A semi-analytic study of self-interacting dark-matter haloes with baryons

Fangzhou Jiang [1,2]★ Andrew Benson [1] Philip F. Hopkins [2] Oren Slone,[3,4] Mariangela Lisanti,[3,5] Manoj Kaplinghat,[6] Annika H. G. Peter [7,8,9] Zhichao Carton Zeng,[7,8] Xiaolong Du,[1] Shengqi Yang[1] and Xuejian Shen [2]

[1]*Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA*
[2]*TAPIR, California Institute of Technology, Pasadena, CA 91125, USA*
[3]*Department of Physics, Princeton University, Princeton, NJ 08544, USA*
[4]*Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA*
[5]*Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Ave, New York, NY 10010, USA*
[6]*University of Irvine, Irvine, CA 92697, USA*
[7]*Department of Physics, The Ohio State University, 191 W. Woodruff Ave., Columbus, OH 43210, USA*
[8]*Center for Cosmology and Astroparticle Physics, The Ohio State University, 191 W. Woodruff Ave., Columbus, OH 43210, USA*
[9]*Department of Astronomy, The Ohio State University, 140 W. 18th Ave., Columbus, OH 43210, USA*

ABSTRACT

We combine the isothermal Jeans model and the model of adiabatic halo contraction into a semi-analytic procedure for computing the density profile of self-interacting dark-matter (SIDM) haloes with the gravitational influence from the inhabitant galaxies. The model agrees well with cosmological SIDM simulations over the entire core-forming stage up to the onset of gravothermal core-collapse. Using this model, we show that the halo response to baryons is more diverse in SIDM than in CDM and depends sensitively on galaxy size, a desirable feature in the context of the structural diversity of bright dwarfs. The fast speed of the method facilitates analyses that would be challenging for numerical simulations – notably, we quantify the SIDM halo response as functions of the baryonic properties, on a fine mesh grid spanned by the baryon-to-total-mass ratio, M_b/M_{vir}, and galaxy compactness, $r_{1/2}/R_{vir}$; we show with high statistical precision that for typical Milky-Way-like systems, the SIDM profiles are similar to their CDM counterparts; and we delineate the regime of core-collapse in the $M_b/M_{vir} - r_{1/2}/R_{vir}$ space, for a given cross section and concentration. Finally, we compare the isothermal Jeans model with the more sophisticated gravothermal fluid model, and show that the former yields faster core formation and agrees better with cosmological simulations. We attribute the difference to whether the target CDM halo is used as a boundary condition or as the initial condition for the gravothermal evolution, and thus comment on possible improvements of the fluid model. We have made our model publicly available at https://github.com/JiangFangzhou/SIDM.

Key words: cosmology: dark matter – galaxies: dwarf – galaxies: evolution – galaxies: haloes – galaxies: structure.

1 INTRODUCTION

Self-interacting dark matter (SIDM) provides appealing revisions on small scales to the standard Λ + Cold Dark Matter (CDM) paradigm of cosmic structure formation. Elastic self-interactions of dark-matter particles transfer heat towards the central regions of dark-matter haloes, creating constant density isothermal cores (e.g. Kochanek & White 2000; Colin et al. 2002; Vogelsberger, Zavala & Loeb 2012; Peter et al. 2013; Rocha et al. 2013). This is a convenient way of explaining the dark-matter cores in some dwarf galaxies (e.g. Blok et al. 2008; Oh et al. 2015), without breaking the large-scale success of the standard cosmology.

Galaxy formation complicates this picture. Hydro-cosmological SIDM simulations, as well as idealized SIDM-only simulations with analytical disc potentials, have shown that the dark-matter density profiles can sometimes be equally cuspy or cuspier than their CDM

counterparts (e.g. Elbert et al. 2018; Sameie et al. 2021). This implies that the response of SIDM haloes to the inhabitant galaxies is diverse and highly sensitive to certain baryonic details. The sensitivity of the SIDM halo response to baryonic details could be advantageous for explaining the small scale puzzles (e.g. Creasey et al. 2017; Kamada et al. 2017; Ren et al. 2019; Kaplinghat, Ren & Yu 2020; Zentner et al. 2022). In fact, there is now compelling observational evidence that the structures of bright dwarf galaxies are diverse, not only in terms of the central dark-matter density slope (e.g. Relatores et al. 2019; Shi et al. 2021) but also straightforwardly in terms of the galaxy size, which ranges from ∼0.5 kpc for compact ellipticals (e.g. Chilingarian & Zolotukhin 2015) all the way to ∼5 kpc for ultra-diffuse galaxies (e.g. Koda et al. 2015). These two aspects of structural diversity may actually be highly correlated, at least in ΛCDM. For example, simulated ultra-diffuse galaxies tend to be hosted by cored dark-matter haloes (e.g. Jiang et al. 2019), where supernovae-driven gas outflows puff up simultaneously the galaxies and the host haloes.

It is therefore interesting to revisit the correlation between galaxy size and host halo structure in the context of SIDM. Can

★ E-mail: fangzhou.jiang@gmail.com

we quantify the halo response to baryons in simple terms? Is it stronger or weaker than that in CDM? Which baryonic process is the most important for establishing the galaxy-SIDM-halo relation? To answer these questions, hydro-cosmological SIDM simulations have been developed; however, they must find a balance between sample size and numerical resolution: zoom-in hydro-cosmological SIDM simulations have so far been limited to a small sample of Milky-Way-like systems and dwarfs (e.g. Cruz et al. 2021; Sameie et al. 2021; Shen et al. 2021), whereas large-box SIDM simulations (e.g. Robertson et al. 2019) which contain large statistical samples still lack the resolution for reliably resolving the innermost few kpc. In this work, we adopt a semi-analytic approach based on the isothermal Jeans model first introduced in Kaplinghat et al. (2014); Kaplinghat, Tulin & Yu (2016). This model solves the Jeans-Poisson equation for the profile of the SIDM isothermal core, given the dark-matter density and velocity dispersion at the centre as well as the baryonic distribution. A recent adaptation of this method has been shown to be remarkably accurate compared to large-box SIDM simulations (Robertson et al. 2021). We improve this model by adding a prescription for adiabatic halo contraction (Gnedin et al. 2004), thus making it more self-consistent in describing the baryonic effect.

This integrated model takes a target CDM halo and baryonic potential as inputs. It computes the contracted CDM halo given the baryonic potential, and stitches an isothermal SIDM core to the CDM-like outskirt by minimizing their differences at the transition radius within which collisions are frequent. As such, this model can quickly compute density profiles for SIDM haloes with inhabitant galaxies, and, as we show below, produce results that are remarkably similar to those from *zoom-in* hydro-cosmological simulations. The speed of this semi-analytic approach enables investigations of SIDM halo response with high statistical precision and with long baselines of input parameters such as baryonic size and mass.

This paper is organized as follows. In Section 2, we recap the model ingredients and combine them into a workflow, summarized in Section 2.4. In Section 3, we compare the model predictions to the results from zoom-in cosmological SIDM simulations, including both dark-matter-only set-ups and hydro-simulations. After demonstrating the accuracy of the model, we use it to study the halo response in Section 4, where we quantitatively relate the inner structure of the SIDM haloes to the compactness and mass fraction of the inhabitant galaxies, and show the importance of considering adiabatic halo contraction. Finally, in Section 5, we compare this model to the other one-dimensional method for SIDM haloes that is extensively studied in the literature – the gravothermal fluid model (Section 5.1), and we also study the facilitation of gravothermal core-collapse by the inhabitant galaxy, providing regions of core-collapse in the space spanned by galaxy mass fraction and galaxy compactness, as a function of the cross section and target halo concentration. For general readers who want to skip the technical details and get to the results sooner, Section 2.4 can be a good starting point.

Throughout, we define the virial radius of a distinct halo as the radius within which the average density is $\Delta = 200$ times the critical density for closure. We also assume spherical symmetry for both the dark-matter haloes and galaxies. We adopt a flat cosmology with the present-day matter density $\Omega_m = 0.3$, baryonic density $\Omega_b = 0.0465$, dark energy density $\Omega_\Lambda = 0.7$, a power spectrum normalization $\sigma_8 = 0.8$, a power-law spectral index of $n_s = 1$, and a Hubble parameter of $h = 0.7$, unless otherwise mentioned.

2 ANALYTIC METHOD FOR COMPUTING THE DENSITY PROFILE OF SIDM HALOES

Scattering between dark-matter particles is prevalent in the centre of a halo where the dark-matter density is high, but is infrequent on the outskirts where the scattering time-scale is longer than the lifetime of the halo. The full profile of an SIDM halo therefore consists of a thermalized core and a CDM-like outer region. The transition is around a characteristic radius r_1, within which an average dark-matter particle has experienced more than one scattering over the lifetime t_{age} of the halo (Kaplinghat et al. 2016):

$$\frac{4}{\sqrt{\pi}} \rho_{dm}(r_1) v(r_1) \sigma_m = \frac{1}{t_{age}}, \tag{1}$$

where the left-hand side is the scattering rate per particle, with ρ_{dm} the DM density, $(4/\sqrt{\pi})v$ the average relative velocity between DM particles for a Maxwellian distribution (where v is the 1D velocity dispersion), and σ_m the self-interaction cross-section per particle mass. Note that, the cross section also carries a radius dependence if it is velocity dependent, which comes in via the velocity dispersion profile, i.e. $\sigma_m = \sigma_m[v(r)]$. Here, we assume constant cross section in the velocity-dispersion regime of interest. This assumption holds when the halo develops its isothermal core within r_1.

The impact of DM self-interactions on the halo density profile can be regarded as a modification to the inner part ($r < r_1$) of a CDM counterpart, and can be computed using the spherical Jeans equation with the assumption that the halo is isothermal within r_1 and in approximate equilibrium.

2.1 Profile of the isothermal core

The density profile of the isothermal dark-matter core can be solved by combining the spherical Jeans equation and the Poisson equation:

$$\frac{d(\rho_{dm}v^2)}{dr} + \frac{2\beta}{r}v^2 = -\rho_{dm}\frac{d\Phi}{dr}, \tag{2}$$

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\Phi}{dr}\right) = 4\pi G\rho = 4\pi G(\rho_{dm} + \rho_b), \tag{3}$$

where Φ is the total gravitational potential, ρ is the total density, and ρ_b is the baryon density. With the assumption of an isotropic ($\beta = 0$) and constant 1D velocity dispersion ($v(r) = v_0$), the Jeans equation has a simple generic solution:

$$\rho_{dm}(r) = \rho_{dm0} \exp\left[-\frac{\Delta\Phi(r)}{v_0^2}\right] \quad \text{or} \quad \Delta\Phi(r) = -v_0^2 \ln\left[\frac{\rho_{dm}(r)}{\rho_{dm0}}\right], \tag{4}$$

where ρ_{dm0} is the central dark-matter density, and $\Delta\Phi(r) = \Phi(r) - \Phi(0)$ is the potential difference between radius r and the centre. Combining equation (3) and (4), we get

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\ln\rho_{dm}(r)}{dr}\right) = -\frac{4\pi G}{v_0^2}[\rho_{dm}(r) + \rho_b(r)]. \tag{5}$$

Following Kaplinghat et al. (2014), we assume a Hernquist profile for the baryon distribution,

$$\rho_b(r) = \frac{M_b/2\pi r_0^3}{\frac{r}{r_0}\left(1 + \frac{r}{r_0}\right)^3}, \tag{6}$$

where M_b is the baryon mass, and r_0 the scale radius. Then, equation (5) can be rewritten as the dimensionless form

$$\frac{d^2h}{dy^2} + \frac{2}{y}\frac{dh}{dy} + \frac{b}{y} + \frac{ae^h}{(1-y)^4} = 0, \tag{7}$$

where $h(y) \equiv \Phi(y)/v_0^2$, $y = (r/r_0)/(1 + r/r_0)$, $a \equiv 4\pi G r_0^2 \rho_{dm0}/v_0^2$, and $b \equiv 2GM_b/r_0 v_0^2$. The boundary conditions for solving this equation are $h(0) = 0$ and $h'(0) = -b/2$. The isothermal core profile can therefore be obtained by integrating equation (7), given the baryon properties (M_b, r_0), the central DM density (ρ_{dm0}), and the constant velocity dispersion within the core (v_0). Rewriting equation (5) into the dimensionless form of equation (7) facilitates the numerical solution. One can assume other profile shapes for the baryon density in equation (5) and derive dimensionless equations similar to equation (7). Here, we opt for the Hernquist profile because it is generally applicable to a wide range of galaxies including dwarfs and massive ellipticals, and more importantly, because when modelling the halo contraction (Section 2.2), we have taken advantage of a fast analytical formalism (Gnedin et al. 2004) that has been optimized for Hernquist profiles.

