Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **519,** 5543–5553 (2023) https://doi.org/10.1093/mnras/stad036
Advance Access publication 2023 January 9

A new discrete dynamical friction estimator based on *N*-body simulations

Linhao Ma [1]★ Philip F. Hopkins [1] Luke Zoltan Kelley [2] and Claude-André Faucher-Giguère [2]

[1] *TAPIR, California Institute of Technology, Mailcode 350-17, Pasadena, CA 91125, USA*
[2] *CIERA and Department of Physics & Astronomy, Northwestern University, Evanston, IL 60208, USA*

Accepted 2023 January 3. Received 2022 December 6; in original form 2022 August 24

ABSTRACT

A long-standing problem in galactic simulations is to resolve the dynamical friction (DF) force acting on massive black hole particles when their masses are comparable to or less than the background simulation particles. Many sub-grid models based on the traditional Chandrasekhar DF formula have been proposed, yet they suffer from fundamental ambiguities in the definition of some terms in Chandrasekhar's formula when applied to real galaxies, as well as difficulty in evaluating continuous quantities from (spatially) discrete simulation data. In this work, we present a new sub-grid DF estimator based on the discrete nature of *N*-body simulations, which also avoids the ambiguously defined quantities in Chandrasekhar's formula. We test our estimator in the GIZMO code and find that it agrees well with high-resolution simulations where DF is fully captured, with negligible additional computational cost. We also compare it with a Chandrasekhar estimator and discuss its applications in real galactic simulations.

Key words: Galaxy: kinematics and dynamics – methods: numerical – black hole physics – quasars: supermassive black holes.

1 INTRODUCTION

An essential element in the study of galactic dynamics is the process of dynamical friction (DF; Chandrasekhar 1943), a statistical effect of numerous two-body scatterings that causes a massive particle to lose its momentum when it travels through a medium of much lighter background particles. DF is believed to be an important effect to drive massive black holes (BHs; from intermediate-mass BHs to supermassive BHs) to galactic centres (see e.g. Ostriker 1999; Chen et al. 2022; Weller et al. 2022), and it plays an essential role in the evolution of globular clusters (see e.g. Portegies Zwart & McMillan 2002; Gürkan, Freitag & Rasio 2004; Alessandrini et al. 2014; Shi, Grudić & Hopkins 2021). Hence, the evaluation of DF is important in studying the evolution of galaxies, globular clusters, and BHs in a wide variety of contexts.

In numerical *N*-body simulations with sufficient resolution (such as in the limit in which all bodies such as stars, BHs, or even dark matter (DM) particles are represented by individual *N*-body particles), DF will be automatically captured. However, as DF is an accumulated effect of many weak encounters in the regime where the 'target' mass is much larger than the mass of the 'background' particles ($M_{\rm target\ particle} \gg M_{\rm background\ particle}$), it is often not possible to fully resolve this background. This is especially true in large-scale simulations of e.g. galactic scales, where a typical '*N*-body particle' can easily have mass much larger than intermediate-mass and supermassive BHs ($\gg 10^4\,{\rm M}_\odot$), let alone the masses of individual stars, DM particles, or hydrogen ions. Specifically, when the *N*-body particle mass becomes comparable to or larger than the 'target' mass, the explicit results of an *N*-body solver will not return the correct DF forces. For example, in the 'high-resolution' simulations of high-redshift galaxies in Ma et al. (2018a, b, 2019), the baryonic mass resolution (Δm_i) is $\sim 7000 m_\odot$ and the DM mass resolution is five times larger, which makes it impossible to resolve DF effects for BHs or other 'sink' particles (e.g. particles that might represent unresolved massive, dense structures such as globular clusters, or hyper-dense exotic DM structures, etc.) less massive than $\sim 10^5\,{\rm M}_\odot$. Hence, in these types of simulations, an additional 'sub-grid' DF force must be added to these 'target' particles to attempt to recover their real dynamics, to replace the lost information of individual two-body encounters in the smoothed-out gravity potential in simulations.

Multiple sub-grid DF models have been proposed in the literature (e.g. Colpi et al. 2007; Dotti et al. 2009; Tremmel et al. 2015; Pfister et al. 2019) based on the classical Chandrasekhar's DF formula (Chandrasekhar 1943, hereafter C43)

$$\boldsymbol{a}_{\rm df}^{\rm C43} = -4\pi G^2 M m \ln\Lambda \int {\rm d}^3\boldsymbol{v}_m\, f(\boldsymbol{v}_m) \frac{\boldsymbol{v}_M - \boldsymbol{v}_m}{|\boldsymbol{v}_M - \boldsymbol{v}_m|^3}, \quad (1)$$

where M and m are the masses of the moving 'target' particle and the background or field particles, respectively. Here, $\boldsymbol{v}_M$ and $\boldsymbol{v}_m$ are their velocities, and Λ is the Coulomb logarithm defined by $\Lambda \equiv b_{\rm max}/b_{\rm min}$, where $b_{\rm max}$ and $b_{\rm min}$ are the maximum and minimum impact factors of scattered particles in weak encounters, respectively. $f(\boldsymbol{v}_m)$ is the velocity distribution of field particles, and, with the usual assumption of a Maxwellian velocity distribution with dispersion σ, the formula reduces to (Binney & Tremaine 2008)

$$\boldsymbol{a}_{\rm df}^{\rm C43} = -\frac{4\pi^2 G^2 M \rho \ln\Lambda}{V_M^3}\left[{\rm erf}\left(\frac{v_m}{\sqrt{2}\sigma}\right) - \sqrt{\frac{2}{\pi\sigma}} {\rm e}^{-v_m^2/2\sigma^2} v_m\right]\boldsymbol{v}_M; \quad (2)$$

i.e. the DF acceleration is proportional to the local field particle density ρ and is in the opposite direction of the particle velocity $\boldsymbol{v}_M$, effectively acting as a 'friction' force. Despite its elegance and (often surprising) accuracy in estimating the DF, Chandrasekhar's formula

★ E-mail: lma3@caltech.edu

suffers from the following shortcomings when applied as a sub-grid model:

(i) In deriving the formula, C43 assumes an isotropic and homogeneous medium of field particles. This is generally not true for real galaxies. For example, it has been pointed out that high-redshift galaxies and low-redshift dwarf galaxies could be chaotic and clumpy (e.g. Weisz et al. 2014; Meng & Gnedin 2020; Flores Velázquez et al. 2021). The existence of such systems makes the physical assumptions behind C43 formula questionable.

(ii) The Coulomb logarithm is ambiguously defined, and is often selected ad hoc in practice, with a case-dependent selection of the minimum and maximum impact parameters (see e.g. Tremmel et al. 2015; Pfister et al. 2019), which introduces a large systematic uncertainty in the sub-grid model.

(iii) The formula has an explicit dependence on the local mass density, which must be evaluated from discrete *N*-body data for collisionless fluid (stars or DM, often 'blended' with gas for which the density is continuously defined, depending on the numerical hydrodynamic method). The choice of how to do so is arbitrary and has no defined 'preferred' scale. Most commonly, it is done with a local kernel density estimator at some multiple of the resolution scale (see e.g. Tremmel et al. 2015), but this is known to be quite noisy, and is not consistent with the unique local gas density available from hydrodynamic calculations.

(iv) The velocity integral and $f(\boldsymbol{v}_m)$ must be estimated with some similar ad hoc local estimator, which is also undefined, and different choices can lead to different *directions* for the DF acceleration. Usually, the choice of a local kernel sampling amplifies numerical noise further here and means that $f(\boldsymbol{v}_m)$ must be assumed to be Maxwellian (since it cannot be fitted to an arbitrary function given just a few local points).

(v) There is no self-consistent way to incorporate force softening, which is necessary in *N*-body simulations to avoid spurious divergences in the forces, as an *N*-body particle does not physically represent a point-mass particle. Failure to incorporate softening can produce inconsistent results between the (often softened) gravitational acceleration and the additional DF acceleration.

(vi) As C43 depends on *local* continuous field parameters but represents long-range forces, there is no way to self-consistently implement it in a way that conserves momentum, while in reality DF should be exactly conservative since it is derived from an infinite superposition of pair-wise *N*-body encounters.

(vii) Evaluating C43 numerically requires operations that are not algorithmically identical to the gravity solver in *N*-body equations, which introduces not only additional inconsistencies, but also substantial computational expense. This also means that numerical convergence for C43 applied to *N*-body particles is undefined: there is no formal guarantee of convergence even on idealized, smooth problems.

To tackle these problems, we develop a new sub-grid DF estimator that can be efficiently embedded into discrete *N*-body calculations in this work. The new estimator is based on a discrete version of the DF formula that can be applied to an arbitrary phase-space distribution of field particles, and avoids the fundamental ambiguity in the definitions of some terms in Chandrasekhar's formula. It also naturally embeds force softening and momentum conservation. It can also easily be generalized to assumptions beyond those of C43 for the nature of DF-like forces. We test our estimator in both on-the-fly simulations and post-processing, and compare our results to those from a Chandrasekhar DF estimator. The paper is written as follows: In Section 2, we derived our discrete DF formula. In Section 3, we describe the methods we use to test the estimator. In Sections 4 and 5, we present and discuss the results.

2 DERIVATION OF OUR DF FORMULA

Here, we present the derivation of our discrete DF formula, and general comments on its application in *N*-body methods.

