell, C. S. Frenk, A. D. Ludlow, A. Fattahi, and A. Ritz, Mon. Not. R. Astron. Soc. **495**, 58 (2020).

[17] J. F. Navarro, V. R. Eke, and C. S. Frenk, Mon. Not. R. Astron. Soc. **283**, L72 (1996).

[18] A. Pontzen and F. Governato, Mon. Not. R. Astron. Soc. **421**, 3464 (2012).

[19] J. D. Burger and J. Zavala, Mon. Not. R. Astron. Soc. **485**, 1008 (2019).

[20] J. D. Burger and J. Zavala, Astrophys. J. **921**, 126 (2021).

[21] A. Zolotov, A. M. Brooks, B. Willman, F. Governato, A. Pontzen, C. Christensen, A. Dekel, T. Quinn, S. Shen, and J. Wadsley, Astrophys. J. **761**, 71 (2012).

[22] S. Garrison-Kimmel, P. F. Hopkins, A. Wetzel, J. S. Bullock, M. Boylan-Kolchin, D. Kereš, C.-A. Faucher-Giguère, K. El-Badry, A. Lamberts, E. Quataert *et al.*, Mon. Not. R. Astron. Soc. **487**, 1380 (2019).

[23] A. Di Cintio, C. B. Brook, A. V. Macciò, G. S. Stinson, A. Knebe, A. A. Dutton, and J. Wadsley, Mon. Not. R. Astron. Soc. **437**, 415 (2014).

[24] R. Errani, J. F. Navarro, R. Ibata, and J. Peñarrubia, Mon. Not. R. Astron. Soc. **511**, 6001 (2022).

[25] M. R. Lovell and J. Zavala, Mon. Not. R. Astron. Soc. **520**, 1567 (2023).

[26] D. N. Spergel and P. J. Steinhardt, Phys. Rev. Lett. **84**, 3760 (2000).

[27] R. Davé, D. N. Spergel, P. J. Steinhardt, and B. D. Wandelt, Astrophys. J. **547**, 574 (2001).

[28] P. Colín, V. Avila-Reese, O. Valenzuela, and C. Firmani, Astrophys. J. **581**, 777 (2002).

[29] M. Vogelsberger, J. Zavala, and A. Loeb, Mon. Not. R. Astron. Soc. **423**, 3740 (2012).

[30] M. Rocha, A. H. G. Peter, J. S. Bullock, M. Kaplinghat, S. Garrison-Kimmel, J. Oñorbe, and L. A. Moustakas, Mon. Not. R. Astron. Soc. **430**, 81 (2013).

[31] G. A. Dooley, A. H. G. Peter, M. Vogelsberger, J. Zavala, and A. Frebel, Mon. Not. R. Astron. Soc. **461**, 710 (2016).

[32] V. H. Robles, T. Kelley, J. S. Bullock, and M. Kaplinghat, Mon. Not. R. Astron. Soc. **490**, 2117 (2019).

[33] S. Tulin and H.-B. Yu, Phys. Rep. **730**, 1 (2018).

[34] M. Vogelsberger, J. Zavala, K. Schutz, and T. R. Slatyer, Mon. Not. R. Astron. Soc. **484**, 5437 (2019).

[35] J. Zavala, M. Vogelsberger, and M. G. Walker, Mon. Not. R. Astron. Soc. **431**, L20 (2013).

[36] P. Creasey, O. Sameie, L. V. Sales, H.-B. Yu, M. Vogelsberger, and J. Zavala, Mon. Not. R. Astron. Soc. **468**, 2283 (2017).

[37] T. Ren, A. Kwa, M. Kaplinghat, and H.-B. Yu, Phys. Rev. X **9**, 031020 (2019).

[38] A. Robertson, R. Massey, and V. Eke, Mon. Not. R. Astron. Soc. **465**, 569 (2017).

[39] A. Robertson, D. Harvey, R. Massey, V. Eke, I. G. McCarthy, M. Jauzac, B. Li, and J. Schaye, Mon. Not. R. Astron. Soc. **488**, 3646 (2019).

[40] D. Harvey, A. Robertson, R. Massey, and I. G. McCarthy, Mon. Not. R. Astron. Soc. **488**, 1572 (2019).

[41] K. E. Andrade, J. Fuson, S. Gad-Nasr, D. Kong, Q. Minor, M. G. Roberts, and M. Kaplinghat, Mon. Not. R. Astron. Soc. **510**, 54 (2022).

[42] D. Eckert, S. Ettori, A. Robertson, R. Massey, E. Pointecouteau, D. Harvey, and I. G. McCarthy, Astron. Astrophys. **666**, A41 (2022).

[43] X. Shen, T. Brinckmann, D. Rapetti, M. Vogelsberger, A. Mantz, J. Zavala, and S. W. Allen, Mon. Not. R. Astron. Soc. **516**, 1302 (2022).