There are four parameters in total that fully determine the isothermal dark-matter profile: two for baryons (M_b, r_0) and two for dark matter (ρ_{dm0}, v_0). For modellers, the baryonic parameters (M_b, r_0) are usually known – for constructing simple toy halo models based on observations, (M_b, r_0) are available from surface photometry; for building more complex semi-analytic or semi-empirical frameworks, they can be set from empirical abundance-matching relations. However, the DM parameters (ρ_{dm0}, v_0) are not readily known. They need to be determined iteratively given the virial mass M_{vir} and concentration c of the target CDM halo, as we will describe in Section 2.3.

We emphasize that, the isothermal Jeans model assumes that the system is in approximate equilibrium. Strictly speaking, an SIDM halo is never in Jeans equilibrium, but constantly evolving by transporting energy from the dynamically hotter region to colder places. For a target system that is initially described by a CDM profile, the dynamically hottest place is where the $v(r)$ profile peaks, so with self-interactions, the heat flows to the centre. As the system evolves, the core temperature gradually becomes the highest and then conducts energy outwards. The full time evolution can be described using the gravothermal fluid equations (see Section 5.1).

2.2 Halo contraction

The dark-matter distribution contracts in response to the condensation of baryons in the halo centre. Blumenthal et al. (1986) described this process assuming circular orbits and an adiabatic invariant of $M(r)r$, where $M(r)$ is the total mass enclosed within radius r. Gnedin et al. (2004) showed that the original adiabatic-contraction treatment overestimates the magnitude of contraction compared to the results of cosmological hydro-simulations, and attributed the mismatch to the oversimplified assumption of circular orbits. To account for orbital eccentricity and orbital phase distributions, they proposed a modified invariant, $M(\bar{r})r$, where $\bar{r}$ is the orbit-averaged radius for particles at instantaneous radius r, approximated by

$$\bar{x} = Ax^w, \tag{8}$$

where $x = r/R_{vir}$, and the parameters $A \approx 0.85$ and $w \approx 0.8$ are calibrated with simulations. There is some halo-to-halo variation in these parameters (Gnedin et al. 2011), which we ignore in this work.[1] With $M(\bar{r})r$ invariant and assuming that the baryons are initially distributed with the same radial profile as the dark matter,

---

[1] We ignore the halo-to-halo variation because there seems to be no systematic trend of w or A with halo mass or concentration. w is weakly dependent on the details of cooling, but usually within 0.6–1.0.

one can show that the final radius r_f of dark-matter particles initially located at $r > r_f$ obeys the equation:

$$\frac{r}{r_f} = 1 - f_b + \frac{M_b(\bar{r}_f)}{M_i(\bar{r})}, \tag{9}$$

where $f_b = M_b/M_{vir}$ is the galactic mass fraction within R_{vir}, $M_b(r)$ is the final baryon mass within r, and $M_i(r)$ is the initial total mass profile.

Assuming that the initial distribution of DM and baryons both follow an NFW profile (Navarro, Frenk & White 1997),

$$\rho(r) = \frac{\rho_s}{cx(1 + cx)^2}, \quad \text{where} \quad \rho_s = \frac{c^3}{3f(c)}\Delta\rho_{crit}, \tag{10}$$

with Δ the average overdensity with respect to the critical density of the Universe $\rho_{crit}(z)$, c the concentration parameter, and $f(c) = \ln(1 + c) - c/(1 + c)$, and that the final baryonic distribution obeys a Hernquist profile, then a solution of equation (9) can be obtained. The details of this step can be found in the appendix of Gnedin et al. (2004). Solving equation (9) for r_f for an initial radius r, we get the enclosed mass profile $M_{dm, f}(r_f) = (1 - f_b)M_i(r)$ of the contracted halo.

The contracted DM mass profile is non-parametric. To facilitate subsequent modelling, such as solving for the characteristic radius r_1, we need simple parametric expressions for the density profile $\rho_{dm}(r)$ and the velocity–dispersion profile $v(r)$. We therefore fit the profile of a contracted halo with the Dekel-Zhao (DZ) profile (Freundlich et al. 2020), which has analytic expressions for $\rho_{dm}(r)$ and $v(r)$, and is flexible enough in the centre to account for the contraction, at the expense of adding just one more degree of freedom than NFW. The enclosed mass of a DZ profile is given by

$$M_{dm}(r) = (1 - f_b)M_{vir}\frac{g(cx, \alpha)}{g(c, \alpha)}, \tag{11}$$

where $g(\xi, \alpha) = [\xi^{1/2}/(1 + \xi^{1/2})]^{2(3-\alpha)}$; and c and α are the free parameters describing the concentration and innermost density slope of the halo. The density profile and the velocity dispersion profile are given by

$$\rho_{dm}(r) = \frac{\rho_{DZ}}{x^\alpha(1 + x^{1/2})^{2(3.5-\alpha)}}, \tag{12}$$

$$v^2(r) = 2V_{vir}^2\frac{c}{g(c, \alpha)}\frac{x^{3.5}}{\chi^{2(3.5-\alpha)}}\sum_{i=0}^{8}\frac{(-1)^i 8!}{i!(8-i)!}\frac{1 - \chi^{4(1-\alpha)+i}}{4(1-\alpha)+i}, \tag{13}$$

where $\rho_{DZ} = [c^3(3 - \alpha)]/[3g(c, \alpha)] \times \Delta\rho_{crit}$, V_{vir} is the circular velocity at the virial radius, and $\chi = x^{1/2}/(1 + x^{1/2})$. We fit the mass profile $M_{dm, f}(r_f)$ of a contracted halo using equation (11) and then solve equation (1) for the transition radius r_1 using the density and velocity dispersion of the best-fitting DZ profile. For typical baryon distributions ($0.01 \lesssim f_b \lesssim 0.2$ and $0.005 \lesssim r_0/R_{vir} \lesssim 0.1$), the best-fitting DZ profile agrees with the non-parametric solution of $M_{dm, f}(r_f)$ to per-cent level.

From now on, we drop the 'dm' in the subscription of the symbol for central DM density ρ_{dm0} and simply denote it by ρ_0.

2.3 Stitching the isothermal core to the CDM outskirt

To obtain the full profile of an SIDM halo with baryons, we determine the parameters (ρ_0, v_0) of the isothermal core iteratively, such that the core joins the contracted CDM halo at radius r_1 smoothly in terms of the local density and the enclosed mass. Specifically, we search
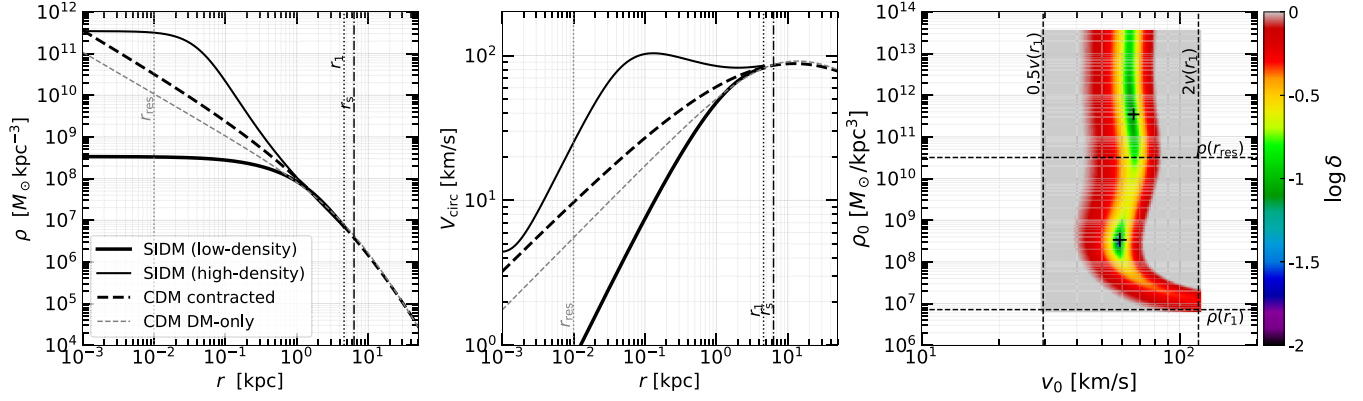
Figure 1. Illustration of the semi-analytical workflow (Section 2.4) – example of finding the SIDM profile for a cross-section of $\sigma_m = 1 \, \mathrm{cm^2 \, g^{-1}}$ and a target CDM halo formed $t_{\mathrm{age}} = 5 \, \mathrm{Gyr}$ ago with a present-day virial mass of $M_{\mathrm{vir}} = 10^{11} \, \mathrm{M_\odot}$ and a concentration of $c = 15$. The inhabitant galaxy has a total cold baryon mass of $M_{\mathrm{b}} = 10^9 \, \mathrm{M_\odot}$, and half-mass radius of $r_{1/2} = 1.9 \, \mathrm{kpc}$ ($r_{1/2}/R_{\mathrm{vir}} = 0.02$). The thin grey dashed lines show the density profile (*left*) and circular–velocity profile (*middle*) of the original dark-matter-only target CDM halo; and the thick dashed lines are the profiles for the adiabatically contracted CDM halo. The thicker solid black lines are the profiles of the best-matching SIDM isothermal core, which corresponds to the low-density solution of (ρ_0, v_0) as marked by the bigger black ' + ' sign in the right-hand panel. The thinner solid black lines are the profiles that correspond to the *discarded* high-density solution, as marked by the smaller ' + ' sign in the right-hand panel. The right-hand panel shows the colour map of the 'stitching error' δ, defined in equation (14), in the space of central density ρ_0 versus central velocity dispersion v_0. Clearly there are two δ minima, but only the low-density solution agrees with simulation results (see Fig. 2). As shown in Appendix A, the two solutions get closer as the system evolves (i.e. as t_{age} or cross-section increases). When they join, gravothermal core-collapse starts to speed up (see Section 5.2). The vertical and horizontal dashed lines indicate the region $\rho_{\mathrm{cdm}}(r_1) < \rho_0 < \rho_{\mathrm{cdm}}(r_{\mathrm{res}} = 10 \, \mathrm{pc})$ and $0.5v(r_1) < v_0 < 2v(r_1)$, which brackets the low-density solution uniquely for a wide range of halo age and cross section.

the space of ρ_0–v_0 to minimize the following objective quantity:

$$\delta^2 = \left[\frac{\rho_{\mathrm{iso}}(r_1) - \rho_{\mathrm{cdm}}(r_1)}{\rho_{\mathrm{cdm}}(r_1)} \right]^2 + \left[\frac{M_{\mathrm{iso}}(r_1) - M_{\mathrm{cdm}}(r_1)}{M_{\mathrm{cdm}}(r_1)} \right]^2, \quad (14)$$

where $\rho_{\mathrm{iso}}(r)$ and $\rho_{\mathrm{cdm}}(r)$ are the density profiles, and $M_{\mathrm{iso}}(r)$ and $M_{\mathrm{cdm}}(r)$ are the enclosed DM mass profiles, of the isothermal core and the contracted CDM halo, respectively. There are two minima of δ^2 in the ρ_0–v_0 space, with similar v_0 values but very different ρ_0. The existence of the two solutions was already noted by Elbert et al. (2018). Here, we illustrate them clearly in the right-hand panel of Fig. 1.[2]

Elbert et al. (2015) only accepted the lower-density solution as it agrees with their simulation results better. We emphasize that both solutions are physical in the sense that they both meet the requirement of constant temperature below r_1. It is just that realistic haloes form with properties closer to the lower-density solution, which is why the lower-density solution agrees better with cosmological simulation results. We find by trial and error that a practical searching range for the lower density solution is $\rho_{\mathrm{cdm}}(r_1) < \rho_0 < \rho_{\mathrm{cdm}}(10 \, \mathrm{pc})$ and $0.5v(r_1) < v_0 < 2v(r_1)$, which, in most cases, brackets a unique minimum of δ^2.

As will be shown below, this simple formalism can capture the onset of gravothermal core-collapse. As the halo age t_{age} increases or as the cross section σ_m becomes larger, the two minima of δ^2 get closer – they first both decrease in ρ_0; then the lower density solution turns around, manifesting the onset of gravothermal core-collapse; and finally the two solutions merge as core-collapse speeds up, beyond which point the isothermal model is no longer

applicable. This is illustrated in Appendix A, and the high-density solution is therefore also useful, as we will address further in Section 5.2.

2.4 Workflow

We summarize the workflow for getting the density profile of an SIDM halo with baryons as follows:

(1) Given a CDM halo described by an NFW profile (i.e. with known virial mass M_{vir}, concentration c, and age t_{age}), and given an inhabitant galaxy described by a Hernquist profile (parametrized by the mass M_{b} and scale size r_0), compute the adiabatically contracted halo profile (Section 2.2).

(2) Given the self-interaction cross-section, σ_m, solve equation (1) for the radius of frequent scattering, r_1, using the density profile and velocity-dispersion profile of the contracted CDM halo.