2.1 Derivation

In C43, the classical DF formula is derived as follows: assume that a test particle with mass M travels through an infinite, homogeneous and isotropic medium (filled with background particles with mass $m \ll M$), and experiences a number of individual two-body encounters. During each encounter, along the direction of relative motion, the test particle velocity in the parallel direction to the initial relative velocity is altered by (after integrating along the encounter path ds from $s \rightarrow -\infty$ to $s \rightarrow +\infty$)

$$\Delta \boldsymbol{v}_{\parallel} = \frac{2\,m\,\boldsymbol{V}}{M+m}\left[1+\frac{b^2\,V^4}{G^2\,(M+m)^2}\right]^{-1} = \frac{2\,m\,\boldsymbol{V}}{(M+m)\,(1+\alpha^2)}, \tag{3}$$

where $\boldsymbol{V} \equiv \boldsymbol{v}_m - \boldsymbol{v}_M$ (i.e. the velocity of m in the rest frame of M), b is the impact parameter, and $\alpha \equiv b\,V^2/G\,(M+m)$ parametrizes the encounter strength. Note that the perpendicular deflection $\Delta \boldsymbol{v}_{\perp}$ will be cancelled by symmetry if the medium is homogeneous and isotropic so we neglect it for now, but we will return to this below. To account for the contributions of *all* encounters, C43 then integrates equation (3), by noting that the encounter rate in a differential time dt is the sum of encounters within a cylindrical slice, with surface area dA in the plane perpendicular to the relative motion and height $V\,\mathrm{d}t$, over all relatively velocities and angles

$$\boldsymbol{a}_{\rm df} \equiv \frac{\mathrm{d}\boldsymbol{v}_M}{\mathrm{d}t} = \int \Delta \boldsymbol{v}_{\parallel}\,V\,\mathrm{d}A\,\mathcal{N}(\boldsymbol{x},\boldsymbol{v})\,\mathrm{d}^3\boldsymbol{v} = \int \frac{2\,\alpha}{1+\alpha^2}\,\frac{G\,m}{b}\,\hat{\boldsymbol{V}}\,\mathcal{N}(\boldsymbol{x},\boldsymbol{v})\,\mathrm{d}p\,\mathrm{d}q\,\mathrm{d}^3\boldsymbol{v}, \tag{4}$$

where $\mathcal{N}(\boldsymbol{x},\boldsymbol{v}) = \mathrm{d}N/\mathrm{d}^3\boldsymbol{x}\,\mathrm{d}^3\boldsymbol{v}$ is the phase-space distribution function (by number) of the background particles; $\hat{\boldsymbol{V}} \equiv \boldsymbol{V}/V$, and p and q are the two spatial coordinates perpendicular to the path-length ds, i.e. characterizing the surface dA (so $\mathrm{d}s\,\mathrm{d}p\,\mathrm{d}q = \mathrm{d}^3\boldsymbol{x}$). The integral can be easily carried out for an isotropic and homogeneous distribution with $\mathcal{N}(\boldsymbol{x},\,\boldsymbol{v}) = n f_{\rm M}(\boldsymbol{v})$, where n is the number density (constant) and $f_{\rm M}(\boldsymbol{v})$ is the Maxwellian velocity distribution, leading to the classical formula.

To generalize the above formula to an arbitrary phase-space distribution sampled by a discrete set of data points as in our simulations, one might naively attempt to directly insert the usual *N*-body approximation, replacing $\mathcal{N}(\boldsymbol{x},\boldsymbol{v}) \rightarrow \sum_i (\Delta m_i/m)\,\delta(\boldsymbol{x}-\boldsymbol{x}_i,\ \boldsymbol{v}-\boldsymbol{v}_i)$. This treats the distribution function as a sum of Dirac δ-functions, i.e. point particles, each with *N*-body particle mass Δm_i, so representing $N = \Delta m_i/m$ 'background' particles of mass m. However, the integral in equation (4) only integrates over the two-dimensional surface (dpdq) as a slice of the full phase space, which makes it impossible to discretize directly. The missing integral parameter reflects the fundamental conceptual difficulty in deriving the DF formula for *arbitrary* phase-space distribution. In deriving equation (4), we actually already performed the integral over the missing degree of freedom when calculating $\Delta \boldsymbol{v}_{\parallel}$, by integrating over path-length ds in each encounter from $-\infty$ to ∞, containing the full effect of one

two-body encounter before we sum them up to get the final result. This is only correct if the background distribution is isotropic and homogeneous, since in principle, the DF process cannot be evaluated in this manner for any given instant of time, without knowing all the history and future of the full dynamics, unless the background profile is static (i.e. isotropic and homogeneous). Nevertheless, it is still suggestive to consider what an inhomogeneous background particle distribution could bring (quantitatively) to this story; hence, we offer an ad hoc derivation here.

The key conceptual requirement to replace equation (4) with one that can be discretized for an arbitrary $\mathcal{N}$ is to re-expand the integral that gave rise to $\Delta\boldsymbol{v}_{\parallel}$ (equation 3) to explicitly account for the contributions of particles at different distances s along their two-body encounter trajectory, i.e. taking $\Delta\boldsymbol{v}_{\parallel} \rightarrow \int \langle \mathrm{d}\Delta\boldsymbol{v}_{\parallel}/\mathrm{d}s\rangle_{\rm deflected}\,\mathrm{d}s$ (see camparison in Fig. 1). Recall that the entire point of our derivation is to develop a formula that can be applied where the *explicit* N-body evolution of the mass M was not followed. Since DF fundamentally arises from the 'back-reaction' of the medium (i.e. the deflection of mass m as it feels gravity from M creating a net 'drag'), we need to identify the *difference* between the contribution to $\mathrm{d}\boldsymbol{v}_M/\mathrm{d}t$ that m *would have* at a distance r along its encounter trajectory with M if it had indeed been deflected by M, relative to the acceleration M would feel if it saw m on an 'undeflected' trajectory. The latter is, of course, just the 'normal' gravitational acceleration on M.[1] The full expressions for this are quite cumbersome and cannot be analytically closed; however, they are still, in any case, approximate (as we still ignore many effects such as other influences on the orbit of m during each stage of its two-body encounter), so we can safely approximate them to the same order of accuracy by noting that asymptotically $\langle \mathrm{d}\Delta\boldsymbol{v}_{\parallel}/\mathrm{d}s\rangle_{\rm deflected} \rightarrow \Delta\boldsymbol{v}_{\parallel}\,b^2/2\,(s^2+b^2)^{3/2}$ at large $r \gg b$ (noting $r^2 \equiv s^2+b^2$), and (for weak encounters, the only case where our derivation is meaningful) near pericentre [$r = b\,(1+\epsilon)$ with $\epsilon \ll 1$] $\langle \mathrm{d}\Delta\boldsymbol{v}_{\parallel}/\mathrm{d}s\rangle_{\rm deflected} \rightarrow \Delta\boldsymbol{v}_{\parallel}\,(1/2\,b)$. Together with the identity $1 = (b/2)\int_{-\infty}^{+\infty} b\,\mathrm{d}s/(s^2+b^2)^{3/2}$, we can replace $\Delta\boldsymbol{v}_{\parallel}$ in equation (4) with this expression, giving

$$\begin{aligned}\boldsymbol{a}_{\rm df} &= \int \frac{2\,\alpha\,G\,m\,\mathcal{N}(\boldsymbol{x},\boldsymbol{v})}{b\,(1+\alpha^2)}\,\hat{\boldsymbol{V}}\,\mathrm{d}p\,\mathrm{d}q\,\mathrm{d}^3\boldsymbol{v}\,\frac{b}{2}\int_s \frac{b\,\mathrm{d}s}{(s^2+b^2)^{3/2}}\\ &\approx \int\int_s \frac{2\,\alpha\,G\,m\,\mathcal{N}(\boldsymbol{x},\boldsymbol{v})}{b\,(1+\alpha^2)}\,\hat{\boldsymbol{V}}\,\mathrm{d}p\,\mathrm{d}q\,\mathrm{d}^3\boldsymbol{v}\,\frac{b}{2}\,\frac{b\,\mathrm{d}s}{(s^2+b^2)^{3/2}}\\ &= \int \frac{\alpha\,b\,G\,m}{(1+\alpha^2)\,r^3}\,\hat{\boldsymbol{V}}\,\mathcal{N}(\boldsymbol{x},\boldsymbol{v})\,\mathrm{d}^3\boldsymbol{x}\,\mathrm{d}^3\boldsymbol{v}, \end{aligned} \tag{5}$$

where we used $\mathrm{d}s\,\mathrm{d}p\,\mathrm{d}q \equiv \mathrm{d}^3\boldsymbol{x}$, and in the $\approx$ step, where we move the integrand, we essentially make a much weaker version of the original C43 approximation, assuming that quantities such as $\mathcal{N}$ do not vary strongly over the time-scale during which most of the $\Delta\boldsymbol{v}_{\parallel}$ is imparted by each two-body encounter. Now, we can insert the discrete N-body form of $\mathcal{N}$ as a sum of δ functions to trivially obtain

$$\begin{aligned}\boldsymbol{a}_{\rm df} &\rightarrow \sum_i \frac{\alpha_i\,b_i\,G\,\Delta m_i}{(1+\alpha_i^2)\,r_i^3}\,\hat{\boldsymbol{V}}_i\\ &= \sum_i \left(\frac{\alpha_i}{1+\alpha_i^2}\right)\left(\frac{b_i}{r_i}\right)\left(\frac{G\,\Delta m_i}{r_i^2}\right)\hat{\boldsymbol{V}}_i. \end{aligned} \tag{6}$$

We have of course made a number of assumptions to derive equation (6), and our final expression is not necessarily unique. However, it has many useful properties. (1) In a spatially homogeneous medium [i.e. anywhere we can write $\mathcal{N} = n\,f(\boldsymbol{v})$], it is trivial to verify by inserting this into equation (5) that equation (6) reproduces *exactly* the expressions from C43 for any $f(\boldsymbol{v})$. (2) equation (6), as intended, can be easily applied to an arbitrary N-body simulation collection of particles of arbitrary types [summing different components such as DM, gas, or stars simply involves carrying out the sum in equation (6) with the appropriate Δm_i and m for each 'species']. (3) equation (6) removes a number of ambiguities: the Coulomb logarithm is removed (it only 'reappears' if indeed the medium is infinite and homogeneous), and $\boldsymbol{V}$ that appears is unambiguous (discussed further below). (4) equation (6) above can be trivially generalized for softened gravity (below). (5) equation (6) at least asymptotically captures the relative contributions of near versus far particles m to the DF force, i.e. the dimensional scaling with r, e.g. correctly capturing the fact that most of the effect comes from when particles are near pericentre.

