



Gravothermal Solutions of SIDM Halos: Mapping from Constant to Velocity-dependent Cross Section

Shengqi Yang¹, Xiaolong Du¹, Zhichao Carton Zeng^{2,3}, Andrew Benson¹, Fangzhou Jiang¹, Ethan O. Nadler^{1,4}, and Annika H. G. Peter^{2,3,5}

¹ Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA; syang@carnegiescience.edu

² Department of Physics, The Ohio State University, 191 W. Woodruff Ave., Columbus, OH 43210, USA

³ Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 W. Woodruff Ave., Columbus, OH 43210, USA

⁴ Department of Physics & Astronomy, University of Southern California, Los Angeles, CA 90007, USA

⁵ Department of Astronomy, The Ohio State University, 140 W. 18th Ave., Columbus, OH 43210, USA

Received 2022 May 26; revised 2023 February 17; accepted 2023 February 17; published 2023 March 28

Abstract

The scale-free gravothermal fluid formalism has long proved effective in describing the evolution of self-interacting dark matter halos with a constant dark matter particle cross section. However, whether the gravothermal fluid solutions match numerical simulations for velocity-dependent cross-section scenarios remains untested. In this work, we provide a fast mapping method that relates the constant-cross-section gravothermal solution to models with arbitrary velocity dependence in the cross section. We show that the gravothermal solutions after mapping are in good agreement with AREPO *N*-body simulation results. We illustrate the power of this approach by applying this fast mapping method to a halo hosting a low-surface-brightness galaxy, UGC 128. We show that this fast mapping method can be used to constrain free parameters in a physically motivated cross-section model and illustrate the parameter space favored by the rotation curve measurement.

Unified Astronomy Thesaurus concepts: Dark matter (353); Milky Way dark matter halo (1049)

1. Introduction

Self-interacting dark matter (SIDM) is a promising class of candidate to explain the cored dark matter (DM) halos implied by observations of local dwarf galaxies (de Blok et al. 2008; Oh et al. 2015), while maintaining the accurate reproduction of large-scale cosmic properties provided by the cold dark matter (CDM) paradigm (Peebles 1982; Planck Collaboration et al. 2011, 2014; Anderson et al. 2012). In the SIDM model, first proposed by Spergel & Steinhardt (2000), in addition to the gravitational interactions, DM particles also scatter with each other. The effect of scattering is negligible on large scales due to the low DM particle number density and the resulting long relaxation time. However, the collisional mean free path of DM particles can become smaller than the orbital Jeans scale near the halo center, causing the formation of an isothermal core and even the gravitational catastrophe (Lynden-Bell et al. 1968) at small scales.

The gravothermal fluid formalism originally developed for studying the gravitational catastrophe of stellar clusters (Lynden-Bell et al. 1968; Eggleton 1971) has been extended to isolated SIDM halos and provides rich physical insights into the nature of SIDM halo evolution (e.g., Lynden-Bell & Eggleton 1980; Balberg et al. 2002; Balberg & Shapiro 2002; Shapiro 2018; Essig et al. 2019; Nishikawa et al. 2020). Although it makes multiple simplifying assumptions, numerical solutions to the gravothermal fluid formalism show surprisingly good agreement with idealized *N*-body simulations and also match reasonably well with cosmological SIDM zoom-in simulations (Koda & Shapiro 2011; Elbert et al. 2015; Essig et al. 2019; Outmezguine et al. 2022). A nice property of the

gravothermal fluid formalism is that the fluid equations themselves can be rewritten into a form that is independent of the halo density and radius scales. Therefore, although computationally expensive, the fluid formalism only needs to be numerically solved once and can be rescaled to SIDM halos with arbitrary masses, sizes, and concentrations.

All previous comparisons between the gravothermal solutions and numerical simulations have assumed a constant DM particle scattering cross section. However, if SIDM is the explanation for the apparent deviations from CDM behavior on small scales, then recent observations indicate that the cross section, σ , for DM particle self-interaction is likely varying among systems of different scales. Specifically, σ is constrained to be of the order of $0.1 \text{ cm}^2 \text{ g}^{-1}$, $1 \text{ cm}^2 \text{ g}^{-1}$, and $10 \text{ cm}^2 \text{ g}^{-1}$ for galaxy cluster, galactic, and dwarf-galaxy halos, respectively, where the velocity dispersion of DM particles decreases from ~ 1000 to $\sim 1 \text{ km s}^{-1}$ (e.g., Kaplinghat et al. 2016; Elbert et al. 2018; Sagunski et al. 2021; Andrade et al. 2022). A velocity-dependent cross section $\sigma \propto v^{-4}$ at high velocity is also motivated by particle physics, assuming a Yukawa or Coulomb potential for the self-interaction (Feng et al. 2009; Ackerman et al. 2009; Ibe & Yu 2010; Kummer et al. 2018). Motivated by this, more recent SIDM numerical simulations have started to focus on velocity-dependent cross-section models (e.g., Nadler et al. 2020). Testing whether the gravothermal fluid solutions are still in good agreement with numerical simulations assuming a velocity-dependent cross-section model becomes crucial in developing better understanding of the SIDM *N*-body simulations, as well as justifying SIDM cross-section constraints derived by gravothermal fluid formalism-based methods.

In this work, we show using an analytic gravothermal fluid formalism that assuming different cross-section values/models effectively rescales SIDM halo evolution rates. In particular, there exists a one-on-one time mapping between any two



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

SIDM halos simulated by the gravothermal fluid approach such that their density, velocity, and all the other physical properties are identical up to some rescaling factors. Based on this finding, we propose a fast mapping method that transfers the gravothermal numerical solutions assuming a constant SIDM cross section to those with arbitrary velocity-dependent cross-section models. To test the performance of this mapping method we perform multiple idealized N -body simulations assuming velocity-dependent cross-section scenarios with AREPO (Springel 2010). We find excellent agreement between the mapped gravothermal solutions and N -body simulation results. This mapping approach is powerful because it eliminates the degrees of freedom introduced by assumptions in the SIDM cross-section model, such that one only needs to solve the gravothermal fluid equations once and one can then rescale the solution to halos of arbitrary sizes, masses, and cross-section models. Due to the high computational efficiency, this method is useful for exploring the continuous parameter space and constraining SIDM cross sections. It is also applicable to semianalytic galaxy formation models to efficiently generate SIDM halo populations. To demonstrate the utility of this approach, we combine the mapping method with rotation curve measurements of the low-surface-brightness (LSB) galaxy UGC 128 (van der Hulst et al. 1993) and illustrate the parameters of the Navarro–Frenk–White (NFW) density profile (Navarro et al. 1996) and velocity-dependent cross-section models favored by observations.

As this work was nearing completion, Outmezguine et al. (2022; hereafter O22) introduced a mapping method with a similar spirit, but which differs in detail from our work. We find that O22 only compared the mapped gravothermal solutions with idealized N -body simulations that assume constant cross sections (Koda & Shapiro 2011). Example applications of their mapping method are also not discussed in detail. We have therefore expanded this work to include comparison of our approach with that of O22.

Calibrating the gravothermal fluid formalism to cosmological zoom-in simulations is an even more interesting topic to explore. However, cosmological simulations generally contain more complicated physics and can break multiple assumptions made in the current fluid model. As a first step, in this work we will only compare gravothermal solutions with idealized SIDM N -body simulations for isolated halos with constant or velocity-dependent cross sections. We leave the more complicated comparisons of gravothermal solution versus cosmological simulation to future works.

The plan of this paper is as follows. In Section 2 we briefly introduce our idealized SIDM halo simulation sets. Section 3 is an introduction to the gravothermal fluid formalism. We briefly review the long-mean-free-path (lmfp) gravothermal solution mapping method among different constant-cross-section scenarios introduced by Balberg et al. (2002), Koda & Shapiro (2011), and Essig et al. (2019) in Section 4. We then explain the methodology of mapping the lmfp gravothermal solutions from constant cross section σ to velocity-dependent cross section $\sigma(v)$ in Section 5 and comparing the mapped gravothermal solutions with N -body simulation results as well as O22 in Section 6. We briefly discuss the generalization of this lmfp mapping method to the short-mean-free-path (smfp) and intermediate regime in Section 7. In Section 8 we combine the mapping method with observational data and show

parameter constraints of the velocity-dependent cross-section models. We conclude in Section 9.

2. Idealized SIDM N -body Simulation

In this work we perform idealized N -body simulations of an isolated Milky Way-sized halo with NFW initial density profile:

$$\rho(r) = \frac{\rho_s}{\frac{r}{r_s} \left(1 + \frac{r}{r_s}\right)^2}, \quad (1)$$

where $\rho_s = 4.2 \times 10^6 M_\odot \text{ kpc}^{-3}$ and $r_s = 24.54 \text{ kpc}$. The halo virial radius and mass, defined with overdensity $\Delta_{\text{vir}} = 99.2$ with respect to the critical density, are $r_{\text{vir}} = 280.6 \text{ kpc}$ and $M_{\text{vir}} = 10^{12.1} M_\odot$. To guarantee a finite halo mass, we apply an exponential cutoff to the halo density profile at r_{vir} . Convergence test results show that such an exponential cutoff at $r \gtrsim 10r_s$ has negligible influence on the gravothermal evolution of the SIDM halo. The properties of this halo are chosen to be consistent with those of the host halo from the CDM zoom-in simulation presented in Nadler et al. (2020), which has also been resimulated in velocity-dependent SIDM models.⁶

However, our selection of these halo properties is an arbitrary choice for testing the performance of the gravothermal solution mapping method introduced in Section 5. Since the gravothermal fluid formalism is scale-free (i.e., independent of ρ_s and r_s besides some rescaling factors), comparisons between the gravothermal solutions and isolated SIDM halo simulations of arbitrary sizes and masses are equivalent.

To avoid extremely long collapsing timescales for halo cores, and therefore expensive simulations, we select a large constant-cross-section factor $\sigma_c = 30 \text{ cm}^2 \text{ g}^{-1}$ and consider a simple velocity-dependent cross-section model introduced in Gilman et al. (2021),

$$\sigma(v) = \frac{\sigma_c}{(1 + v^2/\omega^2)^2}, \quad (2)$$

where ω is a characteristic velocity beyond which the cross section drops as v^{-4} . This asymptotic behavior is motivated by Coulomb or Yukawa interactions (Feng et al. 2009; Ackerman et al. 2009; Ibe & Yu 2010; Kummer et al. 2018). The factor 1 in the denominator is used to avoid a divergent cross section for low-velocity particles in the numerical simulations.

We consider four different cross-section models where $\omega = \infty$ (i.e., $\sigma = \sigma_c = 30 \text{ cm}^2 \text{ g}^{-1}$ is a constant) and $\omega = 1000, 500, \text{ and } 400 \text{ km s}^{-1}$. Notice that since the 1D rms velocity of DM particles at the halo center and at the moment of core formation (when the halo central density reaches minimum) is about 100 km s^{-1} , corresponding to a 3D rms relative velocity of about 250 km s^{-1} , those four scenarios correspond to constant cross section, and weak, intermediate, and strong cross-section velocity dependence, respectively.

We use SPHERIC (Garrison-Kimmel et al. 2013) to generate five initial conditions with identical resolutions and halo properties, but different random seeds. Through this we can quantify the effects of discreteness noise in the initial conditions on the evolution of the halo in the N -body simulations. We set the

⁶ This host halo was selected from the CDM cosmological simulation C1251024, and its corresponding zoom-in simulation was first presented in Mao et al. (2015).

particle mass to $10^6 M_\odot$ so the halo in each simulation is resolved with 1.7×10^6 particles. For each cross-section scenario we simulate five halos with the AREPO SIDM module, and we direct the reader to Vogelsberger et al. (2012, 2014, 2019) for a detailed introduction to the numerical method. This SIDM simulation pipeline has been adopted by many previous works (e.g., Zeng et al. 2022) and has been proved to generate valid simulation results for cross sections up to $60 \text{ cm}^2 \text{ g}^{-1}$.

3. Gravothermal Fluid Formalism

Consider a spherically symmetric and isolated halo. Assuming gravothermal equilibrium is achieved at every radius and every moment, the halo evolution can be described by the following four equations (see also Balberg et al. 2002; Essig et al. 2019):

$$\frac{\partial M}{\partial r} = 4\pi r^2 \rho, \quad (3)$$

$$\frac{\partial(\rho v_{\text{rms}}^2)}{\partial r} = -\frac{GM\rho}{r^2}, \quad (4)$$

$$\frac{L}{4\pi r^2} = -\kappa \frac{\partial T}{\partial r}, \quad (5)$$

$$\frac{\rho v_{\text{rms}}^2}{\gamma - 1} \left(\frac{\partial}{\partial t} \right)_M \ln \frac{v_{\text{rms}}^2}{\rho^{\gamma-1}} = -\frac{1}{4\pi r^2} \frac{\partial L}{\partial r}. \quad (6)$$

Throughout this paper we use $M(r)$ to denote halo enclosed mass at radius less than r . Then, $\rho(r)$, $v_{\text{rms}}(r)$, $L(r)$, $\kappa(r)$, and $T(r) = mv_{\text{rms}}^2(r)/k_B$ correspond to halo density, 1D rms velocity averaged over the Maxwell–Boltzmann (MB) distribution, luminosity, thermal conductivity, and temperature at radius r respectively. Here, m is the DM particle mass and k_B is the Boltzmann constant. Since we do not consider the movement of the halo center throughout this work, the rms velocity and velocity dispersion among DM particles are identical at all radii. Assuming the DM particles act as a monatomic ideal gas, we set the adiabatic index $\gamma = 5/3$. The Jeans equation (Equation (4)) forces gravothermal equilibrium where the outward pressure $\partial(\rho v_{\text{rms}}^2)/\partial r$ is balanced by the gravitational force. Equation (5) defines the heat flux and Equation (6) expresses the first law of thermodynamics. Together Equations (5) and (6) govern the quasi-static time evolution of an SIDM halo.