(3) Integrate the spherical Jeans–Poisson equation, equation (5), to obtain an isothermal core profile (Section 2.1) – do this iteratively to find the central DM density ρ_0 and the central velocity dispersion v_0 by minimizing the relative stitching error defined in equation (14) (Section 2.3).

To illustrate, Fig. 1 shows an example of the density and circular velocity profiles of an SIDM halo obtained with this workflow. In this example, we adopt a self-interaction cross-section of $\sigma_m = 1 \, \mathrm{cm^2 \, g^{-1}}$ and a target CDM halo of $M_{\mathrm{vir}} = 10^{11} \, \mathrm{M_\odot}$, $c = 15$, and $t_{\mathrm{age}} = 10 \, \mathrm{Gyr}$ with a Hernquist baryon distribution of mass $M_{\mathrm{b}} = 10^9 \, \mathrm{M_\odot}$ and half-mass radius $r_{1/2} = 1.9 \, \mathrm{kpc}$ (i.e. a Hernquist $r_0 = r_{1/2}/(1 + \sqrt{2}) \approx 0.8 \, \mathrm{kpc}$). These choices are largely arbitrary for illustration purposes, but are of the same of order as the Large Magellanic Cloud (LMC). In Appendix A, we demonstrate how the two solutions evolve as the halo age increases, and discuss in Section 5.2 that the high-density solution can help us to phenomenologically predict the onset of gravothermal core-collapse. While this procedure is devised for

---

[2]For Figs 1 and A1, we evaluate the objective of stitching error on a regular grid of $\log \rho_{\mathrm{cdm}}$ and $\log v_0$ – this is relatively slow and only for the purpose of illustrating the two minima. For the rest of the work and in the public code that we share, the minimization was performed using the PYTHON module SPICY.OPTIMIZE.MINIMIZE and its default BFGS algorithm.
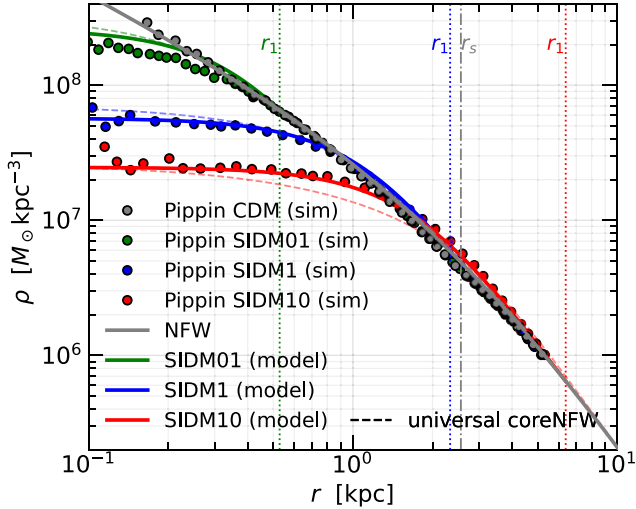
Figure 2. Comparison of the dark-matter density profiles from the model and from the cosmological N-body simulations of the Pippin haloes as in Elbert et al. (2015) at $z = 0$. The grey circles represent the reference-CDM simulation result; and the green, blue, and red circles represent the SIDM simulation results of cross-sections of $\sigma_m = 0.1, 1,$ and $10 \, \mathrm{cm^2 \, g^{-1}}$, respectively (labelled as SIDM01, SIDM1, and SIDM10). The CDM halo is well described by an NFW profile of $M_{\mathrm{vir}} = 10^{9.89} \, \mathrm{M_\odot}$ and $c = 15.8$, as indicated by the grey solid line – this is used as the starting point of the isothermal Jeans model. The model predictions are shown by the solid lines of corresponding colours. The vertical dotted lines show the r_1 radii. The model predictions agree very well with the simulation results across 2 dex in cross-section. The thin dashed lines in pale colours represent a *universal* approximation, which is the coreNFW profile with a scale radius r_c that is 0.45 times the respective r_1.

haloes with baryons, it is fully compatible with dark-matter-only cases, for which one simply sets M_b zero.

3 COMPARISON WITH COSMOLOGICAL SIDM SIMULATIONS

In this section, we show that the aforementioned workflow gives halo profiles closely matching those from cosmological SIDM simulations. We also provide a simple analytical fitting formula for the dark-matter-only cases.

3.1 Comparison with dark-matter only simulations

To compare the model to cosmological dark-matter-only simulations, we use the zoom-in simulations of Elbert et al. (2015) and focus on the 'Pippin' haloes therein. The simulations adopt the *Wilkinson Microwave Anisotropy Probe-7* cosmology (Komatsu et al. 2011), with $h = 0.71$, $\Omega_m = 0.266$, $\Omega_\Lambda = 0.734$, $n_s = 0.963$, and $\sigma_8 = 0.801$. For the high-resolution runs that we compare to, the particle mass is $1.5 \times 10^3 \, \mathrm{M_\odot}$, and the Plummer equivalent force softening length is 28 pc. The Pippin halo was run in both CDM and SIDM with a wide range of velocity-independent cross-sections of $\sigma_m = 0.1$–$50 \, \mathrm{cm^2 \, g^{-1}}$, all starting from the same initial conditions. The SIDM implementation follows that of Rocha et al. (2013). The CDM Pippin halo is accurately described by an NFW profile with a virial mass of $M_{\mathrm{vir}} = 10^{9.89} \, \mathrm{M_\odot}$ and a concentration of $c = 15.8$, as shown by the grey line in Fig. 2. We use this NFW profile as the input of the target CDM profile for our model, and compute the SIDM profiles for $\sigma_m = 0.1, 1,$ and $10 \, \mathrm{cm^2 \, g^{-1}}$, which are then compared to the corresponding simulation results. Since we are dealing with dark-

matter only cases, M_b is set to be infinitesimally small. We find that the model predictions agree well with the simulation results across the cross-section range.

While this semi-analytic procedure is already reasonably fast ($\lesssim 0.1$ sec per system using our publicly available PYTHON implementation), it still requires numerical minimization for determining ρ_0 and σ_0. To accommodate semi-analytic frameworks designed for large ensembles of haloes and subhaloes (e.g. Benson 2012; Jiang et al. 2021), an even faster formula would be useful. We find that a CORENFW profile (Read, Agertz & Collins 2016) with the scale radius being a fixed fraction of r_1 provides decent approximations. The CORENFW profile has an enclosed mass profile given by

$$M(r) = M_{\mathrm{NFW}}(r) \tanh\left(\frac{r}{r_c}\right), \tag{15}$$

where $M_{\mathrm{NFW}}(r)$ is the enclosed mass of the target NFW profile, and r_c is a characteristic core size. We find by trial and error that CORENFW profiles with $r_c = 0.45 r_1$ fit accurately the SIDM haloes derived from the same target CDM halo across 2 dex in cross-section, as shown by the thin dashed lines in Fig. 2. We have verified that this universal approximation holds as long as the system is not in the core-collapse regime, and thus applies to most SIDM haloes with $c \lesssim 20$, $t_{\mathrm{age}} \lesssim 14$ Gyr, and $\sigma_m \lesssim 10 \, \mathrm{cm^2/g}$. It breaks down when the baryonic component is not negligible, or when the halo starts to core-collapse, for which a more complicated profile shape is needed.

3.2 Comparison with hydro simulations

We also compare the model predictions to cosmological hydro simulations, to test its performance when the system is baryon dominated in the centre. We use three Milky-Way-mass systems in the FIRE-2 SIDM suite (Sameie et al. 2021): m12i, m12f, and m12m, which have virial masses of $M_{\mathrm{200m}} = 10^{11.95}$, $10^{12.15}$, and $10^{12.08} \, \mathrm{M_\odot}$, respectively, at $z = 0$. These galaxies are simulated with cross-sections of $\sigma_m = 1$ and $10 \, \mathrm{cm^2 \, g}$, and they all have CDM-only reference runs with matched initial conditions which we can use for the model inputs. Among the three systems, m12i and m12f have Milky-Way-like sizes of $r_{1/2} \approx 4$ kpc and a stellar mass of $M_b \sim 10^{10.7} \, \mathrm{M_\odot}$, while m12m has a slightly higher stellar mass of $M_b \sim 10^{10.9} \, \mathrm{M_\odot}$ and a much more extended stellar distribution of $r_{1/2} \approx 8$ kpc. Table 1 of Sameie et al. (2021) provides more detailed information of these simulations.

Again, following the workflow in Section 2.4, we fit NFW profiles to the CDM-only simulations at $z = 0$ and treat the best-fitting profiles as the target haloes, as shown by the grey lines in Fig. 3. Then we fit Hernquist profiles to their stellar distributions, as represented by the coloured dotted lines in Fig. 3, and use them to model the adiabatic contraction of these haloes. Both the NFW fits and the Hernquist fits are accurate within ~ 5 per cent in the radius range of interest. We assume these systems formed $t_{\mathrm{age}} = 7$ Gyr ago, which is the average formation time of haloes of Milky-Way mass scale. The predicted SIDM profiles, as shown by the coloured solid lines in Fig. 3, match the simulation results fairly accurately. For the SIDM1 runs, the central densities are matched at per cent levels. For the SIDM10 runs, while the model slightly overpredicts the central densities, it still correctly captures the shape of the simulated density profiles: there is a relatively flat central core at $r \lesssim 1$ kpc, a steep decrease at $r \sim 5$ kpc, and a flatter part again at $r \sim r_1 \sim 40$ kpc.

The good agreement between the model and the simulations provides insights into the galaxy–halo connection in the context of SIDM. In CDM, there are two *equally important* competing baryonic effects on halo structure – on the one hand, the galactic potential

Table 1. Comparison of the two 1D models of self-interacting DM haloes – the isothermal Jeans model versus the gravothermal fluid model. See Section 5.1 for details.

	Isothermal	Gravothermal
similarities		
Operation target	CDM halo	CDM halo
Applicable before core-collapse	yes	yes
differences		
Speed	fast	slow
Applicable after core-collapse	no	yes
Captures baryonic effect	yes[a]	no[b]
Support v-dependent σ_m	no[c]	yes

[a]It captures the gravitational effect of the baryonic potential, not the baryonic feedback.

[b]In principle, one can add a static baryonic term in the second equation of equation (16), such that the gravothermal fluid model can also capture the halo response to the baryonic component.

[c]However, velocity dependence effectively makes the cross-section larger in the past, so if given the growth history of the target CDM halo, we can redefine r_1 with equation (17) and perform the isothermal Jeans modeling for each time.

makes the halo contract and become more cuspy; on the other hand, supernovae-driven outflows heat the potential well and flatten the central density. The net effect of the competing mechanisms depend sensitively on details of the subgrid physics for star formation and supernovae (e.g. Bose et al. 2019). The SIDM simulations here also include both of the competing mechanisms, but the model only considers halo contraction and ignores stellar feedback. Hence, the fact that good agreement is still achieved between the model and the FIRE2-SIDM simulations implies that the core-formation effect from supernovae is subdominant and overwhelmed by the effect of the SIDM halo in the presence of the baryonic potential (see also Sameie et al. 2021 for discussion). It is therefore reasonable to speculate that SIDM simulations are not sensitive to the sub-grid baryonic physics for certain ranges of SIDM parameters. This should be better tested with hydro + SIDM simulations with varied strength of feedback.

4 SIDM HALO RESPONSE

In this section, we use the model for quantitative analysis of the SIDM halo response. We express the halo structures as functions of the baryonic mass fraction (M_b/M_{vir}) and the baryonic compactness ($r_{1/2}/R_{vir}$), and also take this opportunity to show the importance of considering adiabatic halo contraction.

4.1 Enhanced structural diversity in SIDM

Zoom-in hydro-simulations have hinted that SIDM haloes are more responsive to the presence of a baryonic distribution (rather than baryonic feedback) than their CDM counterparts. Here, we use the isothermal Jeans model to show this more explicitly.

First, we vary the size of the baryonic component while keeping the total mass and baryon mass fixed at $M_{vir} = 10^{11}\,M_\odot$ and $M_b = 10^9\,M_\odot$ – these values are typical of bright dwarf galaxies such as the LMC or sub-$L^\star$ galaxies which exhibit the most dramatic structural diversity. We also keep the halo age and the target-halo's concentration fixed at typical values of $t_{age} = 10$ Gyr and $c = 10$. We run the model for two cross sections, $\sigma_m = 1$ and $0.1\,\mathrm{cm^2\,g^{-1}}$. We perform control-experiments to get the CDM references, i.e. starting

from the same target halo and the same galaxy as used for the SIDM calculations, and simply compute the adiabatically contracted CDM halo profiles. Fig. 4 shows the comparison. The sensitivity of the halo response in the SIDM models is indeed much higher than that of the reference CDM cases. Notably, the inner SIDM density slope (evaluated at, e.g. $r \sim 0.5$ kpc ≈ 0.5 per cent R_{vir}) can be flat, equally cuspy, or cuspier than that of the reference CDM profile, depending on whether the galaxy is diffuse ($r_{1/2} = 2$ kpc), normal (1 kpc), or compact ($r_{1/2} \simeq 0.5$ kpc). The range of the central densities, e.g. evaluated at $r = 0.1$ kpc, of the CDM results is only 0.5 dex, while that of the SIDM models spans more than an order of magnitude.