2.2 Force softening

To apply equation (6) to numerical simulations, we must account for force softening as in the simulations (since an N-body particle of mass Δm_i represents many individual stars, collocating them at a specific $\boldsymbol{x}_i$, $\boldsymbol{v}_i$ would lead to spurious divergences in the forces). In equation (6), note that all but one term are well behaved: $0 < \alpha_i/(1+\alpha_i^2) < 1/2$, $0 < b_i/r_i < 1$, and $|\hat{\boldsymbol{V}}| = 1$, so numerical divergence entirely arises from the term $G\,\Delta m_i/r_i^2$. However, this is just the Newtonian gravity from a point N-body particle – i.e. *exactly* the same term that is force softened in the simulations. Hence, we insert the same softening kernel $S_i(r_i)$ as used in the actual N-body simulation [taking $G\,\Delta m_i/r_i^2 \rightarrow S_i(r_i)\,G\,\Delta m_i/r_i^2$].

For the specific simulations here, this follows from the adaptive gravitational softening scheme described in Hopkins (2015), corresponding to a cubic spline mass distribution:

$$S_i(r_i) = \begin{cases} \frac{32}{3}q_i^3 - \frac{192}{5}q_i^5 + 32q_i^6 & 0 \le q_i < \frac{1}{2} \\ -\frac{1}{15} + \frac{64}{3}q_i^3 - 48q_i^4 & \\ \quad + \frac{192}{5}q_i^5 - \frac{32}{3}q_i^6 & \frac{1}{2} \le q_i < 1 \\ 1 & q_i \ge 1 \end{cases}, \tag{7}$$

where $q_i \equiv r_i/H_i$, with $H_i \approx 2.8\,\epsilon_i$ the radius of compact support of the kernel and ϵ_i the equivalent Plummer softening. This removes the numerical divergence and gives the correct result for a uniform density distribution sampled by N-body particles.[2]

[1] This contribution will differ depending on the sign of s at a given r, i.e. depending on whether m is 'approaching' or 'receding' from M; however, in our application to N-body simulations, the sign of $\boldsymbol{V}$ for distant m will change frequently, so there is no way to unique identify 'approaching' or 'receding' elements without actually performing the full time integral of every encounter (i.e. doing the full 'live' N-body calculation with M, exactly what we wish to avoid). We therefore simply average between the two, giving $\langle \mathrm{d}\Delta\boldsymbol{v}_{\parallel}/\mathrm{d}s\rangle_{\rm deflected} \equiv (1/2\,|\mathrm{d}s|)\,[\int_{-s-\mathrm{d}s}^{-s}(\boldsymbol{a}' - \boldsymbol{a}^0)\,\mathrm{d}t + \int_s^{s+\mathrm{d}s}(\boldsymbol{a}' - \boldsymbol{a}^0)\,\mathrm{d}t]$, where $\boldsymbol{a}' \equiv \boldsymbol{a}_{Mm}[\boldsymbol{x}_M^{\rm deflected}(t), \boldsymbol{x}_m^{\rm deflected}(t)]$ and $\boldsymbol{a}^0 \equiv \boldsymbol{a}_{Mm}[\boldsymbol{x}_M^{m-\rm undeflected}(t), \boldsymbol{x}_m^{\rm undeflected}(t)]$ are the two-body accelerations assuming m follows the deflected and undeflected trajectories, respectively (note that M still 'sees' m in its undeflected trajectory, but m does not 'see' M in that case).

[2] Note that in principle this softening is not exactly self-consistent with our derivation, since if Δm_i represents an extended spatial distribution of particles, each would be deflected slightly differently in equation (5). However, this is consistent with the simulations: N-body softening for collisionless fluids simply features this ambiguity at a fundamental level, because an individual N-body particle cannot actually deform in a fully Lagrangian manner.

2.3 Perpendicular force

In the above, we only included the parallel DF term ($\propto \hat{\boldsymbol{V}}_i$). However, two-body encounters also produce a perpendicular deflection $\boldsymbol{a}_{\rm df,\perp}$; this only vanishes in the C43 derivation because of the assumption of a homogeneous $\mathcal{N}$ (giving exact cancellation). Because we do not assume homogeneous $\mathcal{N}$, we can (if desired) retain these terms, giving

$$\boldsymbol{a}_{\rm df,\perp} = -\sum_i \left(\frac{1}{1+\alpha_i^2}\right)\left(\frac{b_i}{r_i}\right)\left(S_i(r_i)\,\frac{G\,\Delta m_i}{r_i^2}\right)\hat{\boldsymbol{b}}_i$$
$$\boldsymbol{b}_i \equiv \boldsymbol{r}_i - (\boldsymbol{r}_i\cdot\hat{\boldsymbol{V}}_i)\,\hat{\boldsymbol{V}}_i, \quad (8)$$

This differs from the parallel $\boldsymbol{a}_{\rm df,\parallel}$ only by one power of α_i and, of course, the direction. The power of α_i means that the perpendicular deflection can be stronger (compared to the parallel term) in strong encounters [although $0 < 1/(1+\alpha_i^2) < 1$, so this term is still bounded and cannot produce spurious divergences or forces larger than the regular/external acceleration]. However, because the integrated force is always dominated by weak deflections (where $\alpha_i \gg 1$), then even ignoring cancellations (which further reduce $\boldsymbol{a}_{\rm df,\perp}$ even in inhomogeneous $\mathcal{N}$), this term is generally smaller than the parallel $|\boldsymbol{a}_{\rm df,\parallel}|$ by one power of $\sim G\,M/r\,V^2 \sim M/M_{\rm total,galaxy}(<r) \ll 1$.

We show in an additional set of tests that this term is completely negligible for most galaxy simulation contexts; hence, we do not include them in our final expression and tests below. However, we emphasize that it is trivial to include and imposes no additional cost.

2.4 Final expression

It is straightforward to generalize the above for a spectrum of masses m, i.e. integrating over the stellar initial mass function. However, for any $M \gtrsim 10\,{\rm M}_\odot$, this makes a negligible difference to our results. Since we do not know the 'true' DM particle mass, it is more straightforward to simply assume the limit $M \gg m$, in which case the species masses m completely factor out of the salient expressions. This gives the expression we will use throughout:

$$\boldsymbol{a}_{\rm df} = \sum_i \Delta\boldsymbol{a}^i_{\rm df}$$
$$\Delta\boldsymbol{a}^i_{\rm df} \equiv \left(\frac{\alpha_i\,b_i}{(1+\alpha_i^2)\,r_i}\right)\left(S_i(r_i)\,\frac{G\,\Delta m_i}{r_i^2}\right)\hat{\boldsymbol{V}}_i, \quad (9)$$

with $\alpha_i \approx b_i\,V_i^2/G\,M$.

2.5 Numerical implementation

In the form of equation (9), it is particularly straightforward to implement our estimator. First, noting that α_i and $b_i \equiv r_i\,|\hat{\boldsymbol{r}}_i - (\hat{\boldsymbol{r}}_i\cdot\hat{\boldsymbol{V}}_i)\,\hat{\boldsymbol{V}}_i|$ are a function only of $\boldsymbol{r}_i$ and $\boldsymbol{V}_i$, we see that the only piece of additional information needed to compute equation (9), alongside the usual gravity force, in an N-body solver is the velocity $\boldsymbol{V}$ (already known). In other words, we do not need to construct some estimator for values in the C43 formula, like ρ, Λ, and $\langle V\rangle$ that are not actually computed in standard N-body simulations. Secondly, we also immediately see that it is completely trivial to carry out this sum over any arbitrary set of species (e.g. stars + gas + DM + other BHs).

Comparing the form of equation (9) and the 'regular' gravitational acceleration $\boldsymbol{a}_{\rm ext}$:

$$\boldsymbol{a}_M = \boldsymbol{a}_{\rm ext} + \boldsymbol{a}_{\rm df} = \textstyle\sum_i \Delta\boldsymbol{a}^i_{\rm ext} + \sum_i \Delta\boldsymbol{a}^i_{\rm df} \quad (10)$$

$$\Delta\boldsymbol{a}^i_{\rm ext} \equiv \left(S_i(r_i)\,\frac{G\,\Delta m_i}{r_i^2}\right)\hat{\boldsymbol{r}}_i \quad (11)$$

$$\Delta\boldsymbol{a}^i_{\rm df} \equiv \left(\frac{\alpha_i\,b_i}{(1+\alpha_i^2)\,r_i}\right)\left(S_i(r_i)\,\frac{G\,\Delta m_i}{r_i^2}\right)\hat{\boldsymbol{V}}_i, \quad (12)$$

we immediately see that the operation needed to compute $\boldsymbol{a}_{\rm df}$ is algorithmically identical to that needed to compute the normal gravitational forces. In tree-gravity, tree-PM, direct N-body, or many other methods, implementing exact evaluation of equation (9) in a manifestly conservative manner is especially trivial.[3] In e.g. a tree-walk, as one sums up to compute $\boldsymbol{a}_{\rm ext}$, we simply sum the additional term $\Delta\boldsymbol{a}^i_{\rm df}$, which scales exactly with $|\Delta\boldsymbol{a}^i_{\rm ext}|$ multiplied by the numerical pre-factor $\alpha_i\,b_i/(1+\alpha_i^2)\,r_i$, and oriented in the different direction $\hat{\boldsymbol{V}}_i$. The gravitational force softening is also naturally embedded in equation (9).

Moreover, our equation (9) is well behaved when applied to tree nodes/leaves, not just individual particles: one simply treats each node as a 'superparticle' with the appropriate total Δm_i and mass-averaged $\boldsymbol{V}_i$, $\boldsymbol{r}_i$, in the same manner as done for the usual gravity calculation. It is trivial to verify from the form of equation (9) that the order of the errors from this approach will always be equal to or better than the order of errors in $\boldsymbol{a}_{\rm ext}$ in the tree (i.e. convergence is equal or faster).

To ensure manifest momentum conservation, we simply enforce equal and opposite forces, i.e. apply an acceleration $\Delta\boldsymbol{a}_{M\text{-to-}i} = -(M/\Delta m_i)\,\Delta\boldsymbol{a}^i_{\rm df}$ to each particle i. The scaling of the pre-factor in equation (9) is such that it guarantees this 'back-reaction' term is well behaved and does not produce any spurious numerical divergences in the accelerations of the particles i.[4]

3 NUMERICAL VALIDATION: METHODOLOGY

To study the accuracy of our DF formula, we compare it to both direct high-resolution simulations and calibrated versions of the local Chandrasekhar's DF formula, using both 'on-the-fly' applications in simulations (Section 3.1) and post-processing methods (Section 3.2). Here, we provide details of those methods. In what follows, we refer to the 'target' or 'sinking' particle as a BH of mass $M_{\rm BH}$, since this is a particularly relevant motivating case for our sub-grid model, but of course the 'target' particle could in principle represent any sufficiently compact bound massive object.