For ideal-gas environments where the heat conduction is solely realized by the thermal motion and collisions among gas particles, the kinetic theory of gases gives heat conductivity as

$$\kappa_{\text{smfp}} \propto \frac{n\lambda^2 k_B}{t_r}, \quad (7)$$

and it is conventional to define $\kappa_{\text{smfp}} = 2.1 v_{\text{rms}} k_B / \sigma$, derived from integration over the MB velocity distribution. Here, $n = \rho/m$ is the local DM particle number density, $\lambda = 1/(n\sigma)$ is the mean free path, and $t_r = \lambda/v_{12}$ is the relaxation time. Here we use σ to denote DM particle scattering cross section, which is either a constant or a velocity-dependent quantity. We

use v_{12} to distinguish particle relative velocity from the 1D rms velocity v_{rms} used for the temperature definition.

Equation (7) does not apply to gravothermal systems of low density and low self-interaction cross section, where most of the particles tend to orbit several times before colliding with one another. Using the Jeans length $H = \sqrt{v_{\text{rms}}^2/(4\pi G\rho)}$ to approximate particles' characteristic orbital radii, Equation (7) is only accurate when $\lambda \ll H$ (also referred as the smfp regime). In this work we only care about halo evolution processes that satisfy $\lambda \gg H$, also referred as the lmfp regime. To derive the heat conductivity in the lmfp regime one just needs to replace the mean free path λ with the scale height H (Spitzer 1987):

$$\kappa_{\text{lmfp}} \propto \frac{nH^2 k_B}{t_r}. \quad (8)$$

It is conventional to assume $\kappa_{\text{lmfp}} = (3\beta/2)nH^2 k_B/t_r = 0.27\beta n v_{\text{rms}}^3 \sigma k_B/(Gm)$, where β is an adjustable parameter to match the fluid numerical solutions with N -body simulations. The value of β cannot be derived from first principles and it is therefore treated as a free parameter to be calibrated against N -body simulations. It is known that calibration to N -body simulations results in β of order unity, but with some variation in the range 0.5–1.5 depending on the specific simulation and cross-section parameters (Essig et al. 2019). We will show in Section 6 that in the lmfp regime, it is valid to absorb constant factors resulting from integration over the MB velocity distributions into β when σ is constant. However, we will demonstrate that a more careful treatment is required for cases of relative velocity-dependent cross section $\sigma(v_{12})$.

A nice feature of this gravothermal fluid formalism is that, given the halo initial density profile as NFW with characteristic radius r_s and density ρ_s , the set of Equations (3)–(6) can be rewritten into the scale-free form such that the unitless gravothermal solutions can be applied to an arbitrary halo by rescaling. Specifically, all physical quantities in the equation set x in the fluid formalism can be separated into a unit factor x_0 and a scale-free factor \hat{x} (i.e., $x = x_0 \hat{x}$) such that Equations (3)–(6) can be written into the unitless form:

$$\begin{aligned} \frac{\partial \hat{M}}{\partial \hat{r}} &= \hat{\rho} \hat{r}^2, \\ \frac{\partial(\hat{\rho} \hat{v}_{\text{rms}}^2)}{\partial \hat{r}} &= -\frac{\hat{M} \hat{\rho}}{\hat{r}^2}, \\ \frac{\hat{L}}{\hat{r}^2} &= -\hat{\kappa} \frac{\partial \hat{v}_{\text{rms}}^2}{\partial \hat{r}}, \\ \hat{\rho} \hat{v}_{\text{rms}}^2 \left(\frac{\partial}{\partial \hat{t}} \right)_{\hat{M}} \ln \frac{\hat{v}_{\text{rms}}^2}{\hat{\rho}} &= -\frac{1}{\hat{r}^2} \frac{\partial \hat{L}}{\partial \hat{r}}. \end{aligned} \quad (9)$$

The unit variables are (Essig et al. 2019)

$$\begin{aligned} r_0 &= r_s, \quad \rho_0 = \rho_s, \quad M_0 = 4\pi r_s^3 \rho_s, \\ v_{\text{rms},0} &= \sqrt{4\pi G \rho_s r_s}, \quad t_0 = 1/\sqrt{4\pi G \rho_s}, \\ L_0 &= (4\pi)^{5/2} G^{3/2} \rho_s^{5/2} r_s^5, \quad (\sigma/m)_0 = 1/(\rho_s r_s), \end{aligned} \quad (10)$$

and it can be derived that $\hat{\kappa}_{\text{lmfp}} = 0.27 \times 4\pi \beta \hat{\rho} \hat{v}_{\text{rms}}^3 (\sigma \hat{m})$ and $\hat{\kappa}_{\text{smfp}} = 2.1 \hat{v}_{\text{rms}}/(\sigma \hat{m})$. Hereafter we will abbreviate $(\sigma \hat{m})$ as $\hat{\sigma}$ for simplicity. We assume $\hat{\kappa} = 1/(\hat{\kappa}_{\text{lmfp}}^{-1} + \hat{\kappa}_{\text{smfp}}^{-1})$ to join the lmfp and smfp regimes smoothly.

The steps to solve Equation (9) numerically are (Essig et al. 2019)

1. Determine input parameters β and $\hat{\sigma}$.
2. Grid the halo with NFW density profile $\hat{\rho} = (1/\hat{r})/(1 + \hat{r})^2$ into 150 log radial bins spread uniformly in the range $-2 \leq \log \hat{r} \leq 3$.^{7, 8} We assume boundary conditions $\hat{M} = \hat{L} = 0$ at the inner radius and $\hat{L} = 0$ at the outer radius to solve for the radial profile of \hat{v}_{rms} through the Jeans equation. Following Pollack et al. (2015) and Essig et al. (2019), we take values of \hat{M}_i and \hat{L}_i of the i th shell at the log radial grid \hat{r}_i , while $\hat{\rho}_i$, $\hat{v}_{\text{rms},i}$ are taken as the averaged values between the $(i - 1)$ th and the i th shells.
3. Let the fluid code take a small time step $d\hat{t}$ such that the specific energy $\hat{u} = 3\hat{v}_{\text{rms}}^2/2$ changes by a factor of no more than 10^{-3} among all the shells:

$$\frac{d\hat{u}_i}{d\hat{t}} = -\frac{\partial \hat{L}_i}{\partial \hat{M}_i}. \quad (11)$$

4. After taking one time step forward, perturb \hat{r} , $\hat{\rho}$, and \hat{v}_{rms} of each shell 10 times such that the gravothermal equilibrium (Equation (4)) is established again throughout the halo. During the perturbation, the mass, specific energy, and entropy $s = \ln(\hat{v}_{\text{rms}}^3/\hat{\rho})$ of each shell are conserved.
5. Repeat steps 2, 3, and 4 iteratively until core collapse.

We checked that the basic assumptions made in solving Equation (9), i.e., (i) the halo is spherically symmetric, and (ii) the halo mass and scale do not vary with time, are satisfied in the idealized N -body simulations for isolated halos. However, those assumptions may be not valid for cosmological simulations due to halo interactions, mergers, and mass accretion.

Notice that most of the current gravothermal fluid studies are interested in SIDM models with moderate cross section and β values near unity. Such a setup results in the halo remaining in the Imfp regime for most of its life until core collapse occurs. Moreover, σ decreases drastically with v_{rms} at $v_{\text{rms}} \gtrsim \omega$ under the scenario of a velocity-dependent cross section, leaving the heat conduction more likely to be dominated by the Imfp term. We therefore focus on the gravothermal solutions in the Imfp regime in this work. Specifically, we choose input parameters $\hat{\sigma} = 0.01$, $\beta = 0.6$ in step 1 and solve Equation (9) numerically to derive the gravothermal solutions, which will be used by the mapping method introduced in Sections 4–6. We stop the evolution of the fluid code when $\beta\hat{\sigma}\hat{t}$ reaches 173. At the last time step, the halo is still in the Imfp regime and has entered the core-collapsing stage, where its central density increases rapidly. The time evolution of the ratio between $\hat{\kappa}_{\text{Imfp}}$ and $\hat{\kappa}_{\text{smfp}}$ at the halo center is shown in Figure 1 by the blue solid curve.

We notice that the halos simulated in Section 2 adopt a much greater cross section of $\hat{\sigma} = 0.65$. To check the validity of the Imfp assumption we numerically solve Equation (9) for another set of input parameters, $\{\beta = 0.82, \hat{\sigma} = 0.65\}$, and present the time evolution of the halo central density as the red dashed

curve of Figure 1. This halo still evolves in the Imfp regime during most of its lifetime before core collapse, although the increase in cross section boosts the halo to be closer to the smfp regime since $\hat{\kappa}_{\text{Imfp}}/\hat{\kappa}_{\text{smfp}} \propto \beta\hat{\sigma}^2$. Ideally we would like to simulate a halo with smaller cross section to better match the assumption made in this work. However, this will delay the onset of core collapse and lead to more expensive and unstable idealized N -body simulations. We will show later that even for this large constant cross section $\hat{\sigma} = 0.65$ the Imfp assumption still holds, and the simulated halo evolution matches the Imfp gravothermal solution well. The smfp tension will be further alleviated when we consider the velocity-dependent cross-section scenarios, where the velocity dependence will largely decrease the effective cross section and bring the halo further away from the smfp regime.

The gravothermal evolution of SIDM halos described by Equation (9) is shown in Figure 2. Specifically, at the NFW initial state (shown by the red solid curves) where $\hat{\rho} = (1/\hat{r})/(1 + \hat{r})^2$, the halo temperature is highest at $\hat{r} \approx 0.8$. The heat therefore flows to both the halo center and outskirts. The halo center keeps absorbing heat until local thermal equilibrium is reached. At this point a cored region is formed where the density and the rms velocity are uniform, as shown by the green dashed curves. DM particles in the halo cored region continuously lose energy due to outward heat transfer. In this process their orbital scale decreases and the kinetic temperature increases. As shown by the blue dashed-dotted curves, the halo core density and temperature will keep increasing and will move further away from thermal equilibrium with the halo outskirts. Such a regime is also referred to as gravothermal catastrophe or core collapse.

While the gravothermal model is computationally much more efficient than running an N -body simulation it is still too slow to apply to large ensembles of halos as might be required in a semianalytic model of subhalo evolution, or when generating realizations of subhalo populations for analysis of gravitational lensing systems (e.g., Benson 2012; Gilman et al. 2021). For example, the calculations shown in this section required around 24 CPU hours to compute. While this is trivially manageable for a single halo, it is impractical to apply to ensembles of millions of halos. As such it is useful to be able to map a single solution of the gravothermal model to more general cases. We will explore this possibility in the remainder of this paper.

Additionally, in the Appendix we provide a triple power-law empirical model calibrated to this set of fully Imfp solutions with high precision. The advantage of this analytical model is that it captures the halo core size, core collapsing time, and the radius where the cored density profile joins smoothly to the halo NFW outskirts with simple formulas, and can be easily implemented into semianalytic models of galaxy formation.

4. Mapping Gravothermal Solutions between Constant Cross Sections

In this section we briefly review the mapping method discussed in Balberg et al. (2002), Koda & Shapiro (2011), and Essig et al. (2019) that transfers the gravothermal solutions under a constant cross section $\hat{\sigma}_1$ to another constant cross section $\hat{\sigma}_2$, assuming the halo is mostly evolving in the Imfp regime.

The third and fourth equations in Equation (9) show that the halo evolution rate is degenerated with radius-independent

⁷ In this work log denotes a base-10 logarithm and ln denotes a natural logarithm.

⁸ We have tested that setting the halo outer bound at $\log \hat{r} = 1, 2, 3$ or the inner bound at $\log \hat{r} = -3, -2$ has negligible influence on the fluid solutions.

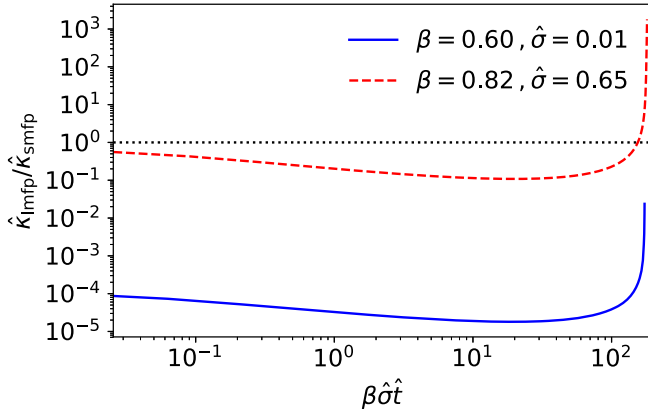


Figure 1. Time evolution of the ratio between the lmfp and the smfp heat conductivities at the halo center, assuming two different sets of parameters $\{\hat{\sigma} = 0.01, \beta = 0.6\}$ (blue solid) and $\{\hat{\sigma} = 0.65, \beta = 0.82\}$ (red dashed). For the $\{\hat{\sigma} = 0.01, \beta = 0.6\}$ set of gravothermal solutions $\hat{\kappa}_{lmfp} \ll \hat{\kappa}_{smfp}$ throughout the halo evolution process, so the total heat conductivity is always dominated by the lmfp term. In the $\{\hat{\sigma} = 0.65, \beta = 0.82\}$ case the halo also evolves mostly in the lmfp regime, but it enters the smfp regime at $\beta\hat{\sigma}\hat{t} = 155$.