This remarkable diversity in halo response is not driven by the difference in the characteristic radius r_1. In fact, for $\sigma_m \gtrsim 1\,\mathrm{cm^2\,g^{-1}}$, the r_1 values are similar across the different galaxy sizes, as shown by the vertical dotted lines in Fig. 4. Only for cross sections as small as $\sigma_m \sim 0.1\,\mathrm{cm^2\,g^{-1}}$, r_1 becomes comparable to the galaxy size and differs significantly depending on the latter. Even here, r_1 occurs where the halo density profiles converge, so the dramatic difference in the inner halo cannot be attributed to that of r_1 or of the local density $\rho(r_1)$. The structural diversity must then arise from the difference in the enclosed mass profile, or $V_{circ}(r)$, as shown in the right-hand panels of Fig. 4. A small change in the baryonic size results in amplified differences in the gradient and the Laplacian of the potential, $d\Phi/dr = V_{circ}(r)^2/r$ and $d^2\Phi/dr^2$, which are leading terms in the Jeans–Poisson equation (equation 7) underlying the whole model.

The structural diversity of bright dwarf galaxies ($M_b \approx 10^{8-9}\,M_\odot$) has drawn a lot of attention recently. Notably, these galaxies span two orders of magnitude in size and exhibit a wide range of morphologies, including compact dwarfs with $r_{1/2}$ as small as ~ 0.1 kpc and ultra-diffuse galaxies with $r_{1/2}$ up to 10 kpc. The structural diversity is also manifested in the logarithmic density slope $s \equiv \mathrm{d}\ln \rho/\mathrm{d}\ln r$ near the centre ($r \lesssim 1$ kpc), as inferred from baryonic kinematics. For example, as Relatores et al. (2019) summarized, s ranges between 0 and 1.5 for galaxies with $M_b \sim 10^9\,M_\odot$. It is challenging for hydro + CDM models to fully explain such a dramatic extent of structural diversity, especially given that both the galaxy size and the inner halo structure exhibit wide ranges. Recently, Zentner et al. (2022) demonstrated that SIDM and feedback-affected CDM models are equally better than a CDM model in explaining the halo structural diversity as seen in the SPARC survey (Lelli, McGaugh & Schombert 2016), however, the prevalence of compact bright dwarfs with $r_{1/2} \lesssim 1$ kpc remains a challenge for hydro-CDM simulations featuring strong feedback (e.g. Jiang et al. 2019). Here, galaxy size is an input of the model, so we do not provide an explanation for the size diversity, but we have clearly shown that *SIDM* models have the virtue of making the two aspects strongly coupled, such that if there is an explanation for the *size diversity*, it explains automatically the range of *DM density slopes*.

Second, we extend the above exercise by scanning a wide range in the space spanned by the baryonic mass fraction and galaxy compactness, and thus more systematically describe the SIDM halo response. Still adopting $\sigma_m = 1\,\mathrm{cm^2\,g^{-1}}$ and a target CDM halo of $t_{age} = 10$ Gyr, $M_{vir} = 10^{11}\,M_\odot$, and $c = 10$, we vary M_b/M_{vir} from 10^{-4} to 0.1, and $r_{1/2}/R_{vir}$ from 0.004 to 0.1. We express the halo structure in terms of the inner density slope $s \equiv -\mathrm{d}\ln \rho/\mathrm{d}\ln r$ evaluated at $r = 1$ kpc, and the central density ρ_0 in units of the NFW scale density ρ_s. The results are shown in Fig. 5. The main panels of Fig. 5 show the contour maps of s and ρ_0/ρ_s in the 2D baryon-property space. The top panels and side panels show the 1D slices of the 2D map with either of the baryonic quantities
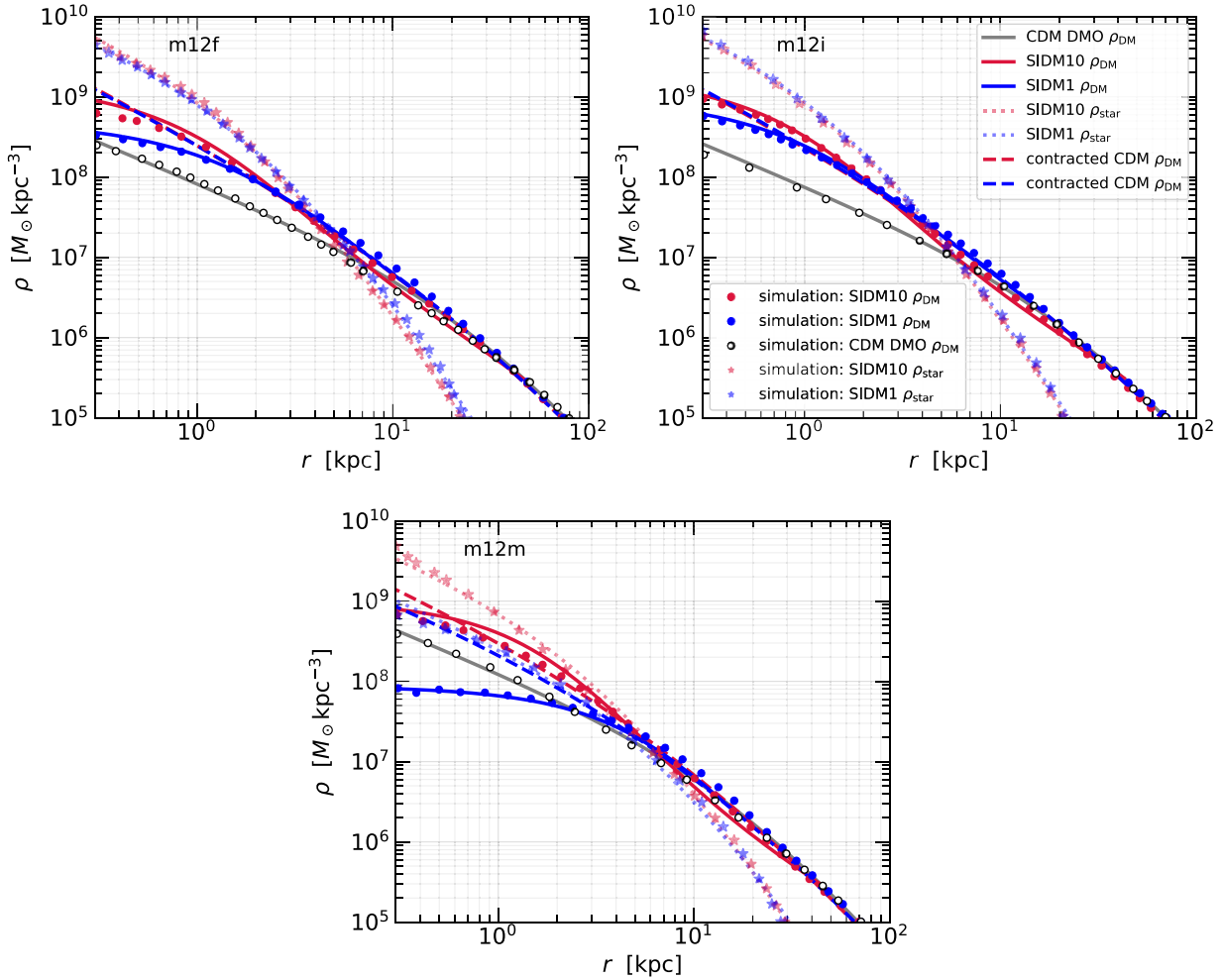
Figure 3. Comparison of the dark-matter density profiles from the model and from the FIRE2-SIDM hydro-cosmological simulations – showing three examples of Milky-Way-mass systems at $z = 0$: m12f, m12i, and m12m, as in Sameie et al. (2021). The open circles represent the density profiles in the reference CDM dark-matter-only (DMO) runs. The solid grey lines show the best-fitting NFW profiles, which are used as inputs in the isothermal Jeans model. The dotted lines represent the best-fitting Hernquist profiles of the stellar density distributions (stars) in the hydro-simulations. The red and blue colours differentiate the SIDM results for $\sigma_m = 10$ (SIDM10) and $1\,\mathrm{cm}^2\,\mathrm{g}^{-1}$ (SIDM1), respectively. The stellar profiles are used as inputs to the model for computing halo contraction. The dashed lines of corresponding colours represent the profiles of the contracted CDM haloes. The filled circles and solid lines of the matching colour represent the profiles from the SIDM simulations and the corresponding models. Overall, the model preditions are in decent agreement with the simulations – for the SIDM1 run, the central densities at $r \sim 1$ kpc agree at per cent level; for SIDM10, the shapes of the simulated profiles are correctly reproduced.

fixed (at $M_b/M_{\rm vir} = 0.01$ or $r_{1/2}/R_{\rm vir} = 0.02$). Clearly, the SIDM halo becomes more dense and cuspy as the galaxy becomes more massive and compact. Although we have used a massive dwarf halo for illustration, the result applies to other mass scales as well since we have expressed the baryonic properties in units of the virial quantities.

Hydro-cosmological zoom-in simulations have shown that, for Milky-Way-like systems, SIDM halo profiles are rather similar to their CDM counterparts down to quite small radii. This can be seen for example in m12f and m12i in Fig. 3, and it has motivated some semi-analytic studies to assume NFW profiles for their Milky-Way sized SIDM host halo when studying the satellite galaxies (e.g. Slone et al. 2022). Here, we can easily check the validity of this assumption in Fig. 5. Abundance-matching studies have shown that a Milky-Way-mass system typically has a stellar-to-total-mass ratio of $\sim$1 per cent (e.g. Moster, Naab & White 2013), and a half mass radius that is $\sim$2 per cent of the host-halo virial radius (e.g. Somerville et al. 2018). For these representative values, as can be seen in Fig. 5, the

SIDM profile indeed has an inner logarithmic density slope very close to the NFW value of $s \sim 1$.

4.2 Necessity of considering adiabatic contraction

Robertson et al. (2021) also studied the isothermal Jeans model in detail and made comparisons with cosmological simulations. There, the authors adopted an *inside-out* fitting scheme. That is, different from what we do here, they start from an isothermal core profile defined by ρ_0 and v_0 in the centre, evaluate r_1 using the core profile, and find the NFW profile on the outskirt that smoothly joins the core at r_1. In this regard, our workflow as described in Section 2.4 is called the *outside-in* approach (e.g. Sagunski et al. 2021). As Robertson et al. (2021) noted, in the inside-out approach, the outer halo is completely determined by the NFW profile and there is no freedom to incorporate contraction. That said, it is still able to capture the effect of baryonic potential on the SIDM profile partially, via the baryonic terms in the Jeans–Poisson equation, equation (5). It is just
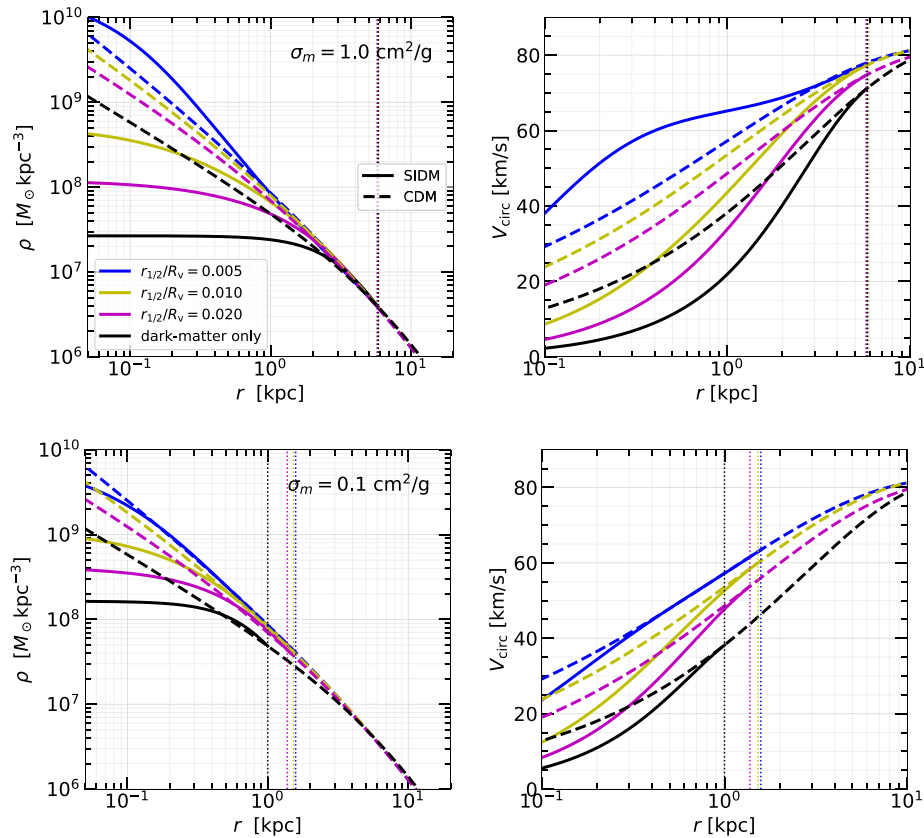
Figure 4. Illustration of the high sensitivity of SIDM halo response to baryonic potentials. The left-hand and right-hand panels show the density profiles and circular velocity profiles, respectively, of SIDM haloes (solid lines) and CDM counterparts (dashed lines). For all the cases, we keep fixed the virial mass of $M_{\rm vir} = 10^{11}\,{\rm M}_\odot$, the target concentration of $c = 10$, and the galaxy mass of $M_{\rm b} = 10^9\,{\rm M}_\odot$, only varying the galaxy size. The colours differentiate the half-mass radii of $r_{1/2} = 0.005, 0.01$, and $0.02 R_{\rm vir}$, as indicated, or equivalently, $r_{1/2} \approx 0.47, 0.95$, and 1.9 kpc – these are representative of compact, normal, and ultra-diffuse dwarf galaxies. The upper and lower panels show the results for cross sections of $\sigma_m = 1\,{\rm cm}^2\,{\rm g}^{-1}$ and $0.1\,{\rm cm}^2\,{\rm g}^{-1}$, respectively. The vertical dotted lines mark the positions of r_1 for the corresponding colour. Clearly, SIDM haloes are more sensitive to baryonic compactness than their CDM counterparts. The strong difference in the inner halo is not driven by the difference in r_1, which is actually negligible for $\sigma_m = 1\,{\rm cm}^2\,{\rm g}^{-1}$ or larger; instead, it arises from the difference in $V_{\rm circ}(r)$, or more precisely in the derivatives of the gravitational potential ${\rm d}\Phi/{\rm d}r = V_{\rm circ}(r)^2/r$ and ${\rm d}^2\Phi/{\rm d}r^2$, as hinted from the right-hand panels.