3.1 On-the-fly simulations

3.1.1 Numerical methods

We have implemented the 'discrete DF estimator' equation (9) in the GIZMO multiphysics code (Hopkins 2015), which uses a standard Barnes–Hut tree algorithm to solve the gravity equations (an improved version of that in Springel 2005). GIZMO is well tested in numerous applications of N-body dynamics problems involving

[3] In PM and related methods, where long-range forces are evaluated via computing the potential from a particle-mesh Fourier method, implementing equation (9) is less trivial: the issue is that the direction $\hat{\boldsymbol{V}}_i$ differs from $\hat{\boldsymbol{r}}_i$, so one cannot simply treat $\boldsymbol{a}_{\rm sf}$ as a scalar correction to the regular external gravitational potential, but must compute a separate potential/field. However, in hybrid tree-PM methods, such as (optionally) implemented in GIZMO, the less accurate PM forces are only used at large distances; given this, we find (consistent with Fig. 7) that the errors from simply truncating the sum for $\boldsymbol{a}_{\rm df}$ by including only the contributions from the tree-walk (ignoring the PM terms in $\boldsymbol{a}_{\rm df}$) are entirely negligible (below normal integration-error level).

[4] That behaviour is *not* guaranteed if one attempts to conserve momentum by simply applying a C43-style formula to M and then ad hoc 'redistribute' the equal-and-opposite momentum change to the neighbouring i around M.

DF, *N*-body resonances, and wake problems (see e.g. Łokas 2019; Bonetti et al. 2020, 2021; Collier 2020; Grudić & Hopkins 2020; Morton, Khochfar & Oñorbe 2021; Bortolas et al. 2022), to which we refer for more detailed descriptions of numerical methods and demonstrations of convergence, test problems, etc. As described earlier, we simply evaluate the DF force $\boldsymbol{a}_{\rm df}$ alongside the 'normal' gravitational force (using the identical softening, etc.) in the tree-walk operation, imposing negligible CPU cost.

3.1.2 Initial conditions

To test the estimator, we have run a series of test problems. In each, we initialize a steady-state 'halo' of collisionless particles (e.g. 'DM' or 'stars') using the GALIC code (Yurin & Springel 2014), with a target/BH particle on an initial orbit expected to decay owing to DF. We have experimented with several different choices for the initial halo density profile, whether the halo velocity distribution is anisotropic or isotropic, and other parameters of the halo and orbit (e.g. eccentric versus circular, and initial position/energy/angular momentum). Our qualitative conclusions and comparison of methods are identical in each case (and of course, this being a pure *N*-body problem it is scale free), so we focus on and show plots from one example with typical cosmological units for the sake of clarity.

In our fiducial example, we adopt a Hernquist (1990)-profile halo with a total mass of $2\times10^{11}\,{\rm M}_{\odot}$ with the Yurin & Springel (2014) concentration parameter of 4 and spin parameter 0.04 [consistent with typical DM halo parameters (Bullock et al. 2001), and sufficient to make the halo mildly anisotropic because of rotation], so that the Hernquist (1990) scale length $a = 30.2$ kpc. The target/BH is placed 5 kpc away from the halo centre and has a tangential velocity of $59\,{\rm km\,s^{-1}}$, which is the circular velocity of the halo at that radius. The BH mass is $10^{8}\,{\rm M}_{\odot}$, much less than the enclosed DM mass inside 5 kpc ($\sim4\times10^{9}\,{\rm M}_{\odot}$), to avoid disrupting the dynamical equilibrium of the galaxy.

3.1.3 Sub-grid versus resolved simulations

As DF should be fully resolved when the target/BH mass $M_{\rm BH}$ is much more than the background ('dark matter' or DM) particle mass $M_{\rm DM}$, one would expect that only in a low-resolution simulation (i.e. $M_{\rm BH}\lesssim M_{\rm DM}$) a sub-grid treatment of DF is necessary.[5] However, if the resolution is too low, the orbital semimajor axis of the BH particle will be smaller than the inter-particle spacing of the *N*-body simulation and the BH will have essentially 'sunk to the centre' already – trivially, if it were just one background/DM particle inside of the initial 5 kpc, then there is no definable smaller scale centre towards which even a 'perfect' sub-grid model could migrate the target/BH. We hence choose $M_{\rm DM}=10^{7}\,{\rm M}_{\odot}$ in the tests with sub-grid DF, so ∼400 DM particles are enclosed inside the initial 5 kpc. We further run a set of 50 simulations with the same background halo, but with the BH particles placed randomly on a 5 kpc-radius sphere with a random direction of velocity in the tangent plane. By choosing the median between these runs, we can smooth out the chaotic motions intrinsic in the problem, as well as the effects of anisotropy (both real, from the halo rotation, and numerical, from *N*-body noise) generating eccentric orbits that produce larger oscillations in the instantaneous BH speed (making the results more difficult to read).

To test our results, we compare a set of reference simulations at varying resolution that do not adopt any sub-grid DF, but with the same set-ups of BH initial conditions. At sufficiently high resolution, these simulations satisfy $M_{\rm BH}\gg M_{\rm DM}$ and so should directly capture the salient effects of DF on the target.

3.1.4 Simulations with a 'fitted' C43 sub-grid model

Finally, we consider a third set of simulations where we again adopt a sub-grid DF estimator, but instead adopt the local Chandrasekhar DF estimator of equation (2) as previously introduced in GADGET in e.g. Cox et al. (2006) updated to be essentially identical to that in Tremmel et al. (2015). Here, we assume a Maxwellian velocity distribution, estimate the mean velocity and dispersion as a kernel-and-cell-mass-weighted mean, and use the BH kernel density estimator from Wellons et al. (2022) to estimate ρ.

We previously noted intrinsic difficulties this method faces: however, for this particular test problem, the background halo is (by construction) smooth and nearly isotropic and single-component and nearly Maxwellian, so this provides a 'best-case scenario' for a C43-like estimator. However, this still leaves unresolved the question of how to estimate the Coulomb logarithm. We find that common choices (e.g. the ratio of virial radius to 'true' inter-particle spacing) are not only impossible to predict a priori in a completely general simulation (they must be put in 'by hand'), but also appear to give DF forces that differ systematically from the resolved solutions by tens of per cent or up to a factor of 2. Therefore, to give this model the best possible chance, we explicitly *fit* the Coulomb logarithm, varying it until we find a model that best matches the BH orbital decay seen in the explicit high-resolution *N*-body calculation. We use this, essentially as a way of detecting how our method compares to a 'best-case' C43 estimator calibrated ahead of time to the *specific* problem being simulated.

The simulation set-ups are summarized in Table 1.

3.2 Post-processing in multiphysics galaxy simulations

While comparing our discrete estimator with the Chandrasekhar estimator in the above idealized test problem can help to test its accuracy, it is of course also important to apply it to some more 'realistic' (or at least more complicated) galaxy simulations that involve multicomponent (gas + star + DM) anisotropic, highly inhomogeneous backgrounds. Full applications to such simulations on the fly can be used to make predictions for e.g. demographics of free-floating BHs, IMBHs, and rates of BH–BH coalescence in galaxy nuclei (e.g. predictions for the Laser Interferometer Space Antenna (LISA)). However, this is clearly beyond the scope of this work. Instead, here we will select some snapshots of *N*-body information from high-redshift galaxies in the Feedback In Realistic Environments (Hopkins et al. 2014, 2018) project, and use these to make some simple post-processing comparisons in order to see how the full on-the-fly application of the estimator used here might differ (or not) from other approaches to including or ignoring DF in these kinds of systems.

4 RESULTS AND DISCUSSIONS

4.1 Validation in on-the-fly simulations

Figs 2–3 show some representative results of our numerical validation

[5]When the resolution lies in between and DF is partially resolved, a sub-grid treatment may cause 'double counting' when calculating DF. While this remains an open question in general, we find that it can be avoided by multiplying a field-mass-dependent function on the DF formula in our tests. See discussions in Section 4.3.

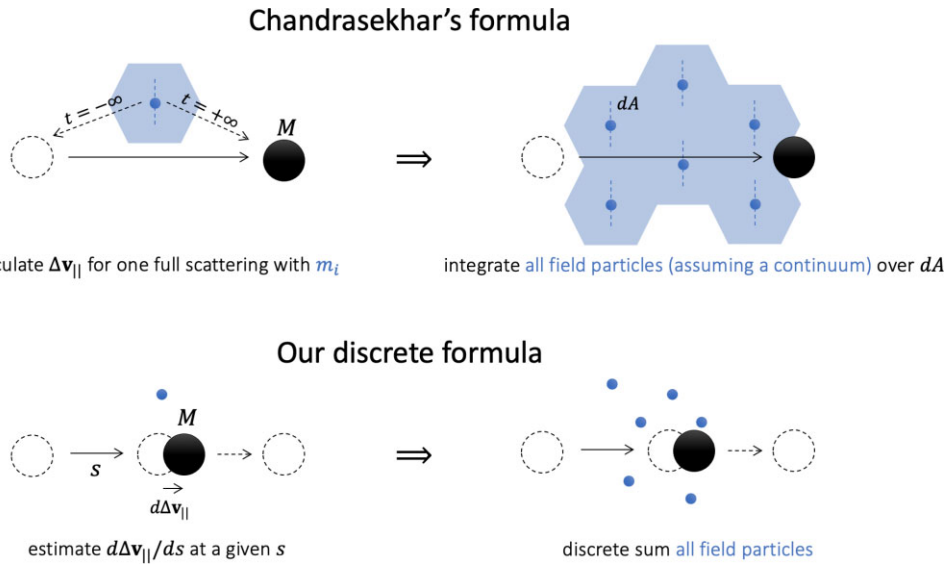

Figure 1. A comparison between the derivation of Chandrasekhar's DF formula (C43) and ours: in C43, Chandrasekhar calculates the change of velocity $\Delta \boldsymbol{v}_{||}$ for one full scattering, and integrates over the remaining two dimensions dA (perpendicular to the direction of motion) for all field particles, assuming a homogeneous and isotropic continuum such that the overall contribution is characterized by the Coulomb logarithm; in our derivation, we estimate the change of velocity d$\boldsymbol{v}_{||}$/ds over the line of motion (coordinated by s) at a given point in the scattering process, which allows us to integrate all field particles over the full configuration space (as the dimension along the line of motion is now recovered), such that a discrete numerical sum is possible.