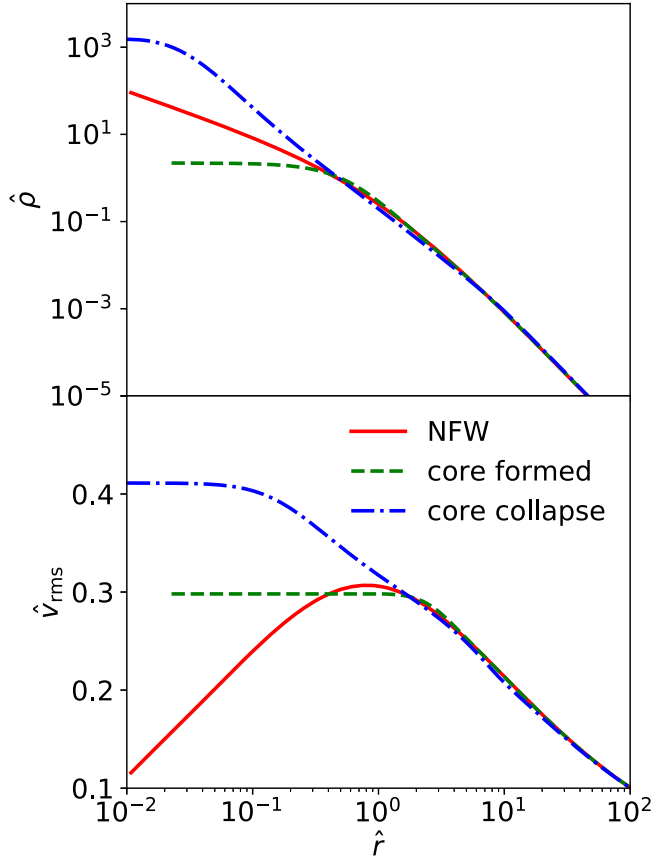


Figure 2. Gravothermal evolution phases of an SIDM halo described by the fluid formalism. The top/bottom panel shows the halo density/rms velocity profiles at different evolution stages. The red solid curves show the NFW initial condition at $\hat{t} = 0$. The green dashed curves present the moment when the halo central density reaches minimum, and an isothermal core is formed at the halo center. The blue dashed-dotted curves show the subsequent process of core collapse.

factors in the heat conductivity $\hat{\kappa}$. In the lmfp regime $\hat{\kappa}_{lmfp} \propto \beta\hat{\sigma}$, therefore $\beta\hat{\sigma}$ and the inverse of the halo evolution time \hat{t} are degenerate such that the time evolution of the gravothermal solution can be parameterized by $\beta\hat{\sigma}\hat{t}$. In other

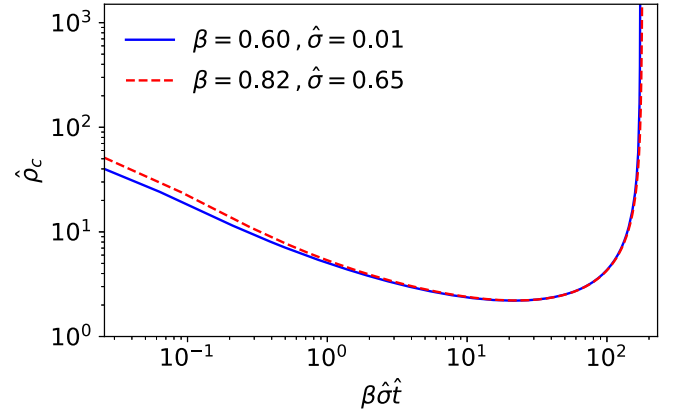


Figure 3. Time evolution of the central density of the SIDM halo given by the gravothermal fluid formalism at parameters $\{\beta = 0.6, \hat{\sigma} = 0.01\}$ (blue solid) and $\{\beta = 0.82, \hat{\sigma} = 0.65\}$ (red dashed). The halo evolution is almost identical with time parameter $\beta\hat{\sigma}\hat{t}$ when κ is dominated by the lmfp term.

words, there exists a one-to-one mapping $\beta_a \hat{\sigma}_a \hat{t}_a \leftrightarrow \beta_b \hat{\sigma}_b \hat{t}_b$ between two sets of gravothermal solutions with parameters $\{\beta_a, \hat{\sigma}_a\}$ and $\{\beta_b, \hat{\sigma}_b\}$ such that

$$\hat{x}_a(\hat{r}, \hat{t}_a) = \hat{x}_b\left(\hat{r}, \hat{t}_a \times \frac{\beta_a \hat{\sigma}_a}{\beta_b \hat{\sigma}_b}\right), \quad (12)$$

where $\hat{x}(\hat{r}, \hat{t})$ stands for an arbitrary scale-free halo property at radius \hat{r} and time \hat{t} , including $\hat{\rho}$, \hat{v}_{rms} , \hat{M} , \hat{L} , etc. We show the scale-free halo central density $\hat{\rho}_c$ versus $\beta\hat{\sigma}\hat{t}$ for two sets of gravothermal solutions with very different input parameters in Figure 3. The time evolution of these two halos is almost identical after the first few time steps, where $\hat{\kappa}_{lmfp}$ is comparable to $\hat{\kappa}_{smfp}$ for the $\hat{\sigma} = 0.65$ case, and the lmfp assumption is not very accurate (see Figure 1).

Although the gravothermal fluid formalism considers a very simplified picture, it provides numerical solutions in surprisingly good agreement with idealized N -body simulations of isolated halos up to a slightly varying β factor. As an example, in Figure 4 we present comparisons of the evolution of the halo central density over time between the gravothermal solutions (red solid curve) and the averaged Arepo N -body simulations (black solid curve) for a Milky Way-sized SIDM halo with $\rho_s = 4.2 \times 10^6 M_\odot \text{ kpc}^{-3}$, $r_s = 24.54 \text{ kpc}$, and $\sigma/m = 30 \text{ cm}^2 \text{ g}^{-1}$ (see Section 2 for simulation details). To test convergence, we reduce the particle number by a factor of 4 in each N -body realization and present the averaged simulation results as the cyan dashed curve. We find the rescaled lmfp gravothermal solution with $\beta = 0.82$ captures the overall halo evolution and reproduces the core collapsing time in the Arepo simulation results. We notice that the fluid numerical solution has a halo central density that drops faster than the N -body results prior to isothermal core formation ($0 \leq t/\text{Gyr} \leq 2$). The central density also grows slightly faster than N -body at $t \gtrsim 18 \text{ Gyr}$. Those differences are, again, caused by the fact that we have selected a very large cross section for the N -body simulations. The basic assumption of this mapping method, that the halo is always in the lmfp regime, is not of high accuracy. Since our focus is to extend the lmfp gravothermal fluid formalism to the velocity-dependent cross-section scenario and capture the overall halo evolution in the N -body simulations, in this work we will accept this slight disagreement. Moreover, the smfp problem is

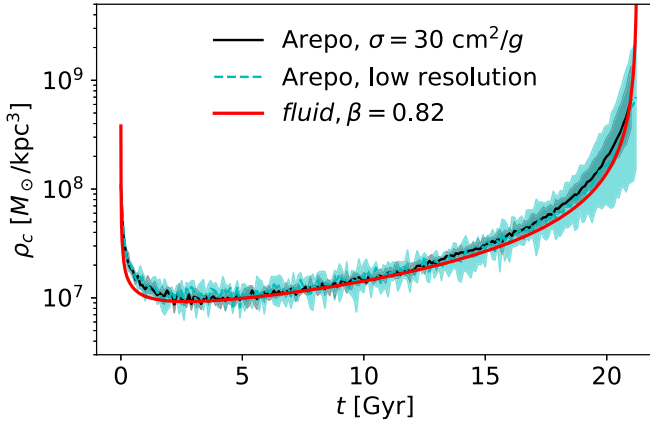


Figure 4. Comparisons of halo central density evolution between the averaged results of the Arepo N -body simulation (black solid curve) and the numerical solutions of the gravothermal fluid formalism (red solid curve). The gray band encloses the maximum and minimum halo central density among the five AREPO realizations. The cyan band shows convergence test results where the particle number in each N -body realization is decreased from 1.7×10^6 to 4.2×10^5 .

alleviated when we simulate velocity-dependent cross-section models.

5. Mapping Gravothermal Solutions from Constant to Velocity-dependent Cross Section

In this section we introduce a mapping method to transfer the gravothermal solutions under a constant cross section $\hat{\sigma}$ to those under a velocity-dependent cross-section model $\sigma(v)$. In numerical simulations the velocity used to determine the scattering probability between two particles is generally the particle relative velocity v_{12} , but in this section we will consider a simpler case where the cross section is determined by the particle rms velocity. We will discuss this mapping method under the more practical $\sigma(v_{12})$ scenario in Section 6.

In this paper we will use the cross-section model introduced by Equation (2) as an example, but this mapping method is applicable to an arbitrary $\sigma(v)$ model.

In Section 4 we have shown that there exists a one-to-one mapping $\beta_a \hat{\sigma}_a \hat{t}_a \leftrightarrow \beta_b \hat{\sigma}_b \hat{t}_b$ between two sets of gravothermal solutions with different β and constant cross sections. Now consider two sets of gravothermal solutions, one with constant cross section $\{\beta_a, \hat{\sigma}_a\}$ and the other with a velocity-dependent cross section $\{\beta_b, \hat{\sigma}_b(\hat{v}_{\text{rms}, b})\}$. The linear time mapping does not hold anymore since $\hat{\sigma}_b(\hat{v}_{\text{rms}, b})$ is a function of radius and time rather than a constant. However, if in every small time interval we may assume a characteristic halo-wide cross section (or, equivalently, a characteristic velocity) such that we can drop the radial dependence of the cross section $\hat{\sigma}(\hat{v}_{\text{rms}, b}(\hat{r}, \hat{t}_b)) \approx \hat{\sigma}(\hat{v}_{\text{rms}, b}^c(\hat{t}_b))$, then there exists the following one-to-one mapping for every small time step:

$$\begin{aligned} \beta_a \hat{\sigma}_a d\hat{t}_a &= \beta_b \hat{\sigma}_b(\hat{v}_{\text{rms}, b}^c(\hat{t}_b)) d\hat{t}_b \\ &= \beta_b \hat{\sigma}_b(\hat{v}_{\text{rms}, a}^c(\hat{t}_a)) d\hat{t}_b. \end{aligned} \quad (13)$$

Here the second equation holds because alternating a time-dependent yet radius-independent SIDM cross section has no influence on the gravothermal fluid solution besides nonlinearly stretching the time axis. In other words, there still exists the following one-to-one mapping between these two sets of

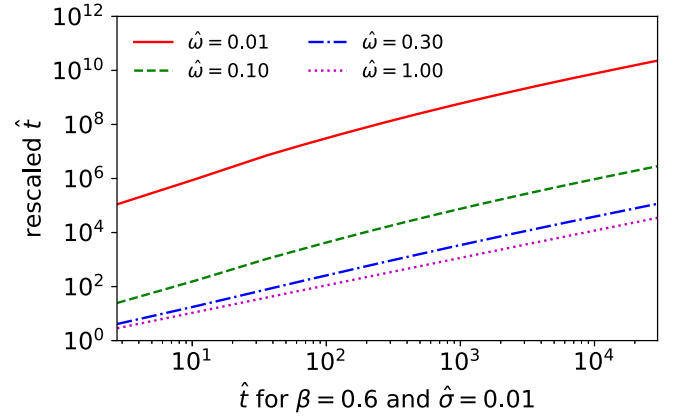


Figure 5. Mapping relation for the time variable in the gravothermal solution. The x -axis present the scale-free time \hat{t} of the gravothermal solution under constant cross section $\hat{\sigma} = 0.01$. We map \hat{t} to the velocity-dependent cross-section model $\hat{\sigma} = \hat{\sigma}_c / (1 + \hat{v}_{\text{rms}}^2 / \hat{\omega}^2)^2$, with $\hat{\sigma}_c = 0.01$ and $\hat{\omega} = 0.01, 0.1, 0.3$, and 1.0 . The rescaled time is presented on the y -axis for each case. For simplicity we have fixed $\beta = 0.6$ in this test.

gravothermal solutions:

$$\begin{aligned} \hat{x}_a(\hat{r}, \hat{t}_a) &= \hat{x}_b(\hat{r}, \hat{t}_b) \\ &= \hat{x}_b\left(\hat{r}, \int_0^{\hat{t}_a} \frac{\beta_a \hat{\sigma}_a}{\beta_b \hat{\sigma}_b(\hat{v}_{\text{rms}, a}^c(\hat{t}'_a))} d\hat{t}'_a\right). \end{aligned} \quad (14)$$

In practice one has the constant-cross-section gravothermal solution that specifies the radial profiles of halo density and rms velocity at an array of time $\hat{t}_a = \{\hat{t}_a^0, \hat{t}_a^1, \dots, \hat{t}_a^N\}$ (here \hat{t}_a^i is the time of the i th step), and aims to map it to some velocity-dependent cross-section scenario. To achieve this one will need to first compute the time step $\delta\hat{t}_a = \{\hat{t}_a^0, \hat{t}_a^1 - \hat{t}_a^0, \dots, \hat{t}_a^N - \hat{t}_a^{N-1}\}$, and then use Equation (13) to compute the mapped time step $\delta\hat{t}_b = [\beta_a \hat{\sigma}_a / \beta_b \hat{\sigma}_b(\hat{v}_{\text{rms}, a}^c(\hat{t}_a))] \delta\hat{t}_a$. Summing $\delta\hat{t}_b$ up, a new time axis \hat{t}_b is derived such that $\hat{x}_b(\hat{r}, \hat{t}_b) = \hat{x}_a(\hat{r}, \hat{t}_a)$.