not entirely self-consistent, as the baryonic potential will affect the entire halo, making the outer part also deviate from NFW.

Here, with the outside-in approach, we can quantify the difference made by including adiabatic halo contraction. We emulate the inside-out model by skipping the halo contraction step of our workflow and only consider the baryonic potential in the Jeans–Poisson equation. The difference is shown in the top and side panels of Fig. 5 – the thick black lines show the halo response from the fiducial model, and the thin grey lines show the result skipping halo contraction (with everything else the same). As can be seen, accounting for adiabatic contraction does not introduce a big difference for galaxies of $M_{\rm b}/M_{\rm vir} \lesssim 0.001$ or for diffuse systems of $r_{1/2}/R_{\rm vir} \gtrsim 0.04$; however, for massive and compact systems, the central density in our fiducial model can be up to four times higher (see e.g. the result at $M_{\rm b}/M_{\rm vir} = 0.04$), and the central density slope can also be different by up to 30 per cent. In short, for massive and compact systems, an explicit adiabatic-contraction treatment must be included for accurate results; for diffuse and dark-matter dominated systems, considering the baryon potential in the Jeans–Poisson equation provides results that are close enough.

5 DISCUSSION

In this section, we first compare the isothermal Jeans model to the more sophisticated gravothermal fluid model, which also predicts SIDM halo profiles and is studied extensively in the literature. Then, we study the facilitation of gravothermal core-collapse by the inhabitant galactic potential, and use the isothermal Jeans model to predict the regime of core-collapse in the space of galaxy mass fraction versus galaxy compactness.

5.1 Comparison with gravothermal fluid evolution

The isothermal Jeans model assumes a system to be in approximate equilibrium, whereas with dark-matter self-interactions, the system is never in strict equilibrium. The full hydrodynamical evolution can be described by the gravothermal fluid model, which is extensively studied in a series of seminal works (Lynden-Bell & Eggleton 1980; Balberg & Shapiro 2002; Koda & Shapiro 2011; Pollack, Spergel & Steinhardt 2015; Essig et al. 2019; Nishikawa, Boddy & Kaplinghat 2020). This method treats SIDM as a gravothermal fluid, and solves a set of coupled partial differential equations for the evolution of
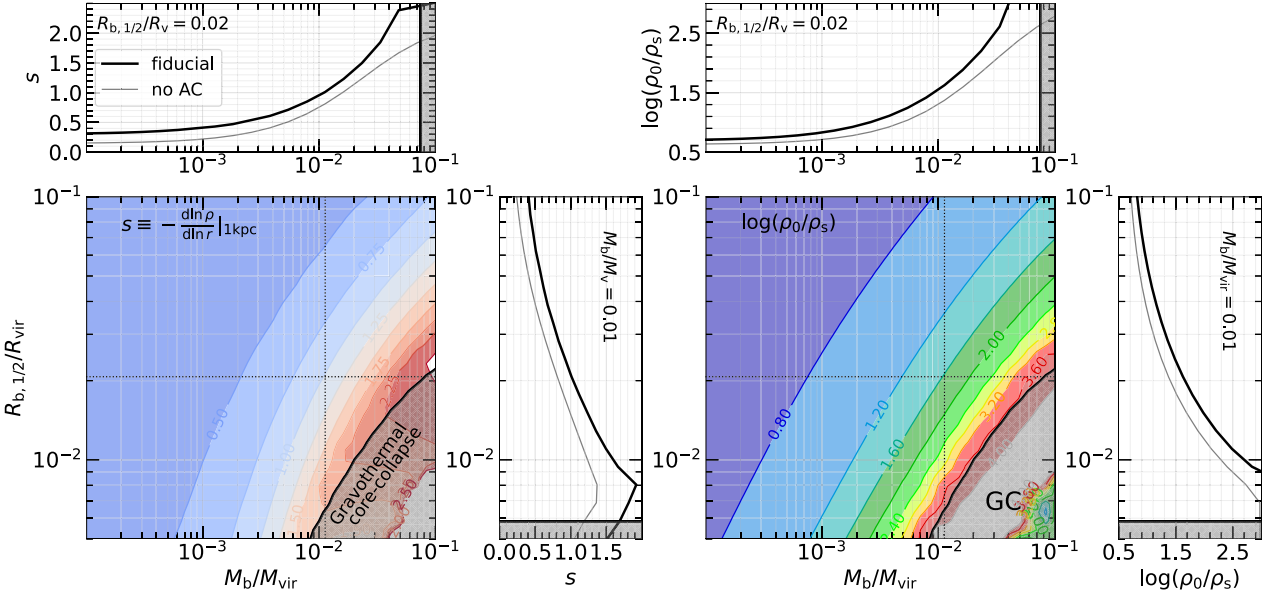
Figure 5. The inner logarithmic density slope (left) and central density (right) of SIDM haloes as functions of the baryonic mass fraction M_b/M_{vir} and the galaxy size in units of the virial radius $r_{1/2}/R_{vir}$. Here, we choose $\sigma_m = 1\,\mathrm{cm^2\,g^{-1}}$ and adopt a target halo formed $t_{age} = 10\,\mathrm{Gyr}$ ago with present-day virial mass $M_{vir} = 10^{11}\,\mathrm{M_\odot}$ and concentration $c = 10$. The slope $s \equiv -\mathrm{d\ln}\,\rho/\mathrm{d\ln}\,r$ is evaluated at $r = 1\,\mathrm{kpc}$, and the central density is expressed in units of the NFW scale density ρ_s. The main panels are the contour maps of s and $\log(\rho_0/\rho_s)$, with the contour-level values indicated. The thick black lines in the top and side panels show one-dimensional slices of the main panels, with one of the baryonic properties fixed at the values indicated by the dotted lines in the main panel. The thin grey lines in the top and side panels show the results without considering adiabatic halo contraction – these are equivalent to the 'inside-out' models of Robertson et al. (2021). Focusing on our fiducial model with adiabatic contraction, the density slope increases from ~0.3 to 2 as M_b/M_{vir} increases from 10^{-4} to 0.04 or as $r_{1/2}/R_{vir}$ decreases from 0.1 to 0.005, for the specific slices. For $M_b/M_{vir} \sim 0.01$ and $r_{1/2}/R_{vir} \sim 0.02$, representative of Milky-Way-mass galaxies according to abundance matching, the SIDM density profile is actually rather similar to the CDM case with an inner slope of ~1. Accounting for adiabatic contraction makes the central density up to four times higher (e.g. at $M_b/M_{vir} = 0.04$) and the central density slope ~30 per cent steeper. In the lower right corner of the space, an isothermal solution can no longer be achieved, manifesting the speed-up of gravothermal core-collapse (GC).

the spherically symmetric profiles of mass $M(r, t)$, density $\rho(r, t)$, velocity dispersion $v(r, t)$, and the luminosity of the radiated heat $L(r, t)$ –

$$\frac{\partial \tilde{M}}{\partial \tilde{r}} = \tilde{r}^2 \tilde{\rho},$$

$$\frac{\partial (\tilde{\rho}\tilde{v}^2)}{\partial \tilde{r}} = -\frac{\tilde{M}\tilde{\rho}}{\tilde{r}^2},$$

$$\frac{\partial \tilde{L}}{\partial \tilde{r}} = -\tilde{r}^2 \tilde{\rho}\tilde{v}^2 \left(\frac{\partial}{\partial \tilde{t}}\right)_{\tilde{M}} \ln\left(\frac{\tilde{v}^3}{\tilde{\rho}}\right),$$

$$\tilde{L} = -\frac{3}{2}\tilde{r}^2 \tilde{v} \left(\frac{a}{b}\tilde{\sigma}_m^2 + \frac{1}{C\tilde{\rho}\tilde{v}^2}\right)^{-1} \frac{\partial \tilde{v}^2}{\partial \tilde{r}}. \quad (16)$$

These equations describe mass conservation, hydrostatic equilibrium, the first law of thermodynamics, and heat conduction, respectively, where $a = 4/\sqrt{\pi}$, $b = 25\sqrt{\pi}/32$, and C is a calibration parameter of order unity. Following Koda & Shapiro (2011) and Nishikawa et al. (2020), we have expressed the equations with the dimensionless quantities: $\tilde{r} \equiv r/r_s$, $\tilde{\rho} \equiv \rho/\rho_s$, $\tilde{M} \equiv M/M_0$ with the mass scale $M_0 = 4\pi r_s^3 \rho_s$, $\tilde{\sigma}_m \equiv \sigma_m/\sigma_{m0}$ with the cross-section scale $\sigma_{m0} = 1/r_s\rho_s$, $\tilde{v} \equiv v/v_0$ with the velocity scale $v_0 = \sqrt{GM_0/r_s}$, $\tilde{L} \equiv L/L_0$ with the luminosity scale $L_0 \equiv GM_0^2/r_s t_0$, and $\tilde{t} \equiv t/t_0$ with the time-scale $t_0 = 1/a\sigma_m v_0 \rho_s$. This assumes that the initial profile $\rho(r, t = 0)$ is NFW, with scale radius r_s and scale density ρ_s. With the dimensionless quantities, we have the convenience that the density-profile evolution is self-similar as long as we are in the long-

mean-free-path[3] regime, and thus the result is almost independent of the cross-section or the initial NFW concentration when expressed in $\tilde{\rho}(\tilde{r}, \tilde{t})$. This is illustrated in Balberg & Shapiro (2002) and in Appendix C of Nishikawa et al. (2020).