Table 1. Representative simulation summary for our idealized tests. Different sets share the same set-up of initial conditions: a $10^8\,M_\odot$ BH particle placed randomly at a 5 kpc radius with a velocity of 59 km s^{-1} in a random tangent direction. The background particles form a Hernquist halo with $M_{\rm halo} = 2\times10^{11}\,M_\odot$. The BH speeds from these tests are shown in Fig. 3 (when multiple runs are present, only the median value is shown).

Set	$M_{\rm DM}/M_{\rm BH}$	Sub-grid DF model	ϵ Criterion	No. of runs
1	10^{-1}	This paper	$\epsilon \sim \Delta x_i$	50
2	10^{-1}	Fitted C43	$\epsilon \sim \Delta x_i$	50
3	10^{0}	None	$\epsilon < b_{\rm min}$	50
4	10^{-1}	None	$\epsilon < b_{\rm min}$	50
5	10^{-2}	None	$\epsilon < b_{\rm min}$	50
6	10^{-3}	None	$\epsilon < b_{\rm min}$	1
7	10^{-4}	None	$\epsilon < b_{\rm min}$	1
8	10^{0}	None	$\epsilon \sim \Delta x_i$	50
9	10^{-1}	None	$\epsilon \sim \Delta x_i$	50
10	10^{-2}	None	$\epsilon \sim \Delta x_i$	50
11	10^{-3}	None	$\epsilon \sim \Delta x_i$	1
12	10^{-4}	None	$\epsilon \sim \Delta x_i$	1
13	10^{-5}	None	$\epsilon \sim \Delta x_i$	1

tests in on-the-fly simulations, specifically focusing on an illustrative trajectory of the BHs as well as the BH velocity as a function of time.

First, we examine the behaviour of pure N-body calculations (*without* sub-grid DF) as a function of resolution. Not surprisingly, when the target mass is similar to the N-body particles (e.g. $m_{\rm DM} \gtrsim M_{\rm BH}$), no DF is captured. Most previous studies arguing for different 'sufficient' resolutions to capture DF refer to this regime (see e.g. van den Bosch et al. 1999; Colpi et al. 2007; Boylan-Kolchin, Ma & Quataert 2008; Hopkins et al. 2018; Pfister et al. 2019; Barausse et al. 2020; Boldrini, Mohayaee & Silk 2020; Ma et al. 2021), depending on the specific problems they are choosing. In our case, at better resolution ($m_{\rm DM} \ll M_{\rm BH}$) we see DF but with an important dependence on how we treat the *spatial* force softening ϵ. If we adopt a fixed Plummer-equivalent ϵ comparable to or smaller than the canonical minimum impact parameter for strong encounters $b_{\rm min} \sim G\,M_{\rm BH}/(2\,\sigma^2 + V_{\rm bh}^2)$ (here ∼60 pc at the initial BH position), we see excellent convergence once $m_{\rm DM} \ll 0.1\,M_{\rm BH}$ (Fig. 3, left-hand panel). However, this is not how force softenings are typically set in N-body simulations that do not resolve the individual point masses: instead, to prevent spurious noise in *other* properties, the 'optimal' softening is usually chosen to roughly match the inter-particle separation $\epsilon \sim \Delta x_i \sim (\Delta m_i/\rho_i)^{1/3}$ (Fig. 3, right-hand panel; Merritt 1996; Romeo 1998; Athanassoula et al. 2000; Dehnen 2001; Rodionov & Sotnikova 2005). When we do this, we see notably worse convergence: in fact, the convergence is logarithmic in $m_{\rm DM}$, because we have $\epsilon > b_{\rm min}$, and the effective Coulomb logarithm is artificially truncated (i.e. we artificially suppress close encounters). This is a known challenge for DF in softened gravity (see e.g. Karl et al. 2015, for more details and extended discussion), and it further emphasizes the importance of a sub-grid model like ours: achieving $\Delta x_i \ll b_{\rm min}$ requires $m_{\rm DM} \ll 10^{-5}\,M_{\rm BH}$, i.e. billions of N-body particles even for a simple, idealized halo like that here.

We then compare our 'sub-grid' DF model (equation 9) calculated on the fly to an extremely low-resolution initial condtion with $m_{\rm DM} = 0.1\,M_{\rm BH}$,[6] using $\epsilon \sim \Delta x_i$ as would be applied in typical cosmological simulations. For this low-resolution case, there is significant variation owing to different eccentric orbits and discreteness noise, so we

[6]We find that the results of our sub-grid DF runs are robust and nearly independent of resolution so long as the dynamical mass of the target/BH particle is at least slightly larger (a factor of ≳2–3) than the mass-weighted median of the 'background' N-body particles. If the BH particle has mass lighter than the background, then either sub-grid DF model (C43 or equation 9) requires additional care or else spurious N-body heating effects can become larger than the true DF forces. So for practical applications where one wishes to evolve the dynamics of targets with very small masses, it is useful to follow standard practice (Di Matteo et al. 2003; Springel & Hernquist 2003; Hopkins et al. 2005) and assign a separate 'true target/BH mass' used for the DF calculation and other physics to the N-body particle 'carrying' the target/BH.

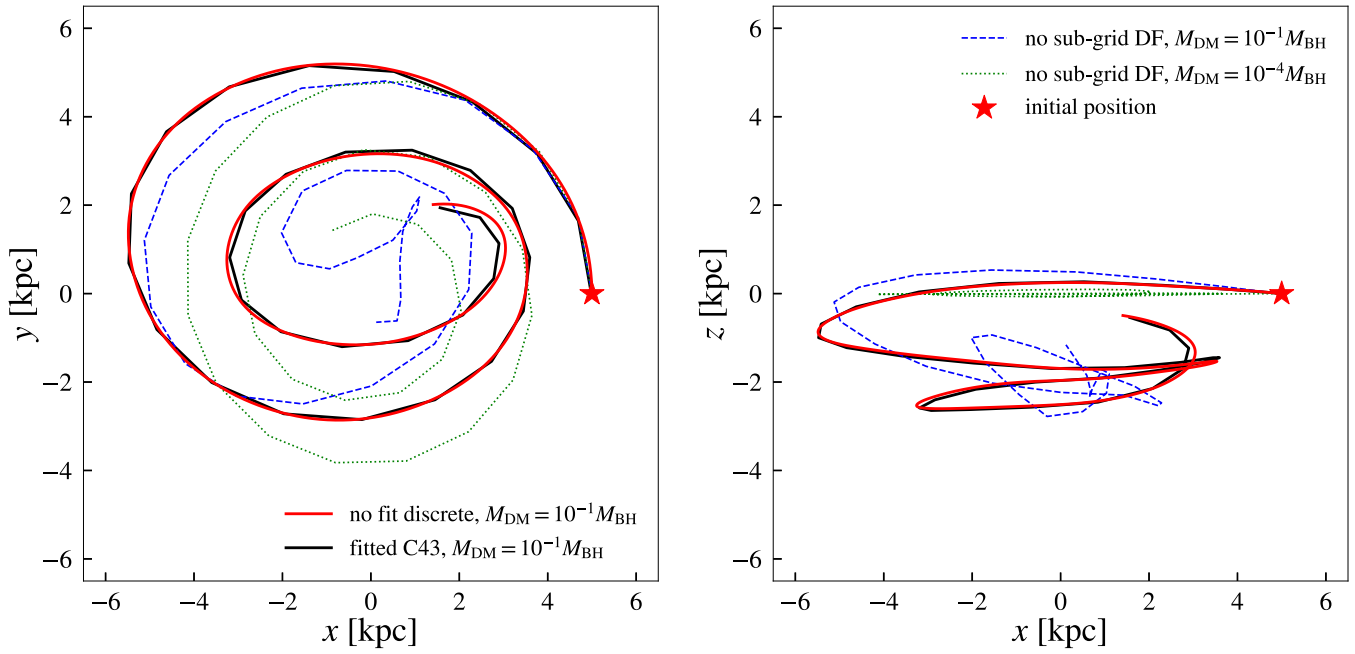

Figure 2. Example trajectories of a BH particle in our simulations. The BH is initially placed 5 kpc away from the halo centre (the coordinate origin) on the x-axis and has a circular velocity of 59 km s^{-1} in the $\hat{y}$ direction. We see that in the high-resolution run (green dashed) the BH sinks to the halo centre in circular orbits as time evolves, which is partially resolved by the low-resolution runs with sub-grid DF (red and black lines, with the discrete DF and the 'calibrated Chandrasekhar' estimator, respectively), but not by the run without it (blue dashed; the BH departs significantly from the halo centre in the z direction). The low-resolution runs (with or without sub-grid DF) suffer from dynamical heating that significantly perturbs the circular orbit. In addition, our discrete estimator matches well with the calibrated Chandrasekhar estimator – but we have not calibrated our discrete DF estimator in any way [we are simply using equation (9) directly, without any input parameter other than the smoothing length ϵ].