[44] A. H. G. Peter, M. Rocha, J. S. Bullock, and M. Kaplinghat, Mon. Not. R. Astron. Soc. **430**, 105 (2013).

[45] G. Despali, L. G. Walls, S. Vegetti, M. Sparre, M. Vogelsberger, and J. Zavala, Mon. Not. R. Astron. Soc. **516**, 4543 (2022).

[46] M. Valli and H.-B. Yu, Nat. Astron. **2**, 907 (2018).

[47] J. I. Read, M. G. Walker, and P. Steger, Mon. Not. R. Astron. Soc. **481**, 860 (2018).

[48] S. Y. Kim and A. H. G. Peter, arXiv:2106.09050.

[49] J. L. Feng, M. Kaplinghat, H. Tu, and H.-B. Yu, J. Cosmol. Astropart. Phys. 07 (2009) 004.

[50] M. R. Buckley and P. J. Fox, Phys. Rev. D **81**, 083522 (2010).

[51] A. Loeb and N. Weiner, Phys. Rev. Lett. **106**, 171302 (2011).

[52] D. Lynden-Bell and R. Wood, Mon. Not. R. Astron. Soc. **138**, 495 (1968).

[53] P. Hut, S. McMillan, J. Goodman, M. Mateo, E. S. Phinney, C. Pryor, H. B. Richer, F. Verbunt, and M. Weinberg, Publ. Astron. Soc. Pac. **104**, 981 (1992).

[54] S. Balberg, S. L. Shapiro, and S. Inagaki, Astrophys. J. **568**, 475 (2002).

[55] S. Balberg and S. L. Shapiro, Phys. Rev. Lett. **88**, 101301 (2002).

[56] J. Koda and P. R. Shapiro, Mon. Not. R. Astron. Soc. **415**, 1125 (2011).

[57] J. Pollack, D. N. Spergel, and P. J. Steinhardt, Astrophys. J. **804**, 131 (2015).

[58] H. Nishikawa, K. K. Boddy, and M. Kaplinghat, Phys. Rev. D **101**, 063009 (2020).

[59] F. Kahlhoefer, M. Kaplinghat, T. R. Slatyer, and C.-L. Wu, J. Cosmol. Astropart. Phys. 12 (2019) 010.

[60] O. Sameie, H.-B. Yu, L. V. Sales, M. Vogelsberger, and J. Zavala, Phys. Rev. Lett. **124**, 141102 (2020).

[61] Z. C. Zeng, A. H. G. Peter, X. Du, A. Benson, S. Kim, F. Jiang, F.-Y. Cyr-Racine, and M. Vogelsberger, Mon. Not. R. Astron. Soc. **513**, 4845 (2022).

[62] M. Silverman, J. S. Bullock, M. Kaplinghat, V. H. Robles, and M. Valli, Mon. Not. R. Astron. Soc. **518**, 2418 (2022).

[63] H. C. Turner, M. R. Lovell, J. Zavala, and M. Vogelsberger, Mon. Not. R. Astron. Soc. **505**, 5327 (2021).

[64] C. A. Correa, M. Schaller, S. Ploeckinger, N. Anau Montel, C. Weniger, and S. Ando, Mon. Not. R. Astron. Soc. **517**, 3045 (2022).

[65] D. Lynden-Bell and R. Wood, Mon. Not. R. Astron. Soc. **138**, 495 (1968).

[66] M. Vogelsberger, J. Zavala, C. Simpson, and A. Jenkins, Mon. Not. R. Astron. Soc. **444**, 3684 (2014).

[67] J. F. Navarro, C. S. Frenk, and S. D. M. White, Astrophys. J. **490**, 493 (1997).

[68] J. F. Navarro, C. S. Frenk, and S. D. M. White, Astrophys. J. **462**, 563 (1996).

[69] J. S. Bullock, T. S. Kolatt, Y. Sigad, R. S. Somerville, A. V. Kravtsov, A. A. Klypin, J. R. Primack, and A. Dekel, Mon. Not. R. Astron. Soc. **321**, 559 (2001).

[70] A. D. Ludlow, J. F. Navarro, R. E. Angulo, M. Boylan-Kolchin, V. Springel, C. Frenk, and S. D. M. White, Mon. Not. R. Astron. Soc. **441**, 378 (2014).

[71] A. D. Ludlow, S. Bose, R. E. Angulo, L. Wang, W. A. Hellwing, J. F. Navarro, S. Cole, and C. S. Frenk, Mon. Not. R. Astron. Soc. **460**, 1214 (2016).

[72] C. Lacey and S. Cole, Mon. Not. R. Astron. Soc. **262**, 627 (1993).

[73] S. Bohr, J. Zavala, F.-Y. Cyr-Racine, and M. Vogelsberger, Mon. Not. R. Astron. Soc. **506**, 128 (2021).

[74] S. Kazantzidis, A. R. Zentner, and A. V. Kravtsov, Astrophys. J. **641**, 647 (2006).