There are a few differences between the isothermal Jeans model and the fluid model. First and foremost, conceptually, the fluid model gives the full (time-dependent) solution to the Boltzmann Equations with an assumed conductivity; while the isothermal model approximates the instantaneous profile as being in equilibrium, and therefore does not have time evolution per se other than a dependence on halo age. Second, the isothermal model is only applicable to the isothermal-coring stage and the onset of gravothermal core-collapse; while the fluid model can follow the evolution well into core-collapse. Third, solving the fluid equations requires discretizing the spherical halo and is relatively computationally expensive; whereas the isothermal model only requires performing the minimization at r_1, and within each iteration, the numerical integration of the Jeans–Poisson equation is quite fast. The speed advantage makes it easier for incorporating into large semi-analytic frameworks. Fourth, the fluid model only considers the dark-matter component, at least as presented in the literature so far; while the isothermal model easily accounts for baryonic effects by including baryonic terms in the Jeans–Poisson equation and by considering adiabatic contraction. For this reason, when we compare the two models, we focus on

---

[3]That is, when the mean free path of scattering, $\lambda = 1/\rho v$, is larger than the gravitational scale length, $H = \sqrt{v^2/4\pi G\rho}$.
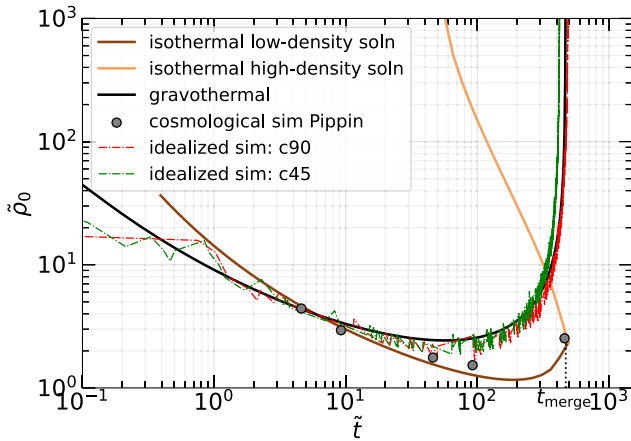
Figure 6. Comparison of the gravothermal fluid model (black solid line) and the isothermal Jeans model (brown solid line) in terms of the (dimensionless) central density $\tilde{\rho}_0 \equiv \rho_0/\rho_s$ as a function of time $\tilde{t} \equiv t/t_0$. See Section 5.1 for the definitions and the details of the calculations. Note that, the details of the target CDM halo or the cross-section have little impact on the dimensionless $\tilde{\rho}_0(\tilde{t})$ track. Simulation results are overplotted for comparison – the grey circles represent the Pippin *cosmological* simulations of different cross-sections $\sigma_m = 0.5, 1, 5, 10$, and $50 \, \mathrm{cm^2 \, g^{-1}}$ at $z = 0$; the red and green dash–dotted lines represent the *idealized* isolated simulations starting from NFW profiles with $c = 45$ and 90, and with $\sigma_m = 10 \, \mathrm{cm^2 \, g^{-1}}$. The isothermal model agrees better with the cosmological results, while the fluid model agrees with the idealized simulations – their difference likely originates from whether the target CDM halo is used as an *initial condition* or as a *boundary condition* (see Section 5.1 for discussion). The orange solid line represents the usually discarded high-density solution of the isothermal model. The point when the high-density and low-density solutions merge coincides with when gravothermal core-collapse speeds up and the core temperature is well above the velocity dispersion of the CDM-like outskirt (see Section 5.2).

the dark-matter-only set-ups.[4] Finally, the fluid model can easily adapt to velocity-dependent cross sections – one can simply plug a v-dependent cross section $\sigma_m(v)$ in the fourth equation of equation (16); while the isothermal model evaluates the r_1 radius using the instantaneous cross-section, and thus ignores any v-dependence. For typical particle-physics models, the v-dependence effectively makes the cross section larger in the past and thus makes the isothermal coring faster (Nadler et al. 2020). That said, if we know the growth history of the target CDM halo including the velocity dispersion profile as a function of redshift $v(r, z)$, then we can solve for an $r_1(t_{\mathrm{age}})$ that includes the time dependence

$$1 = \int_0^{t_{\mathrm{age}}} \rho(r_1, t) v(r_1, t) \sigma_m[v(r_1, t)] \mathrm{d}t, \qquad (17)$$

where $t(z)$ is the lookback time. We can therefore perform the isothermal Jeans modelling for each time and construct a density-profile evolution $\rho(r, z)$ that approximates the case of a v-dependent cross-section.

We summarize these similarities and differences of the two methods in Table 1. Overall, the isothermal Jeans model is simplistic yet much faster. In Fig. 6, we compare the two models in the space of the dimensionless central density $\tilde{\rho}_0$ versus the dimensionless time $\tilde{t}$. For the fluid model, $\tilde{\rho}_0(\tilde{t})$ is simply obtained by solving

equation (16). We have followed the numerical method as detailed in Nishikawa et al. (2020), starting from the NFW profile of the $z = 0$ CDM Pippin halo (i.e. solid grey line in Fig. 2), adopting a cross section of $\sigma_m = 5 \, \mathrm{cm^2 \, g^{-1}}$, and using $C = 0.75$ as calibrated to idealized simulations (Koda & Shapiro 2011). Despite the specific choices, we emphasize that the cross-section, the details of the NFW profile, or the exact value of C as long as it is between 0.5 and 1, has weak impact on the result in this dimensionless space in the core-forming regime. For the isothermal model, in order to construct the 'time evolution', we repeat the exercise for a series of halo age t_{age} and plot $\tilde{\rho}_0$ versus $\tilde{t} \equiv t_{\mathrm{age}}/t_0$. The same target CDM halo and cross-section are used for both methods. Again, these details are largely irrelevant for this dimensionless parameter space due to the self-similar nature of the density evolution in the core-formation regime, and we have verified with the isothermal Jeans model that it predicts a universal track in the $\tilde{\rho}_0$–$\tilde{t}$ space for different σ_m. For the isothermal model, in addition to the default, low-density solution, we also record the high-density solution, and display both solutions in Fig. 6. We reiterate that only the low-density solution is supposed to be comparable to the simulation results or the fluid model predictions.

As can be seen, both models show a similar qualitative behaviour – an isothermal core grows as the density keeps decreasing; then the central density reaches a minimum and turns around, manifesting the onset of gravothermal core-collapse. However, there is a clear difference: with the isothermal model, the core develops faster, and reaches a minimum central density that is ~ 2 times lower than that predicted by the fluid model, at a slightly later time. This difference cannot be attributed to the calibration parameter C. In fact, smaller (larger) C makes the turn-around of $\tilde{\rho}_0$ occur later (earlier), but it has little impact on the steepness of the isothermal-coring stage.

What causes the difference? Which model is more accurate? To get some clues, we compare the model predictions to simulation results of different kinds. First, we compare to the *cosmological* Pippin N-body simulations of Elbert et al. (2015). Following Essig et al. (2019), an 'evolutionary' track $\tilde{\rho}_0(\tilde{t})$ can be constructed using the simulation results all at the same time of $z = 0$. This is because the dimensionless time $\tilde{t} \equiv t(z)/t_0 \propto \sigma_m^{-1}$, where the cosmic time at $z = 0$ is $t = 13.7 \, \mathrm{Gyr}$ for the Pippin cosmology, so different cross-sections correspond to different dimensionless times. Specifically, the Pippin halo was run with cross-sections $\sigma_m = 0.5, 1, 5, 10$, and $50 \, \mathrm{cm^2 \, g^{-1}}$, and the central densities at $z = 0$ are $\rho_0 = 7.5, 5.0, 3.0, 2.6$, and $4.3 \times 10^7 \, \mathrm{M_\odot \, kpc^{-3}}$, respectively. The CDM counterpart has $\rho_s = 1.7 \times 10^7 \, \mathrm{M_\odot \, kpc^{-3}}$ and $r_s = 2.7 \, \mathrm{kpc}$. Hence, the dimensionless central densities are $\tilde{\rho}_0 = 4.4, 2.9, 1.8, 1.5$, and 2.5, which are reached at the dimensionless times of $\tilde{t} = 4.6, 9.2, 46, 92$, and 460, respectively. Interestingly, the isothermal Jeans model, albeit simplistic, agrees with the cosmological Pippin simulations very well. Notably, the steeper isothermal coring at $\tilde{t} \lesssim 100$ is the same, and the last simulation data point at $\tilde{t} = 460$, which exhibits core-collapse, is almost on top of the model prediction. This time happens to be when the low-density solution and the high-density solution merge, beyond which the isothermal Jeans model is no longer applicable. Mathematically, for a continuously evolving quantity (such as the central density $\tilde{\rho}_0$) that has two solutions, any transition between the solutions must be continuous and therefore any continuous parameter (such as time $\tilde{t}$) must enable a smooth transition between the solutions. In this sense, the transition is when the density increases and that is the onset of core collapse. Physically, beyond this time, a negative velocity-dispersion gradient starts to develop so the isothermal assumption breaks (see Section 5.2 and Appendix A for more discussion). In practice, the merging point is manifested by where the 'stitching error' δ^2 can no longer be minimized to zero.

---

[4]It is in principle possible to include a static baryon component in the second equation of equation (16) and thus make the fluid model capture baryon effects as well, but this is beyond the scope of this work.

Second, we compare the model predictions to *idealized* SIDM *N*-body simulations of isolated haloes. To this end, we simulate with the AREPO code (Springel 2010; Weinberger, Springel & Pakmor 2020) two Milky-Way sized haloes, which are initialized with NFW profiles at $z = 0$ with $c = 45$ and 90, respectively, and are evolved with a self-interacting cross section of $10 \, \mathrm{cm^2 \, g^{-1}}$. The details of the simulations are provided in Appendix B. Again, $\tilde{\rho}_0(\tilde{t})$ is not sensitive to the details of the target halo or the cross-section, and the high concentrations are chosen simply to facilitate the gravothermal evolution and shorten the computation time. Clearly, the fluid model agrees well with the idealized simulations, whereas the isothermal model agrees better with cosmological results.

We hypothesize that the difference originates from how the target CDM halo is used in the modelling. Specifically, in the fluid model, the present-day target CDM halo is used to *initialize* the system. That is, there are two implicit assumptions here: first, for the entire history of the target halo up until $t = 0$, dark matter remains collisionless, and only at $t > 0$, DM becomes self-interacting; second, the halo stops mass accretion and evolves in isolation at $t > 0$. As such, the profile we obtain at time t ($t > 0$) is virtually that of an isolated system at a future cosmic time of t + the age of the Universe today, with the effect of self-interactions during the entire assembly of the target halo not taken into account. It is therefore not surprising that the fluid model disagrees with the cosmological results but agrees better with the idealized simulations which essentially make the same implicit assumptions.

In contrast, in the isothermal model, the target CDM halo at $z = 0$ is *not* treated as an *initial condition*, but instead used for the *boundary condition* at r_1. In the context of trying to understand why distinct CDM haloes all have the universal NFW shape, it has been well established that there is a correspondence between the density-profile shape and the shape of the mass assembly history (e.g. Ludlow et al. 2013). In this regard, using the target NFW profile to set the boundary condition means that we have implicitly used some information of the cosmological mass assembly history of the halo. It is therefore reasonable to expect agreement with the cosmological simulations.

It is still remarkable that the simplistic stitching at r_1 results in this high level of agreement and we caution against overinterpretating it physically. But we have verified that altering the detailed definition of r_1 does not change the qualitative agreement. For example, multiplying a constant factor in equation (1) will not change the overall shape of the $\tilde{\rho}_0(\tilde{t})$ track. This implies that, when the isothermal assumption is valid, there is no hysteresis of the core. After all, the isothermal state is a thermodynamic equilibrium, so does not depend on how the state was reached.

We also caution that the comparison in Fig. 6 should not be interpreted as a criticism of the fluid model, but instead a clarification of what it does as implemented in the literature (Lynden-Bell & Eggleton 1980; Balberg & Shapiro 2002; Koda & Shapiro 2011; Pollack et al. 2015; Essig et al. 2019; Nishikawa et al. 2020). To adapt it for better cosmological usage, one may want to explore revisions of the sort of the following. In particular, in order to model the SIDM counterpart of a CDM halo at $z = 0$, it is reasonable to adopt the CDM profile at $z = z_{\mathrm{form}}$ as the initial condition, where z_{form} is a characteristic formation redshift of the halo. Accordingly, the mass, density, and velocity dispersion in equation (16) shall be updated according to cosmological average trends to account for the growth history of the halo. For example, for each time-step, one can add a mass increment to each radius bin self-similarly according to the instantaneous density profile, where the sum of the mass depositions across all the bins is equal to the average mass growth in that time-step. There are well established empirical mass assembly histories
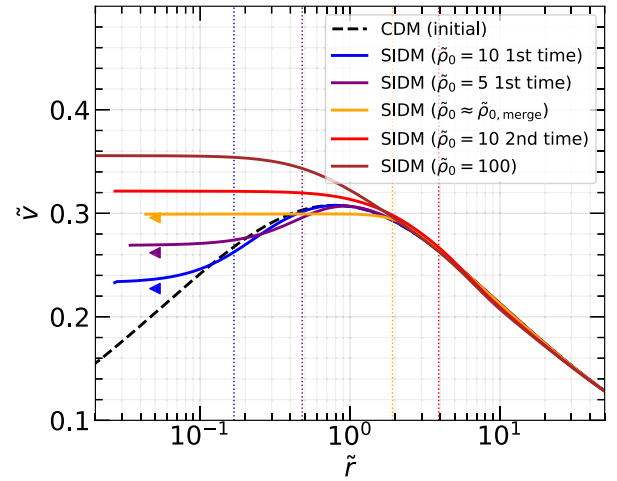

Figure 7. Evolution of the velocity dispersion profile from the fluid model. The dashed black line represents the initial CDM profile. The coloured solid lines represent the SIDM profiles at different times, as indicated. Notably, the orange line represents the result when $\tilde{\rho}_0$ reaches the value when the low-density and the high-density solutions merge in the isothermal Jeans model, $\tilde{\rho}_{0,\mathrm{merge}}$. Since the central density $\tilde{\rho}_0$ initially decreases and later turns around, it will reach certain values twice – e.g. both the blue and red lines here correspond to $\tilde{\rho}_0 = 10$, but the former is during core-formation while the latter is during core collapse. The dotted vertical lines indicate the corresponding r_1 radii. The triangles indicate the isothermal Jeans solutions of the central velocity dispersions (only for the first three cases, because the isothermal model is no longer applicable to the later stages of evolution). Obviously, beyond $\tilde{\rho}_{0,\mathrm{merge}}$, a negative velocity–dispersion gradient develops at $r \sim r_1$, core-collapse speeds up, and therefore the isothermal model stops working as the assumption of constant velocity dispersion at $r < r_1$ is no longer valid.

and mass-concentration-redshift relations from CDM simulations (e.g. McBride, Fakhouri & Ma 2009; Dutton & Macciò 2014). We explore improvements of this sort in a future study (Yang et al. in preparation).