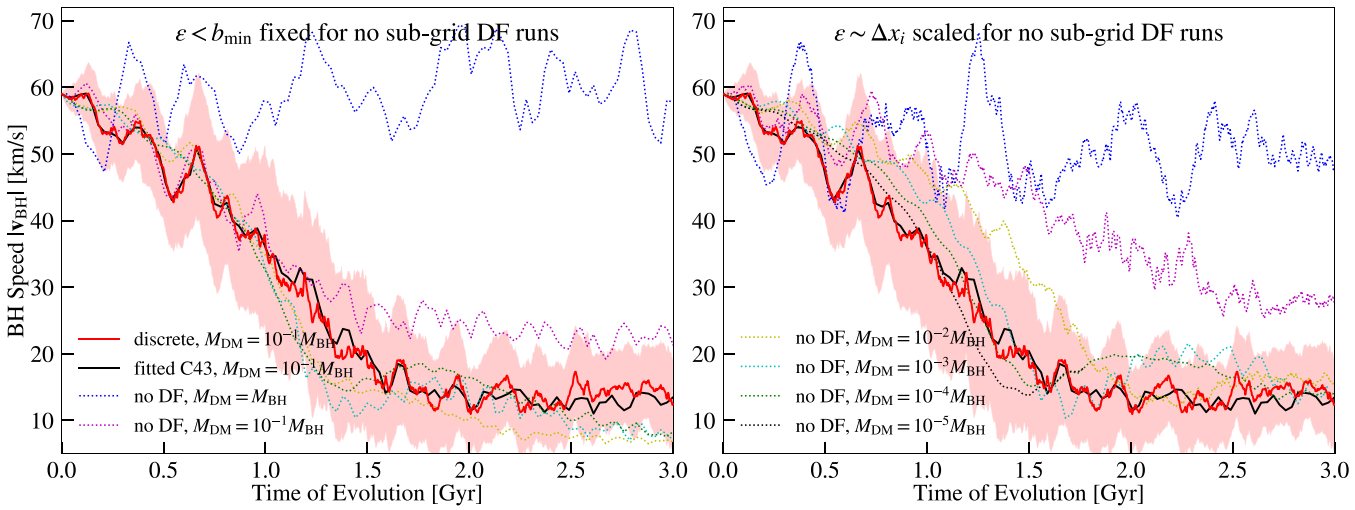

Figure 3. The speed of BH particles upon time of evolution in our test problem. The thick red, thick black, and dotted lines show the (median of) results from low-resolution runs with our discrete estimator, with Chandrasekhar's DF estimator (with a fitted $\ln \Lambda = 4$) and from multiresolution runs without sub-grid DF, respectively (see Table 1). The red shaded area shows the $\pm\sigma$ range for all runs with the discrete estimator. We see that the speed of BH particles decreases significantly as the BHs sink to the halo centre. The results from our discrete estimator match well with the fitted Chandrasekhar estimator, and match the converged results of the no sub-grid DF runs at higher resolution. The convergence is better for no sub-grid DF runs with smoothing length less than the minimum impact parameter ($\epsilon < b_{min}$, left-hand panel) than those with (the usually chosen) length comparable to the inter-particle separation ($\epsilon \sim \Delta x_i$, right-hand panel), as in the latter case the effective Coulomb logarithm is artificially truncated, causing a logarithmic convergence behaviour (see discussions in Section 4.1).

show the median and $\pm 1\sigma$ range of BH velocities. The median agrees remarkably well with the converged solution. We stress that equation (9) contains *no other adjustable parameter* beyond the physically motivated ϵ: this is an actual prediction.

Next, we compare the 'fitted' C43 model equation (2): as described earlier, in addition to the arbitrary choice of kernel estimator size and shape (which we set to the smallest size that reduces noise acceptably), we freely vary the numerical pre-factor ('effective Coulomb

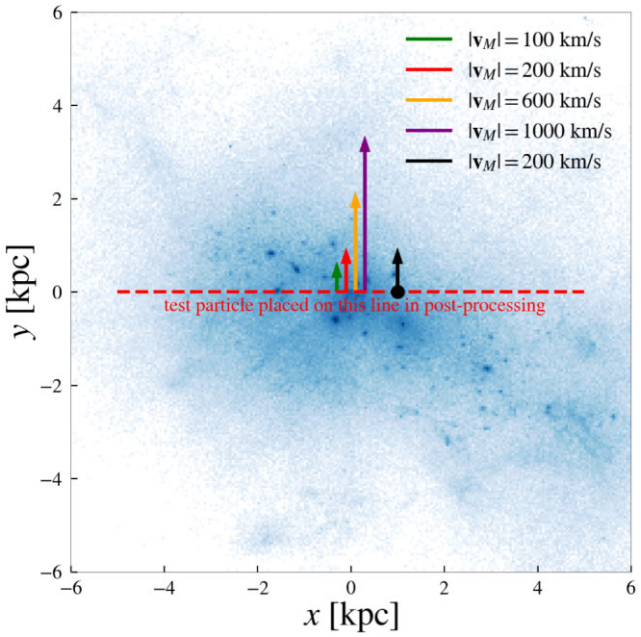

Figure 4. The galaxy snapshot we chose for post-processing analysis. The colour scales with the projected total mass density (DM + stars + gas). It is the '**z5m12b**' galaxy at redshift 7.0 described in Ma et al. (2021), which is clumpy and dynamically unstable. The post-processing tests are also shown: (a) a test particle of 10^5 $M_\odot$ placed on the x-axis with a velocity (V_M) of 200 km s^{-1} $\hat{y}$; (b) the same particle but with different velocities (100, 200, 600, and 1000 km s^{-1}) in the $\hat{y}$ direction; and (c) a fixed particle at (1, 0, 0) with $V_M = 200$ km s^{-1} $\hat{y}$. The implications of these tests are described in the main text.

log') $\ln\Lambda$ in equation (2) until we find a value that best matches our high-resolution simulations. For the best-fitting value, the result is strikingly similar to our equation (9) (perhaps not surprising, given that we start from the same assumptions) – but we stress that even small, ∼10 per cent, differences in $\ln\Lambda$ produce significant disagreement with the high-resolution simulations. Moreover, we have considered a dozen 'standard' estimates of $\ln\Lambda$ widely used in the literature (see references above and Hashimoto, Funato & Makino 2003; Just et al. 2011; Antonini & Merritt 2012; Dosopoulou & Antonini 2017), e.g. $\Lambda \sim |\rho/\nabla\rho|/b_{\rm min}$, and find that *none* of them correctly predicts the best-fitting Λ (usually discrepant by factors of ∼1.3–2). This probably owes at least in part to the fact that the central Hernquist (1990) distribution function is appreciably non-Maxwellian, as discussed in Karl et al. (2015), so the fitted $\ln\Lambda$ is essentially compensating for this error [the 'erf(...)-...' term in equation 2].

As noted earlier, these conclusions are robust to the parameters of the initial halo and orbit, mass profile of the halo assumed, amount of angular momentum (anisotropy in the distribution function), and other choices of the problem set-up: however, we find as expected that the C43 'effective Coulomb logarithm' must be re-calibrated in many cases to fit high-resolution simulations. We have also tested other numerical aspects of the method, including the tree opening criteria (Power et al. 2003; Springel 2005), time-step size/integration accuracy (Hopkins et al. 2018; Grudić & Hopkins 2020), and inclusion/exclusion of the perpendicular force (equation 8): none of these has a significant effect (consistent with previous studies; see e.g. Just et al. 2011; Karl et al. 2015; Mukherjee et al. 2021).

Given the close agreement between the discrete DF and explicitly calibrated Chandrasekhar DF models, it is likely that more detailed

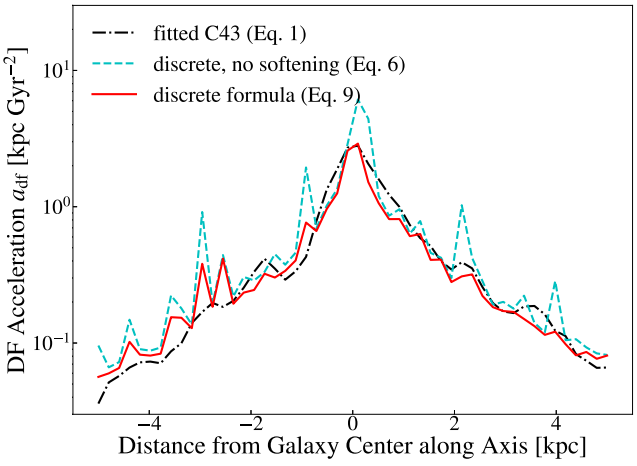

Figure 5. Comparison of the DF amplitude calculated from different DF formulas. The test particle is a 10^5 $M_\odot$ particle with a 200 km s^{-1} velocity in the y direction, put at different positions on the x-axis (Fig. 4, red dashed line and arrow). The black, cyan, and red lines show the results from Chandrasekhar's formula [equation 1, with a (fitted) constant Coulomb logarithm $\ln\Lambda = 5$], our formula without smoothing (equation 6), and our formula with smoothing (equation 9), respectively. Our discrete formula remains very close to Chandrasekhar's approximation. The smoothing removes most of the peaks, which could be caused by numerical divergence.

differences in orbit shape we see comparing *either* of these models and the true, high-resolution simulation owes not to anything we can simply 'further calibrate' (like a Coulomb logarithm), but rather to fundamental resolution effects (e.g. more accurately recovering the shape of the background potential itself, hence the 'correct' elliptical orbit structure, or the treatment of subparsec-scale physics around supermassive BHs, a known issue as discussed in e.g. Rantala et al. 2017; Mannerkoski et al. 2021, 2022), as well as assumptions of the Chandrasekhar-like derivation that our DF derivation also implicitly assumes, e.g., the assumption of linearity (that the net effect on the BH can be approximated via the sum of many independent two-body encounters) or forward/backward asymmetry in the distribution function (implicit in a stronger assumption like homogeneity but present in a weaker form in our derivation as well).

4.2 Post-processing in multiphysics galaxy simulations

While the idealized experiments given above are important for validation, their simplicity means that it is difficult to gain insight into possible differences between our equation (9) and the fitted C43 model. We therefore briefly consider this in post-processing of a multiphysics galaxy formation simulation. The specific (arbitrary) simulation and time we select is the '**z5m12b**' galaxy at redshift 7.0 described in Ma et al. (2021), illustrated in Fig. 4. The simulation is multicomponent, containing DM, stars, multiphase gas, and BHs, with complicated cooling, star formation, and 'feedback' physics all included on the fly. This particular snapshot is chosen because it is dynamically unstable, asymmetric, gas rich, and star forming, and contains several giant star clusters and molecular cloud complexes, all of which complicate the dynamics. We compare the results from the discrete estimator with the Chandrasekhar estimator and discuss their differences and implications.