The remaining problem is to determine how to choose the characteristic rms velocity \hat{v}_{rms}^c at each time. Since the SIDM core evolves the most with time, while the halo outskirts show very little time evolution (see Figure 2), we find that simply using the halo central rms velocity shows very good mapping performance:

$$\hat{v}_{\text{rms}}^c(\hat{t}) = \hat{v}_{\text{rms}}(\hat{r} = \hat{r}_{\text{min}}, \hat{t}). \quad (15)$$

Here $\hat{r}_{\text{min}} = 10^{-2}$ is the innermost radial bin of the fluid formalism.

To test this mapping method, we assume $\hat{\sigma}_c = 0.01$, $\beta = 0.6$, and numerically solve Equation (9) for two cross-section models. For the first test we assume a constant cross section $\hat{\sigma} = \hat{\sigma}_c$, while for the second test we assume $\hat{\sigma} = \hat{\sigma}_c / (1 + \hat{v}_{\text{rms}}^2 / \hat{\omega}^2)^2$. We consider four velocity-dependent cross-section models with $\hat{\omega} = 0.01, 0.1, 0.3$, and 1.0 , corresponding to very strong, strong, intermediate, and weak velocity dependence. Figure 5 presents the unitless time of the constant-cross-section gravothermal solution rescaled by this mapping method. Notice that the 1D velocity dispersion of dark matter particles \hat{v}_{rms} varies from 0.1 to 0.4 near the halo center (see Figure 2 bottom panel), which is small compared to 1.0 . The halo SIDM cross section for the $\hat{\omega} = 1.0$ case is therefore almost a constant $\hat{\sigma} = \hat{\sigma}_c$, corresponding to a nearly trivial time mapping relation shown by the magenta dotted curve.

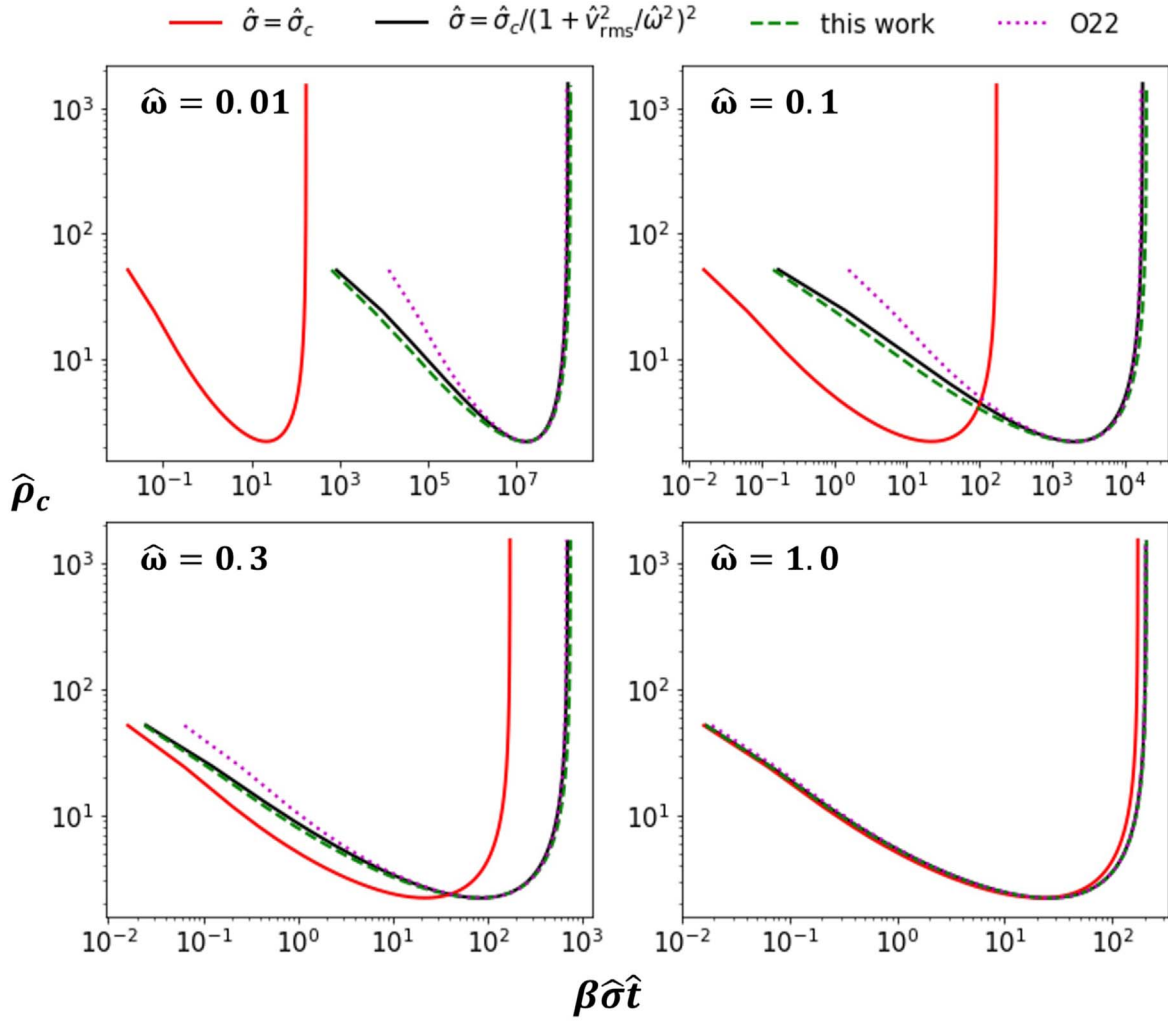


Figure 6. Scale-free evolution of the SIDM halo central density under a constant-cross-section model (red solid) and a velocity-dependent cross-section model (black solid). In each panel the rescaled constant-cross-section gravothermal solutions given by the methods introduced in Section 5 and O22 are shown by the green dashed curve and the magenta dotted curve, respectively. We consider four different cross-section models with $\hat{\omega} = 0.01, 0.1, 0.3$, and 1.0 and compare the numerical solution with the rescaled results in the top left, top right, bottom left, and bottom right panels. Both mapping methods show reasonable agreement with the gravothermal numerical solutions. We notice that O22 is slightly more accurate in mapping the core collapsing time, while this work more accurately reproduces the halo evolution before its central density reaches a minimum.

Decreasing $\hat{\omega}$ suppresses the total cross-section value $\hat{\sigma}$ and also boosts the velocity dependence of the particle scattering probabilities. As a result, the rescaled \hat{t} enlarges for the lower $\hat{\omega}$ cases, and the time mapping nonlinearity becomes more significant.

Comparisons between the mapped constant-cross-section gravothermal solutions (green dashed lines) and the actual velocity-dependent gravothermal solutions (black solid curves) are shown in Figure 6. We find this simple remapping method can successfully reproduce the gravothermal solutions in velocity-dependent cross-section models in all the cases, although it overestimates the core collapsing time by 14%, 13%, 7.5%, and 1.3%, respectively. This overestimation is caused by Equation (15), where we set the halo central rms velocity as a characteristic velocity that determines the overall DM particle cross section at every moment. A true characteristic v_{rms} is generally lower than the central velocity, and therefore our method slightly underestimates the effective cross section at each time and will further overestimate the core collapsing time.

We also check the performance of the mapping method introduced in O22. Instead of using a time-varying characteristic rms velocity \hat{v}_{rms}^c , O22 makes a more simplified assumption to use the rms velocity at the halo center and at the instance of maximum core size as the characteristic value. Effectively the characteristic rms velocity is assumed to be

$$\hat{v}_{\text{rms}}^c = \hat{v}_{\text{rms}}(\hat{r} = \hat{r}_{\text{min}}, \hat{t} = \hat{t}_{\hat{\rho}_{c,\text{min}}}). \quad (16)$$

We show the time evolution of the halo central density mapped by O22 as magenta dotted curves in Figure 6. We show that the O22 method only underestimates the core collapsing time by 6%, 5%, 2.3%, and 0.3% for the $\hat{\omega} = 0.01, 0.1, 0.3$, and 1.0 scenarios, but it cannot correctly capture the halo evolution prior to the instance of maximum core size.

The accuracy of our mapping method can be further boosted if we compute the characteristic \hat{v}_{rms}^c , or, equivalently, the characteristic cross section, at each time step more carefully. Specifically, at every time step we can define the characteristic cross section as the averaged cross section weighted by the

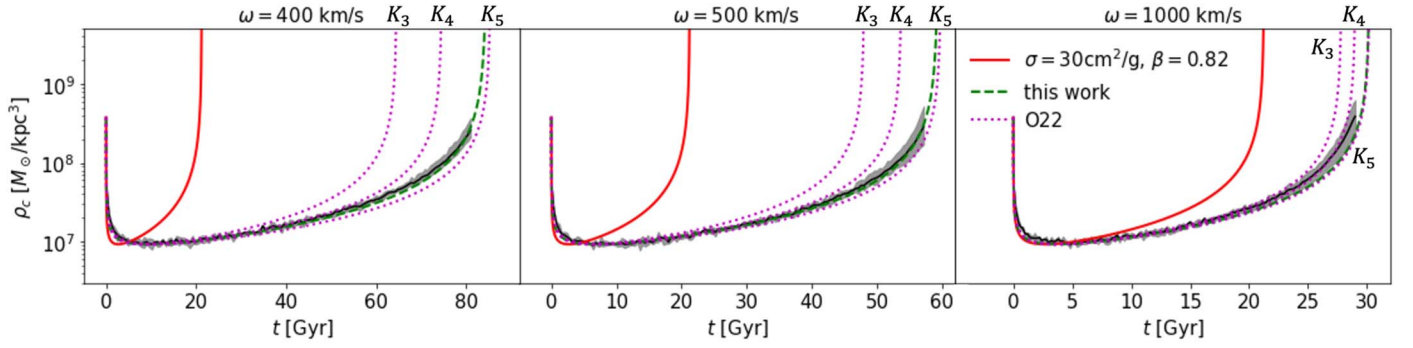


Figure 7. Comparisons of the time evolution of the halo central density between the averaged AREPO N -body simulation results (black solid curve) and the mapped gravothermal solutions (green dashed curves) in different velocity-dependent cross-section models. In each panel, the maximum and minimum halo central density among the five AREPO realizations are enclosed by the gray band. The red curve shows the gravothermal solution that matches with N -body assuming a constant cross section $\sigma = 30 \text{ cm}^2 \text{ g}^{-1}$, which is identical to the red curve shown in Figure 2. We use $\beta = 0.82$ as calibrated in Figure 4, and assume it does not vary with ω . We find the mapping method introduced in this work matches with AREPO N -body simulation results extremely well. It outperforms the $\langle \kappa_{\text{lmfp}} \rangle \propto K_3$ definition, which is assumed as an ansatz in O22. We also find under a linear time rescaling that the definition of the averaged heat conductivity $\langle \kappa_{\text{lmfp}} \rangle \propto K_5$ performs equally as well as Equation (21).

mass of multiple shells:

$$\hat{\sigma}^c = \frac{\sum_{i=0}^{i_{\text{out}}} 4\pi \hat{r}_i^2 \delta \hat{r}_i \hat{\rho}_i \hat{\sigma}_{b,c}(\hat{v}_{\text{rms},i})}{\sum_{i=0}^{i_{\text{out}}} 4\pi \hat{r}_i^2 \delta \hat{r}_i \hat{\rho}_i}, \quad (17)$$

where i is the shell index, and i_{out} corresponds to the shell that joins smoothly to the halo NFW outskirts, beyond which the DM particle self-interaction is negligible. The radius of shell i_{out} is given by Equation (A4) and Equation (A8) in the Appendix. After this step we can map the time step of the constant-cross-section gravothermal solution as $\delta \hat{t}_b = [\beta_a \hat{\sigma}_a / \beta_b \hat{\sigma}_b^c(\hat{t}_a)] \delta \hat{t}_a$. We find that this more careful way of computing the characteristic cross section effectively decreases the fractional error of the mapping method in predicting the core collapsing moment to 6.4%, 5.6%, 2.9%, and 0.5% for $\hat{\omega} = 0.01, 0.1, 0.3$, and 1.0 cases, providing performance equally as good as O22 while maintaining a good match to the actual gravothermal numerical solutions before the instance of maximum core size. However, we do not use this more complicated treatment as the default method because it can be computationally expensive when applied to a large number of halos. Furthermore, in practice the mapping error in the core collapsing time of $\sim 10\%$ at most is small compared to the uncertainties in the β factor.

6. Calibrating the Mapping Method to N -body Simulation

To calibrate the gravothermal solution remapping method introduced in Section 5 to N -body simulations, we consider cross-section models that are determined by the relative velocity v_{12} between two scattering particles. Such velocity-dependent models will influence the heat conductivity averaged over the MB velocity distribution $\langle \kappa_{\text{lmfp}} \rangle$, resulting in additional complexity.