5.2 Gravothermal core-collapse and facilitation by the inhabitant galaxy

As Fig. 6 shows, gravothermal core-collapse occurs, i.e. the central density starts to increase, at $\tilde{t} \gtrsim 100$. Soon after the onset of core-collapse, one can see with the fluid model that the central velocity dispersion increases to a level that is higher than the CDM $v(r_1)$, and thus a steep negative velocity-dispersion gradient occurs at $r \lesssim r_1$, as shown in Fig. 7. Then, the flat isothermal core becomes significantly smaller than r_1 and thus gravothermal core-collapse speeds up. Recall that the key assumption for the isothermal Jeans model is that the core has constant $v = v_0$ throughout the region $r < r_1$. This assumption holds at the onset of core-collapse, which is why the isothermal model is still able to capture the upturn in the central density. But the isothermal method fails as core-collapse continues, because r_1, as defined in equation (1), increases with time, and therefore the assumption of $v = v_0$ within r_1 breaks when v_0 increases to be significantly higher than the peak of the CDM $v(r)$ profile.

Recall that in the isothermal model, we accept the lower-density solution because realistic haloes form with properties closer to it. However, we emphasize that both the low-density and high-density solutions are physical, as they satisfy the Jeans–Poisson equation with constant velocity dispersion within r_1. A smooth
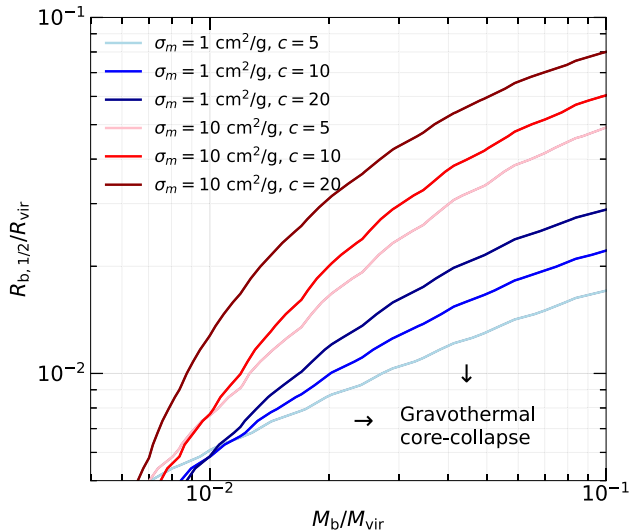
Figure 8. Critical baryonic properties for gravothermal core-collapse – the same as the boundary of the core-collapse region in Fig. 5, but for different halo concentrations and SIDM cross-sections, as indicated. The way to read this is: if the inhabitant galaxy has higher mass or more compact size than this threshold, the host SIDM halo formed $\sim 10\,\mathrm{Gyr}$ ago will start core-collapse. Practically, systems sitting on the threshold have the two isothermal solutions merged into one. Beyond this threshold to the lower right corner, a smooth joint with $\delta^2 < 0.01$ between the SIDM core and the CDM-like outskirt is no longer achievable, or equivalently speaking, it is no longer possible to have constant velocity dispersion within r_1. See Section 5.2 for details.

transition between them is achieved shortly after the central density starts to increase, manifesting core-collapse. Therefore, we can practically use the moment when the two solutions merge as an indicator of gravothermal core-collapse.

The inhabitant galaxy facilitates core-collapse by making the halo contract in the first place. Naturally, this effect is particularly strong when the galaxy is massive and compact. To illustrate this, we highlight the region of core-collapse in the $r_{1/2}/R_{\mathrm{vir}} - M_{\mathrm{b}}/M_{\mathrm{vir}}$ space in Fig. 5. The operational definition of this region is that: for galaxies on the border of this region, the SIDM haloes that formed $10\,\mathrm{Gyr}$ ago have started core-collapse, such that no isothermal solution exists that joins smoothly the CDM outskirt (with $\delta^2 < 0.01$).

The region for core-collapse depends on the target CDM concentration and the self-interacting cross-section, and becomes larger for higher c and σ_m. This is illustrated in Fig. 8. For instance, at $M_{\mathrm{b}}/M_{\mathrm{vir}} = 0.02$ fixed, haloes with $r_{1/2}/R_{\mathrm{vir}} \lesssim 0.01$ will collapse if $\sigma_m = 1\,\mathrm{cm}^2\,\mathrm{g}^{-1}$. The exact size limit is slightly lower for lower concentration. The galaxy-size limit becomes $r_{1/2}/R_{\mathrm{vir}} \lesssim 0.02$ for $\sigma_m = 10\,\mathrm{cm}^2\,\mathrm{g}^{-1}$. Given that numerous galaxies populate the region $r_{1/2}/R_{\mathrm{vir}} \sim 0.01$–$0.02$ observationally, this parameter space may potentially provide useful constraints on SIDM models. To this end, however, we think that it still requires more detailed understanding of how the inhabitant galaxies react to core-collapse. Regardless, the baryon-facilitated core-collapse itself might be a viable way to create compact bright dwarfs, which are common in the real Universe but are difficult to produce in cosmological hydro-simulations.

For dwarf galaxies with $M_{\mathrm{b}}/M_{\mathrm{vir}} \ll 0.01$, there is basically no constraint on how compact the galaxy can get before core-collapse kicks in, as long as c and σ_m are not extremely high.

In short, with the current implementation of the isothermal Jeans model, although we cannot self-consistently describe core-collapse, we can phenomenologically delineate the onset of core-collapse

as a function of the baryonic properties, given the target halo concentration and the cross-section. We caution that, whereas we have expressed the baryonic properties in units of the virial quantities, the SIDM halo response, as illustrated in Figs 5 and 8, is not universal across different virial masses but only applicable to the halo mass adopted, $10^{11}\,\mathrm{M}_\odot$. This is because the scattering rate increases with the velocity scale of the halo, and thus more massive halos (for a given concentration and formation time) would have stronger halo response and earlier core-collapse than lower mass counter parts.

6 CONCLUSION

In this paper, we combine the isothermal Jeans model and the prescription for adiabatic halo contraction into a fast semi-analytic procedure for calculating the density profile of SIDM haloes. This method takes the inputs of (1) a target CDM halo described by an NFW profile, and (2) an observable baryon distribution described by a Hernquist profile. It computes the contraction, fits the contracted CDM halo with a Dekel–Zhao profile, and stitches an isothermal core to the CDM outskirt at the characteristic radius r_1 by minimizing the fractional difference in density and enclosed mass. We have shown that this model works remarkably well compared to cosmological SIDM simulations both in dark-matter-only set-ups (Pippin) and with hydrodynamics and star formation (FIRE2-SIDM). We provide a simple CORENFW approximation formula for the dark-matter-only cases, where the characteristic core size of $r_{\mathrm{c}} = 0.45r_1$ universally applies to a wide range of cross-sections and target CDM halo concentrations.

We use this model to study the response of SIDM haloes to their inhabitant galaxies. We show that the halo response to the baryonic potential is more intensified and more diverse in SIDM than in CDM. Notably, depending on the compactness of the baryonic distribution, the central dark-matter density slope can be cored, equally cuspy, or cuspier than the CDM counterpart – a desirable feature in the context of the structural diversity of bright dwarf galaxies. We note that, while the model does not capture feedback-driven halo expansion and only considers adiabatic contraction, it agrees well with the FIRE2-SIDM simulations which incorporate both effects. We therefore argue that the dominant baryonic effect in the context of SIDM is adiabatic contraction, and that the details of baryonic feedback may be unimportant in SIDM models.

The fast speed of the numerical implementation of the model enables the following analyses that would be otherwise challenging for numerical simulations. We quantify the SIDM halo response on a fine mesh grid spanned by the baryon-to-total mass ratio $M_{\mathrm{b}}/M_{\mathrm{vir}}$ and the ratio between the half mass radius and the virial-radius $r_{1/2}/R_{\mathrm{vir}}$, in terms of the central logarithmic density slope, $s \equiv \mathrm{d}\ln\rho/\mathrm{d}\ln r|_{1\,\mathrm{kpc}}$, as well as the core density in units of the scale density of the reference CDM halo, ρ_0/ρ_{s}. With this, we are able to confirm with unprecedented precision that for typical Milky-Way-like hosts, the SIDM profiles are similar to their CDM counterparts – an assumption often used in semi-analytic or idealized studies of SIDM satellite galaxies.

We also delineate the regime of gravothermal core-collapse in the space of galaxy mass versus galaxy size, $M_{\mathrm{b}}/M_{\mathrm{vir}} - r_{1/2}/R_{\mathrm{vir}}$. This can be done for any choice of the cross-section and the target CDM halo concentration. For any given baryon-to-total ratio, there is a limit on how compact the galaxy can get in terms of $r_{1/2}/R_{\mathrm{vir}}$, beyond which core-collapse will be triggered within the Hubble time. This threshold is lower (i.e. galaxies can be more compact) if the target CDM halo concentration is smaller or if the cross-section is smaller. With $c = 10$ and $\sigma_m = 1\,\mathrm{cm}^2\,\mathrm{g}^{-1}$, galaxy sizes cannot be smaller than $\sim 0.01R_{\mathrm{vir}}$ for typical baryon-to-total ratios of ~ 0.02. Given that

numerous galaxies have $r_{1/2} \sim 0.01 R_{\rm vir}$, we think that this baryon-facilitated gravothermal core-collapse may provide useful constraints on SIDM models, if we can better understand how galaxies react to core-collapse.

Finally, we compare the isothermal Jeans model with the more sophisticated gravothermal fluid model which is extensively studied in the literature. We show that the isothermal model agrees better with cosmological simulations: they both show a steep per central-density decrease in the isothermal coring regime and a later gravothermal core-collapse compared to the fluid model. On the contrary, the fluid model agrees well with idealized simulations of isolated haloes initialized with NFW profiles. We argue that the difference originates from whether the target CDM profile is used for the *boundary condition* (as in the case of the isothermal model) or as the *initial condition* (as in the case of the fluid model).

We have made our programs publicly available, including the programs for computing the profiles of SIDM haloes with baryons, as well as the programs that calculate the threshold for gravothermal core-collapse in the $M_{\rm b}/M_{\rm vir} - r_{1/2}/R_{\rm vir}$ space. They can be downloaded at https://github.com/JiangFangzhou/SIDM. While we stick to Hernquist galaxies in the paper for self-consistency (as equation 5 is based on Hernquist galaxies), the adiabatic contraction model of Gnedin et al. (2004) actually also accommodates exponential disks and is implemented in the code.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The data underlying this article produced by the semi-analytical model are generated using programs available at https://github.com/JiangFangzhou/SIDM. The data of the cosmological simulations are either directly given in the paper or adopted from published works as cited in the paper. The data of the idealized simulations will be available upon reasonable requests.