In Fig. 5, we compare the acceleration amplitude $a_{\rm df} \equiv |\boldsymbol{a}_{\rm df}|$ calculated from different formulae for a test particle of mass 10^5 $M_\odot$ in the representative snapshot. The test particle is placed

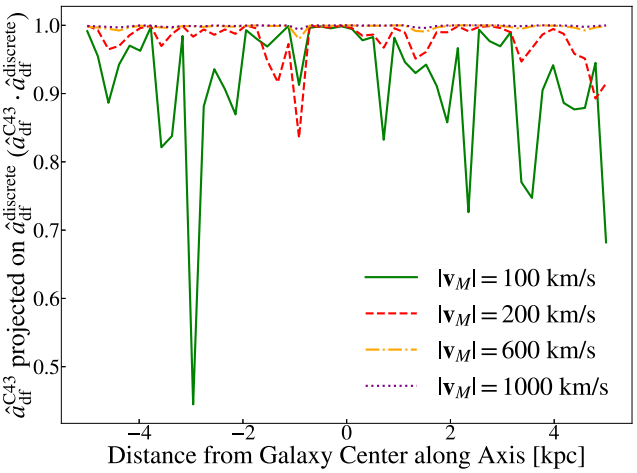

Figure 6. The unit direction vectors calculated from Chandrasekhar's projected on those from our DF formulas. The test particle is the same as in Fig. 5 but with different velocities (still along the y-axis). The difference of this value from 1 shows how misaligned the directions are. At most positions, the directions are perfectly aligned for relative high velocities, yet at low velocities huge error could occur due to contributions from particles far apart.

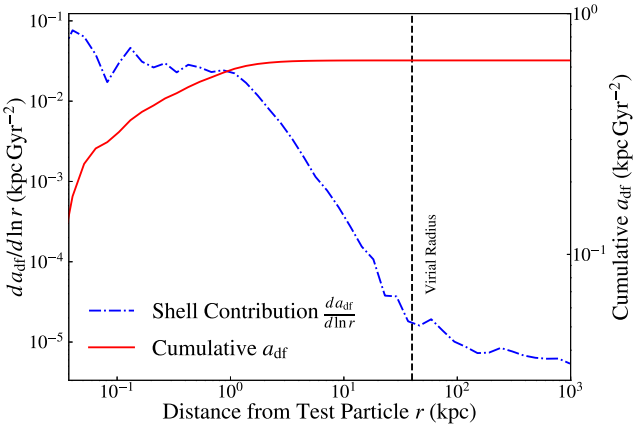

Figure 7. The contributions and cumulative results on DF from slices with different radii around a test particle. The test particle is a 10^5 M$_\odot$ particle at (0, 1, 0) with a 200 km s^{-1} velocity in the y direction. The particle's distance to the virial radius where we cut off the sum is labelled with a black dashed line. We see that contributions are mostly from slices near the particle, while those from slices outside the virial radius are $\gtrsim$1000 times lower, suggesting that our cut-off makes little difference.

along an arbitrary x-axis passing through the galaxy centre with a simulation-frame velocity of $\boldsymbol{V}_M = 200\,{\rm km\,s^{-1}}\,\hat{y}$ (Fig. 4, red dashed line and arrow). We compare the results from our 'full' expression (equation 9), our expression ignoring force softening (equation 6), and the classical C43 expression (equation 1). Equations (9) and (6) can be directly applied to the simulations without any processing. To apply equation (1), we estimate the continuous ρ at each position $\boldsymbol{x}_M$ using a kernel density estimator by averaging through the 0.4 kpc cubic box around $\boldsymbol{x}_M$; we calculate the local velocity integral by converting it into the usual discrete sum in this box, and we take $\ln\Lambda = 5$ to be constant, once again fitting it so that the median/mean acceleration is essentially identical.

The agreement between equations (9) and (1) is reasonable, but again this requires choosing Λ specific for the problem and snapshot (we note, for example, that the effective Λ here differs by almost a factor of 2 from the value fitted to the idealized Hernquist (1990) profile sphere tests in the previous section). Equation (6), which ignores force softening, is also quite similar, except for occasional 'spikes' arising from close proximity to N-body particles producing a spurious large force that is not actually present in the simulations (accounted for correctly in our equation 9).

Fig. 6 similarly compares the direction $\hat{\boldsymbol{a}}_{\rm df}$. Because C43 assume homogeneity, and their $\boldsymbol{a}_{\rm df}^{\rm C43}$ has equal contributions from all scales, a major ambiguity in equation (1) – even after we fit out the Coulomb logarithm – is where/how to evaluate $\boldsymbol{V} = \boldsymbol{V}_M - \boldsymbol{V}_m$. Should we interpolate to the local value at $\boldsymbol{x}_M$, weight by contribution to Λ, or weight by mass (dominated by distant particles)? If we follow the same procedure above to obtain a 'local' $\boldsymbol{V}$, then we see that usually the direction of $\hat{\boldsymbol{a}}_{\rm df}$ from equation (9) and that from equation (1) agree, especially if we assume a test particle M with large lab-frame $|\boldsymbol{V}_M|$ (since then $\boldsymbol{V} \approx \boldsymbol{V}_M$, independent of the background $\boldsymbol{v}_m$). However, when $\boldsymbol{V}_M$ is small (the case of interest for sinking), equation (1) can occasionally 'flip' to point in an unphysical direction in a noisy velocity field.

Our equation (9) allows us to easily quantify the contributions to the total $a_{\rm df}$ from all the mass in radial shells. Fig. 7 shows this (specifically $da_{\rm df}/d\ln r$, integrating the contributions from all particles in logarithmically spaced shells of distance r from M) again for a representative example (with M at $|\boldsymbol{x}_M| = 1$ kpc from the origin on the x-axis) in the same snapshot. At small scales ($r \lesssim 1$ kpc) around M, where the density field is *statistically* homogeneous (there are local fluctuations, but there is not a strong systematic dependence of density on distance r from M), we see the expected Coulomb log behaviour ($da_{\rm df}/d\ln r \sim$ constant). At larger radii, the contribution falls rapidly. We can, for example, truncate the sum in equation (9) at the virial radius (labelled) with negligible loss of accuracy. This is expected if the galaxy follows a realistic density profile, as in e.g. an isothermal sphere, the density is not constant, but at $r \gg |\boldsymbol{x}_M|$ falls rapidly ($\propto r^{-2}$, giving rapid convergence). As expected, the behaviour at larger r does motivate the value of $\ln\Lambda$ we fit: if we take $\Lambda = b_{\rm max}/b_{\rm min}$, with $b_{\rm min} \sim \max[(m/\rho)^{1/3},\ G\,M/V^2] \sim 1$ pc, and $b_{\rm max} \sim 1$ kpc, we obtain $\ln\Lambda \sim 7$, similar to our fitted value.

The above discussions are closely related to cases where the background field particles have a non-negligible physical bulk motion, like a wandering BH in a rotating disc-galaxy set-up. While studying such simulations in detail is beyond the scope of this work, we comment that the rotation of star particles in the disc could largely affect the strength and direction of DF, since their phase-space distribution departs significantly from homogeneity and isotropy. In an extremely dense galactic environment, we may expect that the local disc particles with similar circular velocities contribute most to the BH's DF, such that the BH is boosted by the field particles around it, which is similar to the case we already studied. For a less dense set-up, non-local (halo) particles with different circular velocities could be important, and their combined contribution to DF with local disc particles could make the BH dynamics more complicated. Our DF estimator, which applies to an arbitrary phase-space distribution and counts the DF contribution from each individual field particle, would be ideal for studying such problems. Such topics will be studied in future work.

Briefly, one might wonder whether on sufficiently large scales, where the universe becomes homogeneous and isotropic, $a_{\rm df}$ might begin to grow logarithmically again. However, even if we ignore finite speed-of-gravity effects (i.e. consider pure Newtonian gravity), on these scales the velocity must include the Hubble flow, so $\boldsymbol{v}_{\rm physical} = \boldsymbol{v}_{\rm peculiar} + H(z)\,\boldsymbol{r}$. In an isotropic pure Hubble-flow medium, the DF is

identically zero, as there is always equal-and-opposite contributions to $\boldsymbol{a}_{\rm df}$ from the fact that $\boldsymbol{V} \propto \boldsymbol{r}$ (i.e. because $\langle \boldsymbol{V} \rangle = 0$ on all scales). If we consider a Hubble flow plus peculiar velocities, then expanding equation (9) appropriate for large r where $\langle \rho(r) \rangle \sim$ constant and $H\,r \gg \langle |\boldsymbol{v}_{\rm peculiar}(r)|^2 \rangle^{1/2}$, the contributions to the sum take the form $\sum\, G^2\, M\, \langle |\boldsymbol{v}_{\rm peculiar}(r)|^2 \rangle^{1/2}\, \Delta m_i / H^3\, r^6 \propto \int \rho\, r^{-6}\, {\rm d}^3\boldsymbol{x}$, which converges rapidly as $r \rightarrow \infty$.

4.3 Interpolating the sub-grid model in simulations with variable masses

Finally, one can easily imagine situations such as cosmological simulations with a range of BH masses where the DF forces are well resolved for some targets (e.g. supermassive BHs with $M_{\rm BH} \sim 10^{10}\,{\rm M}_\odot$) but not others (e.g. lower mass BHs). In these cases, applying equation (9) to all BHs would 'double count' for some. A simple (albeit ad hoc) approach to avoid double counting is to multiply $\Delta \boldsymbol{a}^i_{\rm df}$ by a sigmoid or 'switch'-like function $g(\Delta m_i/M_j$, ...) that has the property $g(x$, ...) $\rightarrow 0$ for $x \rightarrow 0$ and $g(x$, ...) $\rightarrow 1$ for $x \rightarrow \infty$. It is beyond the scope of our paper here to develop and test such models, and from Fig. 3 we see one complication is that this should depend on how one treats the force softening (not just particle masses Δm_i), but a quick examination of the idealized tests in Section 4.1 with different $M_{\rm BH}$ suggests that (if we assume $\epsilon \sim \Delta x$, as usually adopted in such simulations) a simple function such as $g = \min(1.0, \max(0, (3/\log(M_{{\rm BH},\,j}/\Delta m_i) - 1)/1.6))$ works reasonably well. Another advantage of our equation (9) is that because it operates in pairwise fashion, it can naturally deal with simulations with a wide range of Δm_i (a common situation), while attempting to apply such a correction factor 'locally' to equation (1) leaves it ill-defined which value of Δm_i to use.