It is worth recalling how $\langle \kappa_{\text{lmfp}} \rangle$ is computed for the case of constant cross section. By convention,

$$\kappa_{\text{lmfp}}(v_{12}) = \frac{3}{2} \beta n H^2 \frac{k_B}{t_r} = \frac{3\beta k_B}{8\pi G} \frac{n\sigma}{m} v_{\text{rms}}^2 v_{12}, \quad (18)$$

and the relative velocity can be defined as $v_{12} = \sqrt{v_1^2 + v_2^2 - 2v_1 v_2 \cos \theta_{12}}$, where θ_{12} is the angle between the velocity vectors \mathbf{v}_1 of particle 1 and \mathbf{v}_2 of particle 2. To

compute the averaged heat conductivity over the MB velocity distribution, previous works compute the averaged v_{12} . Specifically, let us define the MB distribution and the averaged v_{12} as

$$\begin{aligned} f_{\text{MB}}(v) d^3v &= f_{\text{MB}}(v) v^2 dv \sin(\theta) d\theta d\phi = v^2 \left(\frac{m}{2\pi k_B T} \right)^{3/2} \\ &\times \exp\left(-\frac{mv^2}{2k_B T}\right) \sin(\theta) dv d\theta d\phi, \\ \langle v_{12} \rangle &= \int_0^\pi \sin \theta_1 d\theta_1 \int_0^{2\pi} d\phi_1 \int_0^{2\pi} d\phi_{12} \\ &\times \int_0^\infty v_1^2 f_{\text{MB}}(v_1) dv_1 \\ &\times \int_0^\infty v_2^2 f_{\text{MB}}(v_2) dv_2 \int_0^\pi v_{12} \sin(\theta_{12}) d\theta_{12} \\ &= \frac{4}{\sqrt{\pi}} \sqrt{\frac{k_B T}{m}} = \frac{4}{\sqrt{\pi}} v_{\text{rms}} = 2.26 v_{\text{rms}}. \end{aligned} \quad (19)$$

Using this result for $\langle v_{12} \rangle$ in Equation (18) gives the definition assuming a constant DM particle cross section:

$$\langle \kappa_{\text{lmfp}} \rangle = \frac{3\beta k_B}{8\pi G} \frac{n\sigma}{m} v_{\text{rms}}^2 \langle v_{12} \rangle = 0.27 \beta n v_{\text{rms}}^3 \sigma k_B / (Gm). \quad (20)$$

One may argue that a better definition of the averaged heat conductivity might be $\langle \kappa_{\text{lmfp}} \rangle \propto \langle v_1 v_2 v_{12} \rangle$, or $\langle v_{12}^3 \rangle$, or $\langle v^n \rangle / v_{\text{rms}}^{n-3}$, where n is an arbitrary number. The specific choice does not matter for constant-cross-section scenarios because those different integrations will only cause a difference of a constant factor in Equation (20), which can be absorbed into the free parameter β . This is no longer true if a velocity-dependent cross section is assumed. Specifically, the definition $\langle \kappa_{\text{lmfp}} \rangle \propto \langle v^n \sigma(v_{12}) \rangle / (v_{\text{rms}}^{n-3} \sigma_c)$ results in $\langle v^n \rangle / v_{\text{rms}}^{n-3}$ for $\omega \rightarrow \infty$. However, with finite ω the v^n term serves as a weight in the integration. Larger n biases the integration over the MB distribution toward higher velocities, corresponding to a smaller effective cross section and longer halo core collapse timescale. Therefore it is necessary to define $\langle \kappa_{\text{lmfp}} \rangle$ in a physically motivated way that matches N -body results.

Notice that the heat flux is determined by the product of the conductivity and the temperature gradient: $H\partial T/\partial r \sim \Delta T \sim \Delta E$ is the characteristic temperature variation and energy transferred within a particle's orbit. Here temperature is also a statistical property averaged over the MB distribution $T \propto \langle v_{12}^2 \rangle$, so the energy transfer rate is proportional to $v_{12}^3 \sigma(v_{12})$ (Colquhoun et al. 2021). We therefore define the averaged heat conductivity as

$$\langle \kappa_{\text{lmfp}} \rangle = 0.27 \beta n v_{\text{rms}}^3 k_B / (Gm) \frac{\langle v_1 v_2 v_{12}^3 \sigma(v_{12}) \rangle}{\langle v_1 v_2 v_{12}^3 \rangle}. \quad (21)$$

Here the $v_1 v_2$ term in the numerator comes from the scale height H^2 . One v_{12} term is contributed by $1/t_r$, and the other v_{12}^2 term comes from the temperature gradient. The numerator is defined such that Equation (21) reduces to Equation (20) at $\omega \rightarrow \infty$. This definition is slightly different from the ansatz assumed in O22, where $\langle \kappa_{\text{lmfp}} \rangle \propto \langle v_{12}^3 \sigma(v_{12}) \rangle$. We notice that the triple integral $\langle v_1 v_2 v_{12}^3 \sigma(v_{12}) \rangle$ needs to be computed for every time interval, which significantly slows down the mapping calculations. However, it can be shown that the integration results are only sensitive to $\omega/\sqrt{T} \propto \omega/v_{\text{rms}}$ times $\beta \sigma_c$, we therefore create a lookup table for the integration result spanning the parameter range $0.01 \leq \omega/v_{\text{rms}} \leq 100$ and use the linearly interpolated integration results to achieve fast mapping.

To map a set of constant-cross-section gravothermal solutions with parameters $\{\beta_a, \hat{\sigma}_{c,a}\}$ to those assuming a velocity-dependent cross-section model with parameters $\{\beta_b, \hat{\sigma}_{c,b}, \hat{\omega}\}$, we perform the following steps:

1. Compute the time step $\delta \hat{t}_a = \{\hat{t}_a^0, \hat{t}_a^1 - \hat{t}_a^0, \dots, \hat{t}_a^N - \hat{t}_a^{N-1}\}$.
2. Compute the mapped time step $\delta \hat{t}_b = [\beta_a \hat{\sigma}_{c,a} / \beta_b] [\langle v_1 v_2 v_{12}^3 \rangle_{\hat{t}_a} / \langle v_1 v_2 v_{12}^3 \hat{\sigma}_b(v_{12}) \rangle_{\hat{t}_a}] \delta \hat{t}_a$. Here $\langle x \rangle_{\hat{t}_a}$ implies computing the averaged x over an MB distribution where the temperature is given by the constant-cross-section gravothermal solutions at the innermost radius and time \hat{t}_a . We choose the halo central temperature as a characteristic value at every time step with the same argument as discussed in Section 5.
3. Sum over all $\delta \hat{t}_b$ to obtain a new time axis such that $\hat{x}_b(\hat{t}_b) = \hat{x}_a(\hat{t}_a)$.

Assuming that in the idealized N -body simulation β does not vary with ω , we compare the mapped gravothermal solutions and AREPO N -body simulations in velocity-dependent cross-section models introduced by Equation (2) with different ω . We find excellent matches, as presented in Figure 7. We notice that this comparison is only performed in a limited $\omega \geq 400 \text{ km s}^{-1}$ range for this Milky Way-sized halo. Ideally, we would extend this comparison to cross-section models with stronger velocity dependence, i.e., with even lower characteristic velocity ω . However, a halo collapses more slowly as ω decreases, which makes the N -body simulation less stable and computationally more expensive. Furthermore, for the Milky Way-sized halo we simulate in this work $\omega \leq 400 \text{ km s}^{-1}$ is a less interesting scenario, where the halo collapsing time is much longer than the age of the universe.

We also show the gravothermal solution mapped by O22 by the magenta dotted curves in Figure 7. Specifically, O22

suggests to define the averaged heat conductivity as

$$\langle \kappa_{\text{lmfp}} \rangle \propto K_n = \frac{\langle v_{12}^n \sigma(v_{12}) \rangle}{\langle v_{12}^n \rangle} \bigg|_{r_{\text{min}}, t_{\text{pc, min}}}, \quad n = 3. \quad (22)$$

It is emphasized in O22 that the assumption $\langle \kappa_{\text{lmfp}} \rangle \propto K_3$ should be further tested by N -body simulations. We find this assumption overestimates the effective cross section as well as the heat conductivity, leading to a faster core collapse than the N -body results. However, we find assuming $\langle \kappa_{\text{lmfp}} \rangle \propto K_5$ and a linear time rescaling gives mapping performance as precise as Equation (21). This good match is also proved independently by Yang & Yu (2022) through N -body simulations.

7. Extend the lmfp Mapping Method to More General Cases

The lmfp gravothermal solution mapping method introduced in Sections 4–6 can be easily generalized to the smfp regime as well as to the intermediate regime. However, in this work we do not test the validity of the smfp and intermediate-regime mapping methods introduced in this section due to difficulties in both computation and physics. Neither will we apply these mappings in the following sections. From the computational aspect, we find the gravothermal fluid formalism introduced in Section 3 runs into numerical issues when solving systems with very large cross section. Furthermore, the AREPO idealized SIDM simulation pipeline introduced in Section 2 is not tested for isolated halos with cross section greater than $60 \text{ cm}^2 \text{ g}^{-1}$. From the aspect of physics, a very large SIDM cross section may change the formation of structure from that of CDM. The initial density profile of the SIDM halo with large cross section is also uncertain and unlikely to be NFW (Ahn & Shapiro 2003).

Let us first revisit Equation (9) in the fully smfp regime. Now, since we switch focus from $\hat{\kappa}_{\text{lmfp}} \propto \beta \hat{\sigma}$ to $\hat{\kappa}_{\text{smfp}} \propto 1/\hat{\sigma}$, it becomes clear that the time mapping relation changes from $\beta_a \hat{\sigma}_a \hat{t}_a \leftrightarrow \beta_b \hat{\sigma}_b \hat{t}_b$ to $\hat{t}_a/\hat{\sigma}_a \leftrightarrow \hat{t}_b/\hat{\sigma}_b$ for the models with constant SIDM cross section. For the relative velocity-dependent cross-section scenarios one only needs to modify the time interval mapping as

$$\delta \hat{t}_b = \frac{\langle v_{12}^3 \rangle_{\hat{t}_a} / \hat{\sigma}_{c,a}}{\langle v_{12}^3 / \hat{\sigma}_b(v_{12}) \rangle_{\hat{t}_a}} \delta \hat{t}_a, \quad (23)$$

to map gravothermal solutions computed under constant cross sections $\hat{\sigma}_{c,a}$ to those assuming velocity-dependent cross-section models $\hat{\sigma}_b(v_{12})$. Here one v_{12} term in the MB integration comes from the mean free path λ and another v_{12}^2 term comes from the energy transfer as has been discussed in Section 6. The two key differences between the lmfp and smfp mapping methods are: (1) The smfp heat conductivity and therefore the gravothermal solutions are no longer dependent on the free parameter β . (2) In the lmfp regime a halo collapses faster with a larger cross section, but the situation is reversed in the smfp regime.

To consider the intermediate regime where the mean free path and the scale height are comparable, the time mapping relation among constant-cross-section models has the form $\hat{\kappa}_a \hat{t}_a \leftrightarrow \hat{\kappa}_b \hat{t}_b$, where $\hat{\kappa} = 1/(\hat{\kappa}_{\text{lmfp}}^{-1} + \hat{\kappa}_{\text{smfp}}^{-1})$. For the relative velocity-dependent cross-section scenarios the time interval

mapping relation is

$$\delta \hat{t}_b = \frac{\langle \hat{\kappa}_a v_{12}^2 \rangle_{\hat{t}_a}}{\langle \hat{\kappa}_b v_{12}^2 \rangle_{\hat{t}_a}} \delta \hat{t}_a, \quad (24)$$

where the v_{12} terms in the MB integration again come from the energy transfer. The challenge of computing Equation (24) compared to the asymptotic lmf and smf scenarios is that constant parameters β , $\hat{\sigma}_c$, as well as the scale-free density $\hat{\rho}$ can no longer be pulled out from the integration due to this more complicated expression for heat conductivity. It is therefore necessary to prepare a 4D lookup table that simultaneously covers the variations of $\{\beta, \hat{\sigma}_c, \hat{\omega}/\hat{v}_{\text{rms}}, \hat{\rho}\}$ to achieve fast mapping calculations.

8. Observational Constraints on the Cross-section Model

Both the gravothermal fluid formalism and idealized N -body simulations are computationally expensive in the sense that they cannot be directly used to explore the continuous SIDM particle parameter space. The strength of the lmf mapping method introduced in this work is that one needs to numerically solve for the gravothermal fluid equations only once, and can then extend this solution to halos with arbitrary cross-section models, as long as the halo is mostly evolving in the lmf regime. It therefore serves as a powerful tool for deriving constraints on SIDM cross-section model parameters.