REFERENCES

Balberg S., Shapiro S. L., 2002, Phys. Rev. Lett., 88, 101301
Benson A. J., 2012, New Astron., 17, 175
Blok W. J. G. d., Walter F., Brinks E., Trachternach C., Oh S.-H., Kennicutt R. C., 2008, AJ, 136, 2648
Blumenthal G. R., Faber S. M., Flores R., Primack J. R., 1986, ApJ, 301, 27
Van den Bosch F. C., Ogiya G., 2018, MNRAS, 475, 4066
Bose S. et al., 2019, MNRAS, 486, 4790
Chilingarian I., Zolotukhin I., 2015, Sci., 348, 418
Colin P., Avila-Reese V., Valenzuela O., Firmani C., 2002, ApJ, 581, 777
Creasey P., Sameie O., Sales L. V., Yu H.-B., Vogelsberger M., Zavala J., 2017, MNRAS, 468, 2283
Cruz A. et al., 2021, MNRAS, 500, 2177
Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359
Elbert O. D., Bullock J. S., Garrison-Kimmel S., Rocha M., Oñorbe J., Peter A. H. G., 2015, MNRAS, 453, 29

Elbert O. D., Bullock J. S., Kaplinghat M., Garrison-Kimmel S., Graus A. S., Rocha M., 2018, ApJ, 853, 109
Essig R., McDermott S. D., Yu H.-B., Zhong Y.-M., 2019, Phys. Rev. Lett., 123, 121102
Freundlich J. et al., 2020, MNRAS, 499, 2912
Gnedin O. Y., Kravtsov A. V., Klypin A. A., Nagai D., 2004, ApJ, 616, 16
Gnedin O. Y., Ceverino D., Gnedin N. Y., Klypin A. A., Kravtsov A. V., Levine R., Nagai D., Yepes G., 2011, preprint (arXiv:1108.5736)
Jiang F., Dekel A., Freundlich J., Romanowsky A. J., Dutton A. A., Macciò A. V., Cintio A. D., 2019, MNRAS, 487, 5272
Jiang F., Dekel A., Freundlich J., van den Bosch F. C., Green S. B., Hopkins P. F., Benson A., Du X., 2021, MNRAS, 502, 621
Kamada A., Kaplinghat M., Pace A. B., Yu H.-B., 2017, Phys. Rev. Lett., 119, 111102
Kaplinghat M., Keeley R. E., Linden T., Yu H.-B., 2014, Phys. Rev. Lett., 113, 021302
Kaplinghat M., Tulin S., Yu H.-B., 2016, Phys. Rev. Lett., 116, 041302
Kaplinghat M., Ren T., Yu H.-B., 2020, J. Cosm. Astropart. Phys., 2020, 027
Kochanek C. S., White M., 2000, ApJ, 543, 514
Koda J., Shapiro P. R., 2011, MNRAS, 415, 1125
Koda J., Yagi M., Yamanoi H., Komiyama Y., 2015, ApJ, 807, L2
Komatsu E. et al., 2011, ApJS, 192, 47
Lelli F., McGaugh S. S., Schombert J. M., 2016, AJ, 152, 157
Ludlow A. D. et al., 2013, MNRAS, 432, 1103
Lynden-Bell D., Eggleton P. P., 1980, MNRAS, 191, 483
McBride J., Fakhouri O., Ma C.-P., 2009, MNRAS, 398, 1858
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Nadler E. O., Banerjee A., Adhikari S., Mao Y.-Y., Wechsler R. H., 2020, ApJ, 896, 112
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Nishikawa H., Boddy K. K., Kaplinghat M., 2020, Phys. Rev. D, 101, 063009
Oh S.-H. et al., 2015, AJ, 149, 180
Peter A. H. G., Rocha M., Bullock J. S., Kaplinghat M., 2013, MNRAS, 430, 105
Pollack J., Spergel D. N., Steinhardt P. J., 2015, ApJ, 804, 131
Read J. I., Agertz O., Collins M. L. M., 2016, MNRAS, 459, 2573
Relatores N. C. et al., 2019, MNRAS, 887, 94
Ren T., Kwa A., Kaplinghat M., Yu H.-B., 2019, Phys. Rev. X, 9, 031020
Robertson A., Harvey D., Massey R., Eke V., McCarthy I. G., Jauzac M., Li B., Schaye J., 2019, MNRAS, 488, 3646
Robertson A., Massey R., Eke V., Schaye J., Theuns T., 2021, MNRAS, 501, 4610
Rocha M., Peter A. H. G., Bullock J. S., Kaplinghat M., Garrison-Kimmel S., Oñorbe J., Moustakas L. A., 2013, MNRAS, 430, 81
Sagunski L., Gad-Nasr S., Colquhoun B., Robertson A., Tulin S., 2021, J. Cosm. Astropart. Phys., 2021, 024
Sameie O. et al., 2021, MNRAS, 507, 720
Shen X., Hopkins P. F., Necib L., Jiang F., Boylan-Kolchin M., Wetzel A., 2021, MNRAS, 506, 4421
Shi Y., Zhang Z.-Y., Wang J., Chen J., Gu Q., Yu X., Li S., 2021, ApJ, 909, 20
Slone O., Jiang F. Lisanti M., Kaplinghat M., 2022, Phys. Rev. D, 107, 4
Somerville R. S. et al., 2018, MNRAS, 473, 2714
Springel V., 2010, MNRAS, 401, 791
Vogelsberger M., Zavala J., Loeb A., 2012, MNRAS, 423, 3740
Weinberger R., Springel V., Pakmor R., 2020, ApJS, 248, 32
Zeng Z. C., Peter A. H. G., Du X., Benson A., Kim S., Jiang F., Cyr-Racine F.-Y., Vogelsberger M., 2022, MNRAS, 513, 4845
Zentner A., Dandavate S., Slone O., Lisanti M., 2022, J. Cosm. Astropart. Phys., 2022, 031

APPENDIX A: TWO SOLUTIONS OF THE ISOTHERMAL JEANS STITCHING

In Section 2 and Fig. 1, we illustrated the workflow of the isothermal Jeans model and showed that there are two islands of minima of the 'stitching error' at r_1, in the space of central dark-matter density ρ_0
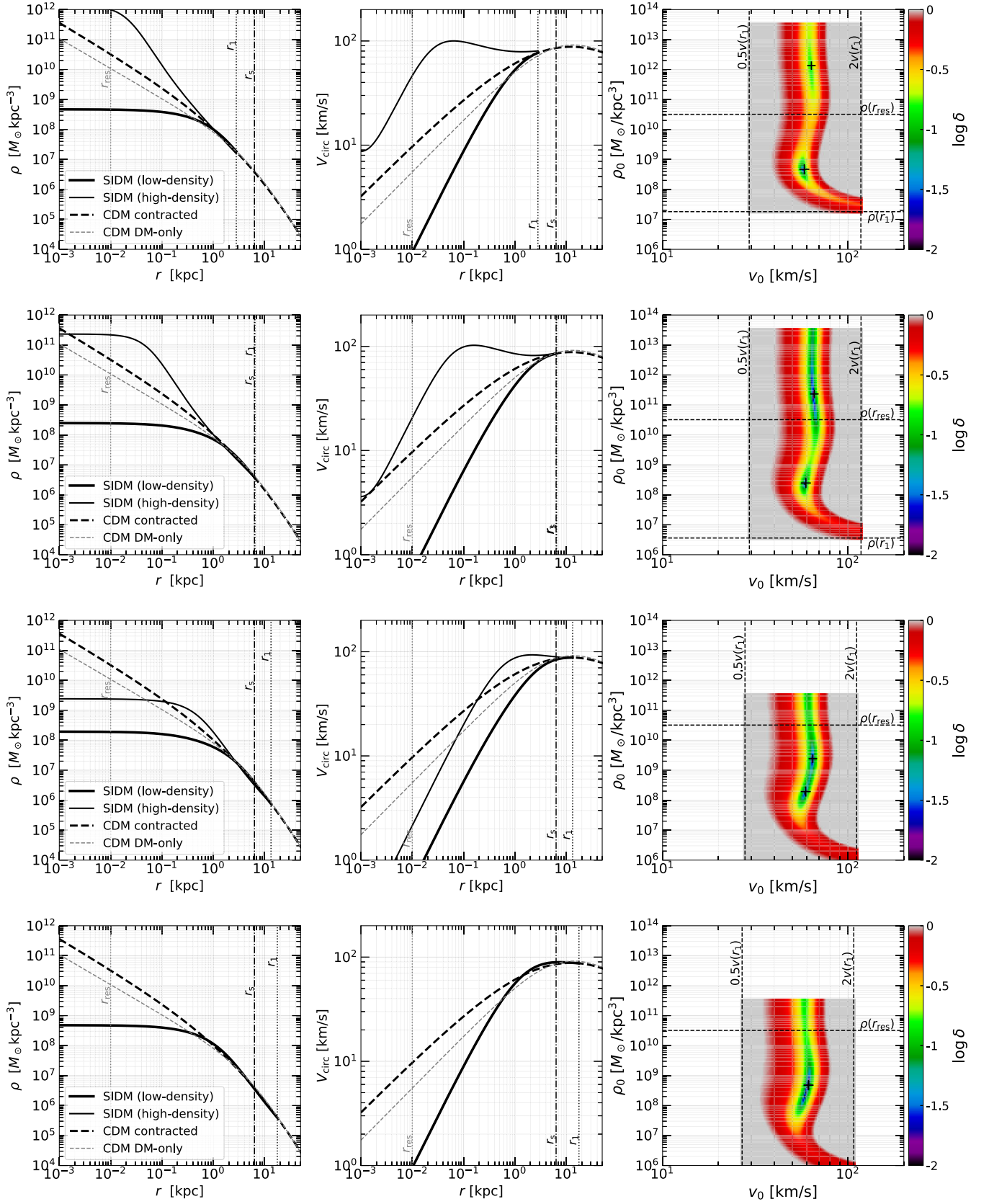
Figure A1. The same as Fig. 1, but for a series of different halo ages: $t_{age} = 2, 10, 50,$ and 100 Gyr. Note that: (1) the central density ρ_0 decreases first at $t_{age} \lesssim 50$ Gyr and then increases between $t_{age} = 50$ Gyr and 100 Gyr; (2) the low-density and high-density solutions get closer as the system evolves and finally merge – this is when the isothermal assumption starts to break, with v_0 reaching the highest value allowed by the isothermal assumption, and this is when gravothermal core-collapse kicks in.

and the core velocity dispersion v_0. There, we showed an example of a system of $t_{age} = 5\,\mathrm{Gyr}$, $M_{vir} = 10^{11}\,\mathrm{M_\odot}$, $c = 15$, $M_b = 10^9\,\mathrm{M_\odot}$, and $r_{1/2} = 1.9\,\mathrm{kpc}$, for a cross section of $\sigma_m = 1\,\mathrm{cm^2\,g^{-1}}$. Here, as shown in Fig. A1, we extend the exercise to a series of different halo ages, $t_{age} = 2, 10, 50$, and $100\,\mathrm{Gyr}$, with everything else the same. This effectively shows the evolution of the system.

As the system evolves, the two minima of δ^2 first both decrease in ρ_0 ($t_{age} = 2$ and $10\,\mathrm{Gyr}$); then, the lower-density solution turns around ($t_{age} = 50\,\mathrm{Gyr}$) and finally the two solutions merge ($t_{age} = 100\,\mathrm{Gyr}$), marking the onset of gravothermal core-collapse.

This trend actually holds as long as the system 'evolves' in terms of the dimensionless time $\tilde{t} \equiv t_{age}/t_0 = 8\sqrt{G}\sigma_m\rho_s^{3/2}r_s t_{age}$, so it can be achieved also by increasing σ_m or c. For example, the central density track of the Pippin simulations as we showed in Fig. 6 is obtained by increasing σ_m with everything else fixed.

APPENDIX B: IDEALIZED SIMULATIONS

For Section 5.1, in addition to comparing with the published cosmological Pippin simulations, we also compared the models to idealized simulations of isolated SIDM haloes using the AREPO code (Springel 2010; Weinberger et al. 2020). AREPO comes with a default module of dark matter self-interactions with the form of two-body scattering (Vogelsberger et al. 2012). This code is intensively used in the recent study of Zeng et al. (2022) on SIDM subhaloes. The initial conditions are generated with NFW profiles of $M_{vir} = 10^{12}\,\mathrm{M_\odot}$ with a concentration parameter of $c = 45$ or 90, using the code SPHERIC. High concentration values are adopted to facilitate the gravothermal evolution. The particle mass is $10^7\,\mathrm{M_\odot}$. The gravitational softening length ϵ of each halo is decided following the criteria of Van den Bosch & Ogiya (2018) such that:

$$\epsilon = r_s f(c)\sqrt{\frac{0.32(N_p/1000)^{-0.8}}{1.12c^{1.26}}}, \qquad (B1)$$

where r_s is the scale radius of the initial NFW halo, c is the initial concentration, $f(c) = \ln(1+c) - c/(1+c)$, and N_p is the number of simulation particles. The haloes are evolved with self-interaction cross section $\sigma_m = 10\,\mathrm{cm^2\,g^{-1}}$ until a core is well developed in the centre. We emphasize that for the dimensionless $\tilde{\rho}_0$–$\tilde{t}$ space (Fig. 6) in which we compare the results, the mass, and the concentration of the halo or the cross-section has little impact on the results. The central density ρ_0 is defined as the average density of the innermost 100 particles.

This paper has been typeset from a TEX/LATEX file prepared by the author.