5 CONCLUSIONS

In numerical simulations, especially of star and galaxy formation, it is common to encounter the limit where DF should be experienced ($M \gtrsim m$) by some explicitly evolved objects M (e.g. BHs and massive stars), but it cannot be *numerically* resolved ($\Delta m_i \gtrsim M$). As a result, there have been several attempts to develop and apply 'on-the-fly' sub-grid DF models. Almost all of these amount to some attempt to calculate and apply the traditional C43 formula (equation 1) to the masses M at each time (see e.g. Colpi et al. 2007; Dotti et al. 2009; Tremmel et al. 2015; Pfister et al. 2019). However, this can introduce a number of problems in practice, namely the ambiguity of kernel-dependent locally defined quantities, inconsistency in applying force softening and momentum conservation, the semi-arbitrary choice of Coulomb logarithm, the necessity of assuming Maxwellian velocity distribution functions, and additional computational expenses for kernel estimates.

In this paper, we derive a new discrete expression for the DF force, $\boldsymbol{a}_{\rm df}$, given in equation (9). This formula is specifically designed for application to numerical simulations, either in post-processing or 'on the fly' when the DF forces cannot be resolved (e.g. when N-body particle masses are comparable to the BH mass M, as a 'sub-grid' DF model). While still approximate, this has a large number of advantages compared to the traditional C43 analytical expression, including (1) it allows for an arbitrary distribution function, without requiring an infinite homogeneous time-invariant medium with constant density, Maxwellian velocity distribution, etc. (but it does reduce *identically* to a discrete form of the C43 expression, when these assumptions are actually satisfied); (2) it is designed specifically for simulations, so it is represented only as a sum over quantities that are always well defined in the simulation for all N-body particles (e.g. positions, velocities, and masses), and does not require the expensive and fundamentally ill-defined evaluation of quantities such as a 'smoothed' density, background mean velocity/dispersion/distribution function, Coulomb logarithm, etc.; (3) it trivially incorporates force softening exactly consistent with how it is treated in code, and generalizes to arbitrary multicomponent N-body simulations with different species and an arbitrary range of particle masses; (4) it manifestly conserves total momentum, unlike N-body implementations of C43; and (5) it can be evaluated directly alongside the normal gravitational forces with negligible cost, and automatically inherits all of the desired convergence and accuracy properties of the N-body solver. We have implemented this 'live' evaluation of equation (9) in GIZMO, and verified that all of the properties above apply, that it agrees well with our N-body simulations, and that the computational overhead of evaluating it alongside gravity in the tree is immeasurably small.

There are still uncertainties in our work. In our derivation of the discrete formula, we inserted an approximate integral kernel, which is not necessarily unique or best behaved. We found that even if our discrete estimator closely agrees with the calibrated Chandrasekhar DF estimator in our test problems, it still differs from the high-resolution simulation results in terms of the detailed particle trajectories, which might be related to the fundamental Chandrasekhar-like assumptions we have made in our formula. We also note that it remains an open question how to accurately avoid 'double counting' when some of the DF may be captured self-consistently by the N-body code, while additional DF is modelled using our sub-grid model. This is especially the case when the system evolves (such as when supermassive BHs grow) and the fraction of 'resolved' DF changes with time. Future work will be needed to make improvements on these points.

ACKNOWLEDGEMENTS

We thank Sophia Taylor for early contributions to the development of the discrete DF estimator, and Daniel Anglés-Alcázar for useful discussions. Support for LM and PFH was provided by NSF Collaborative Research Grants 1715847 and 1911233,NSF CAREER grant 1455342, and NASA grants 80NSSC18K0562 and JPL 1589742. LZK was supported by NSF-AAG-1910209, and by the Research Corporation for Science Advancement through a Cottrell Fellowship Award. CAFG was supported by NSF through grants AST-1715216, AST-2108230, and CAREER award AST-1652522; by NASA through grants 17-ATP17-0067 and 21-ATP21-0036; by STScI through grants HST-AR-16124.001-A and HST-GO-16730.016-A; by CXO through grant TM2-23005X; and by the Research Corporation for Science Advancement through a Cottrell Scholar Award. Numerical calculations were run on the Caltech compute cluster 'Wheeler', allocations FTA-Hopkins supported by the NSF and TACC, and NASA HEC SMD-16-7592.

DATA AVAILABILITY

The data and source code supporting the plots within this article are available on reasonable request to the corresponding author.

REFERENCES

Alessandrini E., Lanzoni B., Miocchi P., Ciotti L., Ferraro F. R., 2014, ApJ, 795, 169
Antonini F., Merritt D., 2012, ApJ, 745, 83

Athanassoula E., Fady E., Lambert J. C., Bosma A., 2000, MNRAS, 314, 475
Barausse E., Dvorkin I., Tremmel M., Volonteri M., Bonetti M., 2020, ApJ, 904, 16
Binney J., Tremaine S., 2008, Galactic Dynamics, 2nd edn. Princeton Univ. Press, Princeton, NJ
Boldrini P., Mohayaee R., Silk J., 2020, MNRAS, 495, L12
Bonetti M., Bortolas E., Lupi A., Dotti M., Raimundo S. I., 2020, MNRAS, 494, 3053
Bonetti M., Bortolas E., Lupi A., Dotti M., 2021, MNRAS, 502, 3554
Bortolas E., Bonetti M., Dotti M., Lupi A., Capelo P. R., Mayer L., Sesana A., 2022, MNRAS, 512, 3365
Boylan-Kolchin M., Ma C.-P., Quataert E., 2008, MNRAS, 383, 93
Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001, ApJ, 555, 240
Chandrasekhar S., 1943, ApJ, 97, 255
Chen N., Ni Y., Tremmel M., Di Matteo T., Bird S., DeGraf C., Feng Y., 2022, MNRAS, 510, 531
Collier A., 2020, MNRAS, 492, 2241
Colpi M., Callegari S., Dotti M., Kazantzidis S., Mayer L., 2007, in di Salvo T., Israel G. L., Piersant L., Burderi L., Matt G., Tornambe A., Menna M. T., eds, AIP Conf. Proc. Vol. 924, The Multicolored Landscape of Compact Objects and Their Explosive Origins. Am. Inst. Phys., Italy, p. 705
Cox T. J., Di Matteo T., Hernquist L., Hopkins P. F., Robertson B., Springel V., 2006, ApJ, 643, 692
Dehnen W., 2001, MNRAS, 324, 273
Di Matteo T., Croft R. A. C., Springel V., Hernquist L., 2003, ApJ, 593, 56
Dosopoulou F., Antonini F., 2017, ApJ, 840, 31
Dotti M., Ruszkowski M., Paredi L., Colpi M., Volonteri M., Haardt F., 2009, MNRAS, 396, 1640
Flores Velázquez J. A. et al., 2021, MNRAS, 501, 4812
Grudić M. Y., Hopkins P. F., 2020, MNRAS, 495, 4306
Gürkan M. A., Freitag M., Rasio F. A., 2004, ApJ, 604, 632
Hashimoto Y., Funato Y., Makino J., 2003, ApJ, 582, 196
Hernquist L., 1990, ApJ, 356, 359
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F., Hernquist L., Martini P., Cox T. J., Robertson B., Di Matteo T., Springel V., 2005, ApJ, 625, L71
Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581
Hopkins P. F. et al., 2018, MNRAS, 480, 800
Just A., Khan F. M., Berczik P., Ernst A., Spurzem R., 2011, MNRAS, 411, 653
Karl S. J., Aarseth S. J., Naab T., Haehnelt M. G., Spurzem R., 2015, MNRAS, 452, 2337
Łokas E. L., 2019, A&A, 629, A52
Ma X. et al., 2018a, MNRAS, 477, 219
Ma X. et al., 2018b, MNRAS, 478, 1694
Ma X. et al., 2019, MNRAS, 487, 1844
Ma L., Hopkins P. F., Ma X., Anglés-Alcázar D., Faucher-Giguère C.-A., Kelley L. Z., 2021, MNRAS, 508, 1973
Mannerkoski M., Johansson P. H., Rantala A., Naab T., Liao S., 2021, ApJ, 912, L20
Mannerkoski M., Johansson P. H., Rantala A., Naab T., Liao S., Rawlings A., 2022, ApJ, 929, 167
Meng X., Gnedin O. Y., 2020, MNRAS, 494, 1263
Merritt D., 1996, AJ, 111, 2462
Morton B., Khochfar S., Oñorbe J., 2021, preprint (arXiv:2103.15848)
Mukherjee D., Zhu Q., Trac H., Rodriguez C. L., 2021, ApJ, 916, 9
Ostriker E. C., 1999, ApJ, 513, 252
Pfister H., Volonteri M., Dubois Y., Dotti M., Colpi M., 2019, MNRAS, 486, 101
Portegies Zwart S. F., McMillan S. L. W., 2002, ApJ, 576, 899
Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, MNRAS, 338, 14
Rantala A., Pihajoki P., Johansson P. H., Naab T., Lahén N., Sawala T., 2017, ApJ, 840, 53
Rodionov S. A., Sotnikova N. Y., 2005, Astron. Rep., 49, 470
Romeo A. B., 1998, A&A, 335, 922
Shi Y., Grudić M. Y., Hopkins P. F., 2021, MNRAS, 505, 2753
Springel V., 2005, MNRAS, 364, 1105
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Tremmel M., Governato F., Volonteri M., Quinn T. R., 2015, MNRAS, 451, 1868
van den Bosch F. C., Lewis G. F., Lake G., Stadel J., 1999, ApJ, 515, 50
Weisz D. R., Dolphin A. E., Skillman E. D., Holtzman J., Gilbert K. M., Dalcanton J. J., Williams B. F., 2014, ApJ, 789, 147
Weller E. J., Pacucci F., Hernquist L., Bose S., 2022, MNRAS, 511, 2229
Wellons S. et al., 2022, preprint (arXiv:2203.06201)
Yurin D., Springel V., 2014, MNRAS, 444, 62

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