[75] M. Boylan-Kolchin and C.-P. Ma, Mon. Not. R. Astron. Soc. **349**, 1117 (2004).

[76] K. Ahn and P. R. Shapiro, Mon. Not. R. Astron. Soc. **363**, 1092 (2005).

[77] O. Fakhouri, C.-P. Ma, and M. Boylan-Kolchin, Mon. Not. R. Astron. Soc. **406**, 2267 (2010).

[78] J. L. Feng, M. Kaplinghat, and H.-B. Yu, Phys. Rev. Lett. **104**, 151301 (2010).

[79] E. L. Łokas and G. A. Mamon, Mon. Not. R. Astron. Soc. **321**, 155 (2001).

[80] M. Petač, P. Ullio, and M. Valli, J. Cosmol. Astropart. Phys. 12 (2018) 039.

[81] M. Vogelsberger and J. Zavala, Mon. Not. R. Astron. Soc. **430**, 1722 (2013).

[82] N. J. Outmezguine, K. K. Boddy, S. Gad-Nasr, M. Kaplinghat, and L. Sagunski, arXiv:2204.06568.

[83] R. Essig, S. D. McDermott, H.-B. Yu, and Y.-M. Zhong, Phys. Rev. Lett. **123**, 121102 (2019).

[84] H. Xiao, X. Shen, P. F. Hopkins, and K. M. Zurek, J. Cosmol. Astropart. Phys. 07 (2021) 039.

[85] J. Choquette, J. M. Cline, and J. M. Cornell, J. Cosmol. Astropart. Phys. 07 (2019) 036.

[86] W.-X. Feng, H.-B. Yu, and Y.-M. Zhong, Astrophys. J. Lett. **914**, L26 (2021).

[87] W.-X. Feng, H.-B. Yu, and Y.-M. Zhong, J. Cosmol. Astropart. Phys. 05 (2022) 036.

[88] M. Rocha, A. H. G. Peter, and J. Bullock, Mon. Not. R. Astron. Soc. **425**, 231 (2012).

[89] C. A. Correa, Mon. Not. R. Astron. Soc. **503**, 920 (2021).

[90] J. Peñarrubia, A. J. Benson, M. G. Walker, G. Gilmore, A. W. McConnachie, and L. Mayer, Mon. Not. R. Astron. Soc. **406**, 1290 (2010).

[91] L. Jiang, S. Cole, T. Sawala, and C. S. Frenk, Mon. Not. R. Astron. Soc. **448**, 1674 (2015).

[92] M. A. Sánchez-Conde and F. Prada, Mon. Not. R. Astron. Soc. **442**, 2271 (2014).

[93] Z. C. Zeng, A. H. G. Peter, X. Du, A. Benson, S. Kim, F. Jiang, F.-Y. Cyr-Racine, and M. Vogelsberger, Mon. Not. R. Astron. Soc. **513**, 4845 (2022).

[94] A. Marasco, G. Cresci, L. Posti, F. Fraternali, F. Mannucci, A. Marconi, F. Belfiore, and S. M. Fall, Mon. Not. R. Astron. Soc. **507**, 4274 (2021).

[95] M. J. Bustamante-Rosell, E. Noyola, K. Gebhardt, M. H. Fabricius, X. Mazzalay, J. Thomas, and G. Zeimann, Astrophys. J. **921**, 107 (2021).

[96] J. R. Jardel and K. Gebhardt, Astrophys. J. **746**, 89 (2012).

[97] V. Lora, F. J. Sánchez-Salcedo, A. C. Raga, and A. Esquivel, Astrophys. J. Lett. **699**, L113 (2009).

[98] T. Sawala, C. S. Frenk, A. Fattahi, J. F. Navarro, T. Theuns, R. G. Bower, R. A. Crain, M. Furlong, A. Jenkins, M. Schaller *et al.*, Mon. Not. R. Astron. Soc. **456**, 85 (2016).

[99] J. Zavala and C. S. Frenk, Galaxies **7**, 81 (2019).

[100] A. F. Neto, L. Gao, P. Bett, S. Cole, J. F. Navarro, C. S. Frenk, S. D. M. White, V. Springel, and A. Jenkins, Mon. Not. R. Astron. Soc. **381**, 1450 (2007).

[101] A. Kamada, H. J. Kim, and T. Kuwahara, J. High Energy Phys. 12 (2020) 202.

[102] D. Yang, E. O. Nadler, and H.-b. Yu, arXiv:2211.13768.

[103] J. E. Greene, J. Strader, and L. C. Ho, Annu. Rev. Astron. Astrophys. **58**, 257 (2020).

[104] M. MacLeod, J. Guillochon, E. Ramirez-Ruiz, D. Kasen, and S. Rosswog, Astrophys. J. **819**, 3 (2016).

[105] O. D. Elbert, J. S. Bullock, S. Garrison-Kimmel, M. Rocha, J. Oñorbe, and A. H. G. Peter, Mon. Not. R. Astron. Soc. **453**, 29 (2015).