Halos hosting dwarfs and LSB galaxies are promising laboratories for constraining SIDM models because these small galaxies serve as powerful tracers of the DM halo density profile, while their shallow baryonic potentials have little influence on the DM dynamics (although see Pontzen & Governato 2012; Read et al. 2016, for example). As a demonstration of the utility of this work, we constrain the cross-section model parameter space favored by the measurement of the rotation curve of the LSB dwarf galaxy UGC 128 (van der Hulst et al. 1993; Kamada et al. 2017). We select UGC 128 because in this system the circular velocity is mostly determined by the dark matter component, while the stellar disk has negligible effect (Kamada et al. 2017). The fact that our model does not capture baryonic physics is therefore not expected to significantly influence the SIDM parameters inferred from UGC 128’s rotation curve. We notice that the rotation curves of dozens of dwarf/LSB galaxies have been measured carefully (e.g., Kuzio de Naray et al. 2008; Oh et al. 2015; Essig et al. 2019). Some of the measurements have been used to derive stringent upper bounds on cross-section models (e.g., Read et al. 2018; Jiang et al. 2021). While a careful and thorough exploration of the SIDM cross-section model parameter space using all the available dwarf/LSB observational data will be an important step to take, in this work we choose to show just an example application of the gravothermal solution mapping method. We therefore illustrate regions of parameter space favored by the data on UGC 128 alone and defer a more complete analysis of a larger sample of dwarf/LSB galaxies to a subsequent work. In this work we trust the measurement of UGC 128’s rotation curve and ignore potential underestimation of the error caused by beam smearing, coherent turbulence, the lack of determination of pressure support, and the treatment of inclination errors (Oman et al. 2019; Sellwood et al. 2021). The derived constraints can therefore be overtightened.

We assume $\beta = 0.5$ and a halo evolution time of 10 Gyr. We then use EMCEE (Foreman-Mackey et al. 2013) to conduct the Markov Chain Monte Carlo (MCMC) sampling over the 4D parameter space $-2.5 \leq \log(\sigma_c/\text{cm}^2 \text{g}^{-1}) \leq 4.0$, $0.0 \leq \log(\omega/\text{km s}^{-1}) \leq 4.0$, $5.0 \leq \log(\rho_s/M_\odot \text{kpc}^{-3}) \leq 8.0$, and $-1.0 \leq \log(r_s/\text{kpc}) \leq 2.0$, assuming uniform priors on these logarithmic variables. Here σ_c and ω are parameters introduced by the velocity-dependent cross-section model $\sigma(v_{12}) = \sigma_c/(1 + v_{12}^2/\omega^2)^2$, and $\{\rho_s, r_s\}$ are the NFW profile scaling parameters. In each MCMC step we use the gravothermal mapping method introduced in Section 6 to compute the halo density profile at $t = 10$ Gyr, and transfer the dark matter density profile to the rotation curve through $v_{\text{circ}}(r) = \sqrt{GM(<r)/r}$. Here $v_{\text{circ}}(r)$ is the circular velocity at radius r . We adopt a Gaussian likelihood function to direct 20 MCMC random walkers with the measurement of UGC 128’s rotation curve. To get a quantitative sense of how strongly the baryonic gravitational potential in this system may influence the parameter fitting results, we perform two independent MCMC fits for the total rotation curve and the rotation curve of the dark matter halo, respectively. Contributions to the rotation curve from stars and gas are fit by Kamada et al. (2017). A triangle plot for the parameter constraints is presented in Figure 8(left panel).

The MCMC fitting results show that the NFW parameters ρ_s and r_s are more sensitive to the differences between the total and DM-only rotation curves, while the $\log \sigma_c - \log r_s$ parameter constraints show little response to this data variation. This is fortunate as we are generally more interested in learning about the cross-section parameters rather than fitting NFW parameters of different halos with high accuracy. However, the generality of this feature should be further tested using a larger observational sample.

The MCMC results also show that the NFW scaling parameters ρ_s and r_s are degenerate. This is expected since the increase in amplitude of the density profile caused by increasing ρ_s can be compensated by decreasing r_s , which effectively shifts the halo density profile to smaller radii. Despite this parameter degeneracy we are still able to constrain $\{\rho_s, r_s\}$ within a range much narrower than the flat priors. The best-fit NFW parameters for this halo fitted from the total rotation curve are $\log(\rho_s/M_\odot \text{kpc}^{-3}) = 6.49^{+0.17}_{-0.16}$ and $\log(r_s/\text{kpc}) = 1.34 \pm 0.09$ (1σ errors). For the rotation curve with the fitted stellar disk and gas components subtracted, the best-fit NFW scales are $\log(\rho_s/M_\odot \text{kpc}^{-3}) = 6.40 \pm 0.17$ and $\log(r_s/\text{kpc}) = 1.35 \pm 0.10$. There is a “L”-shaped degeneracy between the cross-section parameters $\log \sigma_c$ and $\log \omega$. One can also see such a degeneracy from the asymptotic behavior of the cross-section model. Specifically, the cross-section model reduces to the constant factor σ_c for high ω , and the parameter constraints become independent of ω . On the other hand, lowering ω effectively decreases the cross section, resulting in a weaker upper bound on σ_c . The turnover of this “L”-shaped band occurs on a scale comparable to the characteristic rms velocity of the halo. Therefore, the degeneracy between σ_c and ω can be at least partially broken by combining similar parameter constraints for systems of different sizes, masses, and characteristic velocity dispersions, and we defer a more careful analysis to a future work. We also emphasize that in this fit we have assumed $\beta = 0.5$ and a halo evolution time of 10 Gyr. Due to the $\beta\hat{\sigma}_c$ degeneracy introduced in Sections 4–6, assuming alternative values for β and halo evolution time

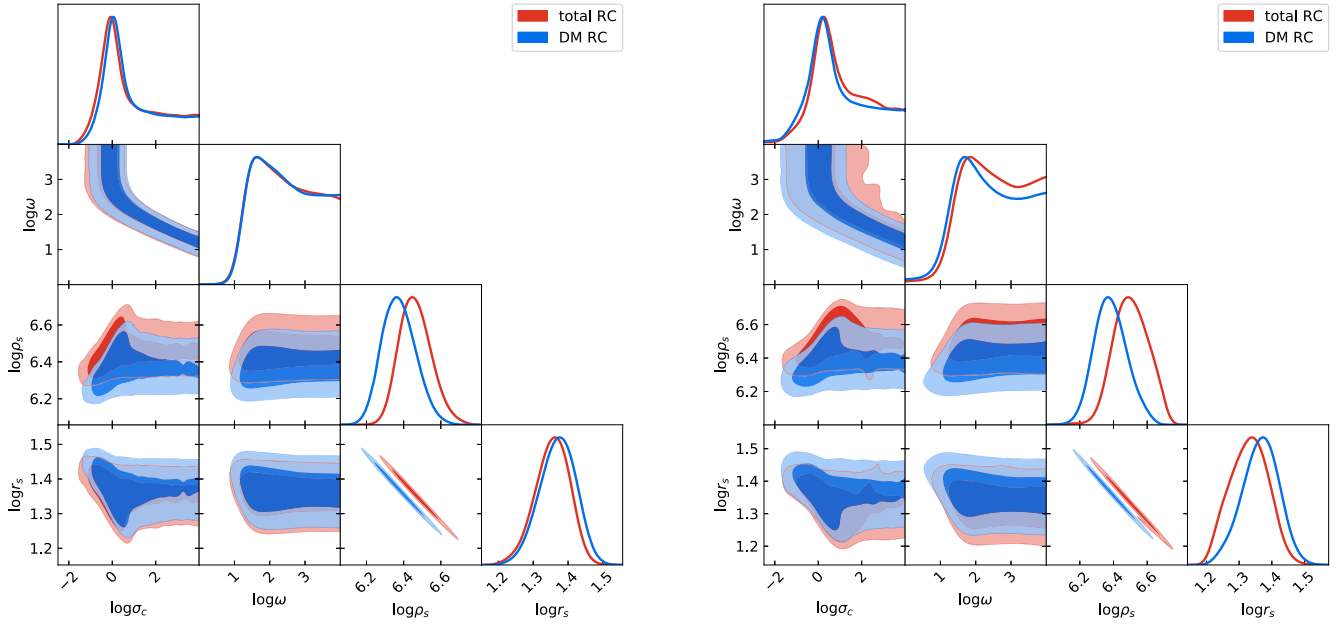


Figure 8. Parameter constraints for the NFW initial condition Equation (1) and the cross-section model Equation (2) resulting from measurements of the rotation curve of dwarf galaxy UGC 128 (van der Hulst et al. 1993; Kamada et al. 2017). We fix $\beta = 0.5$ and a halo evolution time of 10 Gyr. We assume Gaussian likelihood for the total/DM-only rotation curve measurements and show the parameter space favored by observation in red/blue regions. The contours enclose parameter constraints at 68% and 95% confidence levels. The left panel shows MCMC fitting results derived from the original UGC 128 rotation curve, while for the right panel the error bars on the rotation curve for the innermost three radius bins ($r \leq 5$ kpc) are enlarged by a factor of 2 to account for potential underestimation of measurement uncertainty.

would rescale the σ_c constraints by a factor of $1/(\beta t)$. The constraints on parameters $\{\omega, \rho_s, r_s\}$ are insensitive to the assumptions on β and halo age.

The single “L”-shaped band in the $\log \sigma_c - \log \omega$ parameter space favored by the observed rotation curve was a surprise to us. Specifically, we have assumed an NFW initial density profile, for which the halo central density first decreases, reaching a maximum core size, and then increases with time. There will be in general two $\beta \sigma t$ instances where the cored halo density profiles are very similar to each other (unless the halo is measured to be precisely at the moment of maximum core size, or the halo has entered the late core collapsing stage where its central density is higher than in the initial state): one from the core formation process and another one from the core collapsing process. The existence of two gravothermal solutions with similar halo density profiles but very different $\beta \sigma t$ is one of the causes of the huge dispersion in SIDM cross-section constraints among different works (Kamada et al. 2020). For example, SIDM cross sections of the Milky Way’s dwarf satellite galaxies constrained by isothermal models such as those of Kamada et al. (2017) and Valli & Yu (2018) only pick out the low-cross-section solution because the isothermal models by design only work for the SIDM halo core formation process. Cross sections constrained by isothermal-based methods are therefore relatively low, ranging from 0.1 to 40 $\text{cm}^2 \text{g}^{-1}$. On the other hand, previous studies such as Correa (2021) that use the gravothermal fluid formalism to constrain SIDM cross sections may only focus on the halo core collapsing process, and give higher constraints of 30–200 $\text{cm}^2 \text{g}^{-1}$ for similar systems. We find the “L”-shaped band favored by UGC 128’s rotation curve points to one solution during the core formation process. As an example, we pick one point in the best-fit parameter space $\sigma_c = 1 \text{ cm}^2 \text{g}^{-1}$, $\omega = 10^4 \text{ km s}^{-1}$, and compare the mapped halo density profile as well as the rotation curve at $t = 10$ Gyr with observations in Figure 9.

We show the best-fit gravothermal solutions to the total and DM-only rotation curves as black and blue solid curves, respectively. We find fixing the best-fit ρ_s, r_s , and ω values, but enlarging σ_c to about $100 \text{ cm}^2 \text{g}^{-1}$ gives very similar mapped halo density profile at $t = 10$ Gyr, although under such a large cross section the halo has started core collapse. The gravothermal solutions with identical halo central density to the best-fit solution, but a much larger cross section, are shown by dashed lines in Figure 9. The red/green dashed line corresponds to the case of total/DM-only rotation curve. We show in the right panel of Figure 9 that, for either the total or the DM-only case, rotation curve measurements of UGC 128 at radii less than about 5 kpc do not favor either one of those two gravothermal solutions, owing to their similar halo central densities. To clarify the origin of this tight cross-section constraint from UGC 128, in the left panel of Figure 9 we compare these two sets of gravothermal solutions to NFW profiles with the best-fit $\{\rho_s, r_s\}$ parameters to the total/DM-only rotation curve, shown by the black/blue dotted curve. During the halo core formation process where $\sigma_c = 1 \text{ cm}^2 \text{g}^{-1}$, the halo density profile is only suppressed in the cored region ($r \lesssim 5$ kpc) compared to the NFW initial condition. However, for the $\sigma_c \approx 100 \text{ cm}^2 \text{g}^{-1}$ case the halo density is altered by DM self-interactions over a larger radial range ($r \lesssim 100$ kpc). As a result the halo density profile outside the cored region becomes steeper than the initial NFW profile. This subtle difference is distinguishable by the rotation curve measurements outside the isothermal core. However, we emphasize that all the above analysis relies on the assumption that the data on UGC 128’s rotation curve and its measurement uncertainties are statistically valid. The error bars on the rotation curve can be underestimated, especially in the central region of the galaxy, due to center offsets, noncircular motions, inclination, the effects of pressure support, and many other observational uncertainties (see de Blok 2010 for a review). While a careful

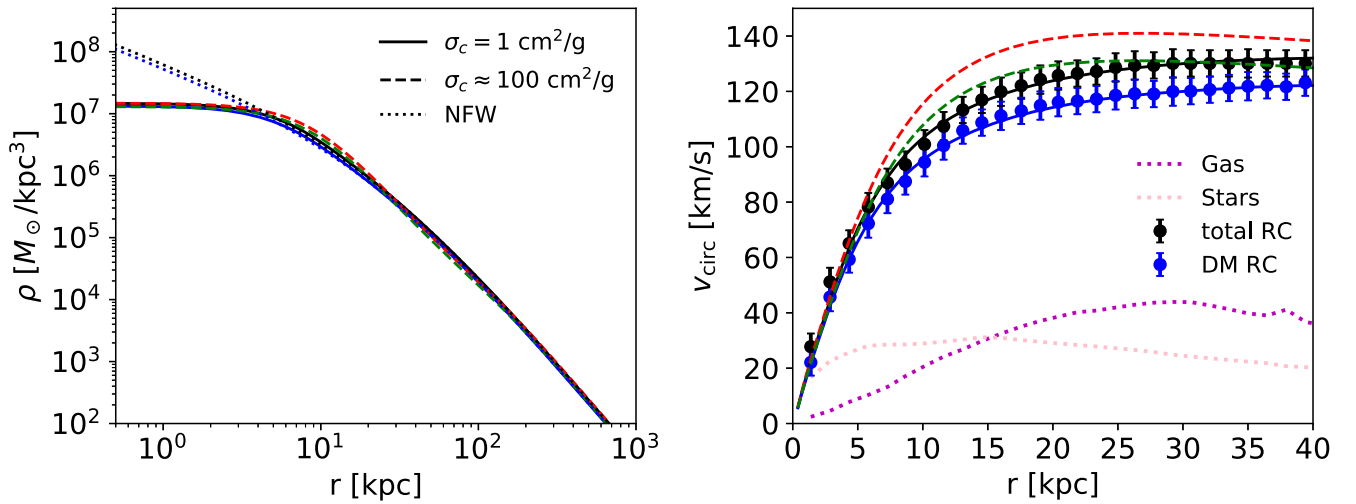


Figure 9. The best-fit halo density profiles (left) and rotation curves (right) at 10 Gyr, assuming $\beta = 0.5$ but different SIDM cross sections. The black/blue solid curves show the best-fit mapped gravothermal solution to the total/DM-only rotation curve. These two best-fit solutions correspond to a small cross section $\sigma_c = 1 \text{ cm}^2 \text{ g}^{-1}$ and $\omega = 10^4 \text{ km s}^{-1}$. The red/green dashed lines show mapped gravothermal solutions under a larger cross section $\sigma_c \approx 100 \text{ cm}^2 \text{ g}^{-1}$ and $\omega = 10^4 \text{ km s}^{-1}$, but identical halo central densities to the best-fit solutions to the total/DM-only rotation curve. Compared to the NFW profiles with the best-fit $\{\rho_s, r_s\}$ parameters for the total/DM-only rotation curve shown by the black/blue dotted curves, density profiles for the $\sigma_c = 1 \text{ cm}^2 \text{ g}^{-1}$ case are only suppressed in the cored region, while a larger region in the halo is influenced by the DM self-interaction for the $\sigma_c \approx 100 \text{ cm}^2 \text{ g}^{-1}$ case. Only the solutions for small cross section match the observed total/DM-only rotation curves, shown by black/blue data points. The purple and pink dotted curves show rotation curves contributed by gas and stars in UGC 128, fitted by Kamada et al. (2017).

reprocessing of the data on UGC 128’s rotation curve is beyond the scope of this work, we attempt to account for potentially underestimated uncertainties by increasing the error bars on UGC 128’s rotation curve of the innermost three radius bins, where the radius is less than 5 kpc, by a factor of 2. The MCMC fitting results are shown in Figure 8(right panel). Assuming larger measurement uncertainties near the galaxy center, the “L”-shaped bands in the $\log \sigma_c - \log \omega$ space become significantly wider such that cross sections as large as $\sigma_c \sim 100 \text{ cm}^2 \text{ g}^{-1}$ can be preferred by the rotation curve measurement at 95% confidence level.

We note that dissipative cooling processes (Essig et al. 2019), the effects of a baryonic disk (Jiang et al. 2022), and tidal effects (Zeng et al. 2022) can all speed up the onset of halo core collapse. On the other hand, Meshveliani et al. (2022) suggests that halo mass accretion and merger can delay the core collapse process, and proposes a correction term to the halo core collapsing time in order to account for this effect. However, the goal of this work is simply to introduce a convenient nonlinear method for rescaling time for the gravothermal fluid formalism without considering any of the above effects. Nevertheless, such a preliminary constraint already shows the advantage of our mapping method in the sense that it can rapidly explore the continuous parameter space and clearly present the parameter degeneracies.

9. Conclusion and Discussion

In this work we introduce a fast mapping method that transfers the gravothermal fluid formalism solutions from the constant-cross-section scenario to arbitrary velocity-dependent cross-section models. Such a mapping is intrinsically built into the gravothermal fluid formalism. Specifically, when a halo is evolved fully in either the lmfp or the smfp regime, the cross section only influences the heat conductivity of the isolated halo at each radius and moment, which further determines the halo evolution rate. Varying the assumptions of the cross-section model therefore only remaps the time axis of the halo

evolution, and has no effect on the radial profile amplitudes/shapes of halo density, velocity, enclosed mass, etc. Such a time remapping is linear between two different constant cross sections, but becomes nonlinear between constant and velocity-dependent cross-section models.

The thermal conductivity of an SIDM halo varies at different radii and times. To achieve fast mapping we use the thermal conductivity at the halo center at each instance as a characteristic value. We test the performance of this mapping method through comparing the mapped gravothermal solutions, assuming a constant cross section, with actual gravothermal numerical solutions assuming different velocity-dependent cross-section models. We find that the fast mapping method introduced in this work nicely captures the full time evolution of an SIDM halo, although it slightly overestimates the core collapsing time due to the simplified assumption about the instantaneous heat conductivity of the halo. We show that the accuracy of this mapping method can be further boosted through a more careful treatment of the calculation of instantaneous heat conductivity, but argue that since the fractional error of the default mapping is small compared to the uncertainties of halo evolution time and the free parameter β , the more complicated mapping method is unnecessary.

Calibrating gravothermal solutions with N -body simulations under the assumption of a velocity-dependent cross section is more challenging since in numerical simulations the scattering probabilities of particles are generally determined by their relative velocities, while the constant-cross-section gravothermal fluid formalism only cares about the particle rms velocity/velocity dispersion in each radial shell. Assuming the SIDM particle velocity in each halo shell follows the Maxwell-Boltzmann distribution, we define the averaged thermal conductivity as $\langle \kappa_{\text{lmfp}} \rangle \propto \langle v_1 v_2 v_{12}^3 \sigma(v_{12}) \rangle$, where v_1, v_2 are the velocities of two particles, and v_{12} is their relative velocity. The velocity terms in the averaging bracket of the MB distribution are contributed by particle orbiting scales, collisional relaxation times, and energy transfer. We show that this physically

motivated definition of the averaged heat conductivity provides an excellent match between the mapped gravothermal solutions and AREPO idealized N -body simulation results, assuming multiple different velocity-dependent cross-section models.

The advantage of this mapping method lies in its high computational efficiency. Specifically, it takes about 24 CPU hours of computation with our gravothermal fluid code to reach the core-collapse regime for an SIDM halo in the lmfp regime, and introducing a velocity-dependent cross-section model brings at least two extra degrees of freedom. It is therefore impractical to continuously explore the cross-section model parameter space directly with the gravothermal fluid code. Luckily, by adopting the mapping method introduced in this work one only needs to solve the fluid code once, and one can fast map this set of numerical solutions to halos with arbitrary assumptions on cross section. Besides deriving parameter constraints, this mapping method is also suitable to be implemented into semianalytic galaxy formation models in order to achieve fast generation of SIDM halo populations.

As an example application, we select the LSB system UGC 128 where the baryonic disk potential has negligible influence on the galaxy rotation curve, and use the mapping method to constrain four free parameters in the NFW and velocity-dependent cross-section models. We have assumed $\beta = 0.5$ and a halo evolution time of 10 Gyr, but we show analytically that the βt assumption will only rescale constraints on the cross-section scaling parameter σ_c . The other three parameters $\{\omega, \rho_s, r_s\}$ are not sensitive to the assumed values for β and halo age. To get a quantitative sense of how much the baryonic gravitational potential in this system may influence the parameter fitting results, we perform two independent MCMC fits for the total rotation curve and the dark matter halo rotation curve, respectively, finding that the SIDM parameters of interest are unaffected. We also show that the accurately measured rotation curve of UGC 128 can distinguish two very similar gravothermal solutions during the core formation and core collapsing phases, and favors the low-cross-section solution. We notice that the constraints derived in this work do not account for other possible nonlinear effects that can influence the halo core collapsing time, including cooling, an additional gravitational potential contributed by a baryonic disk or black hole, tidal effects, halo mass accretion, and merger. The constraints can be too tight since we have trusted the measurements of UGC 128's rotation curve, and have ignored possible underestimation of the error. Nevertheless the advantage of this fast gravothermal solution mapping method is recognized in the sense that it can quickly explore the continuous 4D parameter space with the first-principles gravothermal solutions, and effectively capture degeneracies among the SIDM halo parameters of interest.

We also compare the performance of this mapping method with a similar strategy proposed in O22. We find the simpler method introduced in O22, where the characteristic heat conductivity is assumed to be a time- and radius-independent constant, outperforms the method introduced in this work in reproducing the core collapse time when mapping between gravothermal solutions with v_{rms} -dependent cross-section models. However, this work performs better in mapping the full halo evolution process, while O22 fails to capture the evolution of the halo central density before the instant of maximum core size if the cross-section model has strong velocity dependence. Correctly mapping the SIDM halo core formation process is as

important as capturing the core collapse time because: (1) most current SIDM cosmological simulations evolve halos with mild self-interaction cross section for only around 10 Gyr, such that the simulation stops before an isothermal core has been completely formed (e.g., Elbert et al. 2015; Nadler et al. 2020), and (2) observations may prefer an SIDM model with small cross section. If this is the case, most halos hosting observable galaxy clusters/galaxies/dwarfs are still approaching their maximum core size. O22 propose to estimate the averaged thermal conductivity in a particle relative velocity-dependent scenario as $\langle \kappa_{\text{lmfp}} \rangle \propto \langle v_{12}^n \sigma(v_{12}) \rangle$, where n is to be calibrated to N -body simulations. We find that a physically motivated definition $\langle \kappa_{\text{lmfp}} \rangle \propto \langle v_1 v_2 v_{12}^3 \sigma(v_{12}) \rangle$ outperforms the ansatz assumed in O22, where $n=3$. However, we show that combining the linear time rescaling method introduced by O22 and the definition $\langle \kappa_{\text{lmfp}} \rangle \propto \langle v_{12}^5 \sigma(v_{12}) \rangle$, also suggested by Yang & Yu (2022), gives an equally good match to the N -body simulation results as $\langle \kappa_{\text{lmfp}} \rangle \propto \langle v_1 v_2 v_{12}^3 \sigma(v_{12}) \rangle$.

We thank Yiming Zhong for sharing his gravothermal fluid code and for many helpful discussions. We thank Mark Vogelsberger for sharing the AREPO N -body simulation package including the SIDM module. Computing resources used in this work were made available by a generous grant from the Ahmanson Foundation. This work was supported in part by the NASA Astrophysics Theory Program under grant No. 80NSSC18K1014.

Appendix An Empirical Model for the lmfp Gravothermal Density Profiles

In this appendix we introduce a convenient empirical model for the density profile of SIDM halos, calibrated to the lmfp gravothermal solution used in this work.

Balberg et al. (2002) has proved that the gravothermal density profile evolves with time self-similarly, and the density in the halo outskirts falls with radius as $\hat{\rho} \propto \hat{r}^{-2.19}$. This derivation is elegant, but not directly applicable to most of the numerical SIDM simulations. Specifically, in most idealized N -body as well as some cosmological simulations the halo starts from an NFW initial state where $\hat{\rho} \propto \hat{r}^{-3}$ at large radii. As a result, when a core has been formed the halo density still drops as $\hat{\rho} \propto \hat{r}^{-2.19}$ slightly outside the core radius. At even larger radii the density profile joins smoothly with the NFW initial condition, and drops as $\hat{\rho} \propto \hat{r}^{-3}$. At this stage the halo density profile can be captured by a triple power law.

We first define the instant of halo core collapse as the time at which the halo central density diverges:

$$\log(\beta \hat{\sigma} \hat{t}) = E = 2.238, \quad (\text{A1})$$

and the instant of maximum core size as the time when the halo central density reaches its minimum:

$$\log(\beta \hat{\sigma} \hat{t}) = F = 1.341. \quad (\text{A2})$$

We find at the early stage of the halo evolution, i.e., $\log(\beta \hat{\sigma} \hat{t}) < F$, the halo density profile can still be described by the NFW profile multiplied by a tanh central cut:

$$\hat{\rho} = \frac{\tanh(\hat{r}/r_{\text{core}})}{\hat{r}(1 + \hat{r})^2}, \quad (\text{A3})$$

where we use free parameter r_{core} to capture the core size. We find the following empirical model described the time evolution of the halo core size at $\log(\beta\hat{\sigma}\hat{t}) \leq 1.341$ with high accuracy:

$$\log r_{\text{core}} = A_{r_{\text{core}}}(\log(\beta\hat{\sigma}\hat{t}))^2 + B_{r_{\text{core}}}\log(\beta\hat{\sigma}\hat{t}) + C_{r_{\text{core}}}. \quad (\text{A4})$$

Here $A_{r_{\text{core}}} = -0.1078$, $B_{r_{\text{core}}} = 0.3737$, and $C_{r_{\text{core}}} = -0.7720$ are calibrated to the gravothermal solutions.

As mentioned before, in later phases of halo evolution, $\log(\beta\hat{\sigma}\hat{t}) > 1.341$, when a core has been formed, the density profile can be described by triple power law:

$$\hat{\rho}(\hat{r}) = \frac{\rho_{\text{core}}}{1 + (\hat{r}/r_{\text{core}})^s(1 + \hat{r}/r_{\text{out}})^{3-s}}. \quad (\text{A5})$$

Here $s = 2.19$ is provided by the gravothermal solution self-similar analysis of Balberg et al. (2002). Again we use free parameter r_{core} to trace the characteristic halo core size at each time step. Similarly, the parameter ρ_{core} captures the halo central density, and r_{out} traces the radius where the density profile influenced by self-interaction joins smoothly to the NFW outskirts. We develop the following empirical model for ρ_{core} , r_{core} , and r_{out} , calibrated to the lmfp gravothermal solution:

$$\begin{aligned} \log \rho_{\text{core}} = & A_{\rho_{\text{core}}}(\log \beta\hat{\sigma}\hat{t} - (E + 3))^2 + C_{\rho_{\text{core}}} \\ & + \frac{D_{\rho_{\text{core}}}}{(E + 0.0001 - \log \beta\hat{\sigma}\hat{t})^{0.02}}, \end{aligned} \quad (\text{A6})$$

where $A_{\rho_{\text{core}}} = 0.05771$, $C_{\rho_{\text{core}}} = -21.64$, $D_{\rho_{\text{core}}} = 21.11$,

$$\begin{aligned} \log r_{\text{core}} = & A_{r_{\text{core}}}(\log \beta\hat{\sigma}\hat{t} - (E + 2))^2 + C_{r_{\text{core}}} \\ & + \frac{D_{r_{\text{core}}}}{(E + 0.0001 - \log \beta\hat{\sigma}\hat{t})^{0.005}}, \end{aligned} \quad (\text{A7})$$

where $A_{r_{\text{core}}} = -0.04049$, $C_{r_{\text{core}}} = 43.07$, $D_{r_{\text{core}}} = -43.07$, and

$$\begin{aligned} \log r_{\text{out}} = & A_{r_{\text{out}}}(\log \beta\hat{\sigma}\hat{t} - (E + 2))^2 + C_{r_{\text{out}}} \\ & + \frac{D_{r_{\text{out}}}}{(E + 0.04 - \log \beta\hat{\sigma}\hat{t})^{0.005}}, \end{aligned} \quad (\text{A8})$$

where $A_{r_{\text{out}}} = 0.02403$, $C_{r_{\text{out}}} = -4.724$, $D_{r_{\text{out}}} = 5.011$.

We compare the halo density profile given by the gravothermal solution and this empirical model at five characteristic moments in Figure 10. This simple empirical

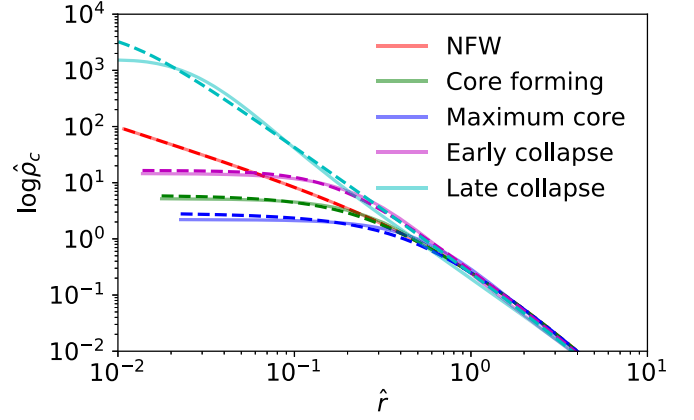


Figure 10. Comparisons of density profile between the gravothermal solutions and the empirical model introduced in the appendix. We compare the numerical solution and empirical model at the halo initial state $\beta\hat{\sigma}\hat{t} = 0$ (red), core formation state $\beta\hat{\sigma}\hat{t} < F = 1.341$ (green), maximum core moment $\beta\hat{\sigma}\hat{t} = F$ (blue), early core collapse $\beta\hat{\sigma}\hat{t} = 150$ (magenta), and the late core collapse state $\beta\hat{\sigma}\hat{t} = E = 2.238$ (cyan). The density profiles of the gravothermal solution are shown as faint solid curves, while the empirical model predictions are shown by the dashed lines.

model accurately captures the overall time evolution and radial trend of the gravothermal solutions throughout the halo evolution process. We also show the halo central density using the exact result from the gravothermal fluid model and our empirical model (along with their fractional difference) in Figure 11. The fractional error of this empirical model is better than 10% at most times. Even at the very late core collapsing stage its accuracy is still better than 50%. This degree of accuracy is sufficient for many applications, considering the fact that the random fluctuations in N -body simulation results can cause the halo central density to fluctuate at a similar level. The fractional error shows noncontinuous behavior at the maximum core moment $\beta\hat{\sigma}\hat{t} = 1.341$ because we switch the empirical description of the density profile from a tanh cutoff to the triple power law here.

This analytic empirical model has wide applications. For example, it can be implemented into semianalytic galaxy formation models for fast SIDM halo simulation. As another example, we show in Section 5 that the characteristic radius r_{out} provided by this model can be used to boost the accuracy of the mapping method introduced in this work.

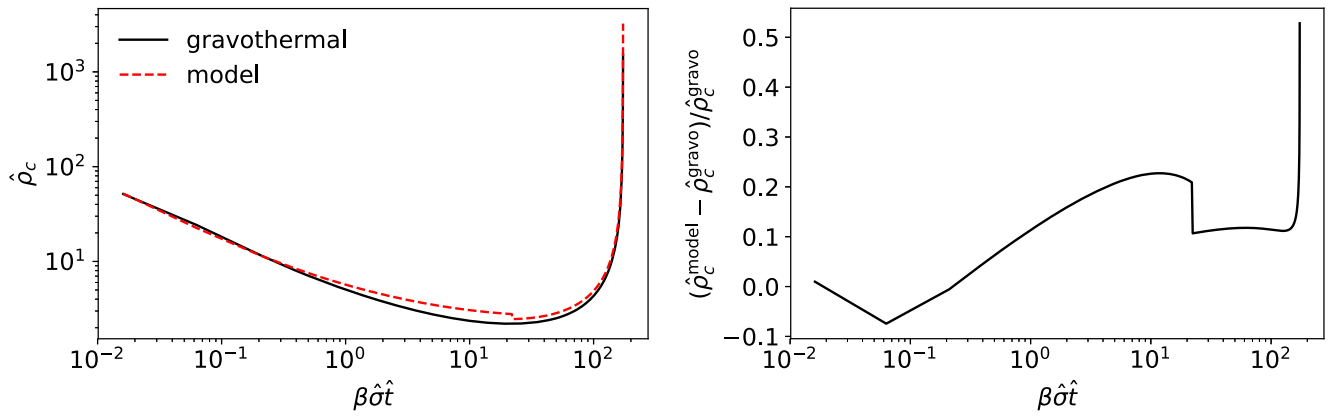


Figure 11. Left: comparison of the time evolution of the halo central density between the gravothermal solutions (black solid curve) and empirical model (red dashed curve) introduced in this section. Right: fractional error of the empirical model in reproducing the halo central density given by the gravothermal solutions.

ORCID iDs

Shengqi Yang <https://orcid.org/0000-0002-0782-9116>

Xiaolong Du <https://orcid.org/0000-0003-0728-2533>

Zhichao Carton Zeng <https://orcid.org/0000-0002-6008-5144>

Andrew Benson <https://orcid.org/0000-0001-5501-6008>

Ethan O. Nadler <https://orcid.org/0000-0002-1182-3825>

Annika H. G. Peter <https://orcid.org/0000-0002-8040-6785>

References

- Ackerman, L., Buckley, M. R., Carroll, S. M., & Kamionkowski, M. 2009, *PhRvD*, **79**, 023519
- Ahn, K., & Shapiro, P. 2003, *JKAS*, **36**, 89
- Anderson, L., Aubourg, E., Bailey, S., et al. 2012, *MNRAS*, **427**, 3435
- Andrade, K. E., Fuson, J., Gad-Nasr, S., et al. 2022, *MNRAS*, **510**, 54
- Balberg, S., & Shapiro, S. L. 2002, *PhRvL*, **88**, 101301
- Balberg, S., Shapiro, S. L., & Inagaki, S. 2002, *ApJ*, **568**, 475
- Benson, A. J. 2012, *NewA*, **17**, 175
- Colquhoun, B., Heeba, S., Kahlhoefer, F., Sagunski, L., & Tulin, S. 2021, *PhRvD*, **103**, 035006
- Correa, C. A. 2021, *MNRAS*, **503**, 920
- de Blok, W. J. G. 2010, *AdAst*, **2010**, 789293
- de Blok, W. J. G., Walter, F., Brinks, E., et al. 2008, *AJ*, **136**, 2648
- Eggleton, P. P. 1971, *MNRAS*, **151**, 351
- Elbert, O. D., Bullock, J. S., Garrison-Kimmel, S., et al. 2015, *MNRAS*, **453**, 29
- Elbert, O. D., Bullock, J. S., Kaplinghat, M., et al. 2018, *ApJ*, **853**, 109
- Essig, R., McDermott, S. D., Yu, H.-B., & Zhong, Y.-M. 2019, *PhRvL*, **123**, 121102
- Feng, J. L., Kaplinghat, M., Tu, H., & Yu, H.-B. 2009, *JCAP*, **2009**, 004
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Garrison-Kimmel, S., Rocha, M., Boylan-Kolchin, M., Bullock, J. S., & Lally, J. 2013, *MNRAS*, **433**, 3539
- Gilman, D., Bovy, J., Treu, T., et al. 2021, *MNRAS*, **507**, 2432
- Ibe, M., & Yu, H.-B. 2010, *PhLB*, **692**, 70
- Jiang, F., Benson, A., Hopkins, P. F., et al. 2022, arXiv:2206.12425
- Jiang, F., Kaplinghat, M., Lisanti, M., & Slone, O. 2021, arXiv:2108.03243
- Kamada, A., Kaplinghat, M., Pace, A. B., & Yu, H.-B. 2017, *PhRvL*, **119**, 111102
- Kamada, A., Kim, H. J., & Kuwahara, T. 2020, *JHEP*, **2020**, 202
- Kaplinghat, M., Tulin, S., & Yu, H.-B. 2016, *PhRvL*, **116**, 041302
- Koda, J., & Shapiro, P. R. 2011, *MNRAS*, **415**, 1125
- Kummer, J., Kahlhoefer, F., & Schmidt-Hoberg, K. 2018, *MNRAS*, **474**, 388
- Kuzio de Naray, R., McGaugh, S. S., & de Blok, W. J. G. 2008, *ApJ*, **676**, 920
- Lynden-Bell, D., & Eggleton, P. P. 1980, *MNRAS*, **191**, 483
- Lynden-Bell, D., Wood, R., & Royal, A. 1968, *MNRAS*, **138**, 495
- Mao, Y.-Y., Williamson, M., & Wechsler, R. H. 2015, *ApJ*, **810**, 21
- Meshveliani, T., Zavala, J., & Lovell, M. R. 2022, arXiv:2210.01817
- 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. 1996, *ApJ*, **462**, 563
- Nishikawa, H., Boddy, K. K., & Kaplinghat, M. 2020, *PhRvD*, **101**, 063009
- Oh, S.-H., Hunter, D. A., Brinks, E., et al. 2015, *AJ*, **149**, 180
- Oman, K. A., Marasco, A., Navarro, J. F., et al. 2019, *MNRAS*, **482**, 821
- Outmezguine, N. J., Boddy, K. K., Gad-Nasr, S., Kaplinghat, M., & Sagunski, L. 2022, arXiv:2204.06568
- Peebles, P. J. E. 1982, *ApJL*, **263**, L1
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011, *A&A*, **536**, A1
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, *A&A*, **571**, A16
- Pollack, J., Spergel, D. N., & Steinhardt, P. J. 2015, *ApJ*, **804**, 131
- Pontzen, A., & Governato, F. 2012, *MNRAS*, **421**, 3464
- Read, J. I., Agertz, O., & Collins, M. L. M. 2016, *MNRAS*, **459**, 2573
- Read, J. I., Walker, M. G., & Steger, P. 2018, *MNRAS*, **481**, 860
- Sagunski, L., Gad-Nasr, S., Colquhoun, B., Robertson, A., & Tulin, S. 2021, *JCAP*, **2021**, 024
- Sellwood, J. A., Spekkens, K., & Eckel, C. S. 2021, *MNRAS*, **502**, 3843
- Shapiro, S. L. 2018, *PhRvD*, **98**, 023021
- Spergel, D. N., & Steinhardt, P. J. 2000, *PhRvL*, **84**, 3760
- Spitzer, L. 1987, *Dynamical Evolution of Globular Clusters* (Princeton, NJ: Princeton Univ. Press)
- Springel, V. 2010, *MNRAS*, **401**, 791
- Valli, M., & Yu, H.-B. 2018, *NatAs*, **2**, 907
- van der Hulst, J. M., Skillman, E. D., Smith, T. R., et al. 1993, *AJ*, **106**, 548
- Vogelsberger, M., Zavala, J., & Loeb, A. 2012, *MNRAS*, **423**, 3740
- Vogelsberger, M., Zavala, J., Schutz, K., & Slatyer, T. R. 2019, *MNRAS*, **484**, 5437
- Vogelsberger, M., Zavala, J., Simpson, C., & Jenkins, A. 2014, *MNRAS*, **444**, 3684
- Yang, D., & Yu, H.-B. 2022, *JCAP*, **2022**, 077
- Zeng, Z. C., Peter, A. H. G., Du, X., et al. 2022, *MNRAS*, **513**, 4845