Accelerating the Next Frontier of Gravitational Waves: Advanced Computational Methods & Simulations

Diptajyoti Mukherjee

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
at
Carnegie Mellon University
Department of Physics
Pittsburgh, Pennsylvania

Thesis Committee:
Prof. Hy Trac (chair)
Prof. Tiziana Di Matteo
Prof. Katelyn Breivik
Prof. Carl L. Rodriguez

April 10, 2025

Abstract

We are entering an exciting era in gravitational wave (GW) astrophysics, with upcoming space-based detectors like LISA set to explore the millihertz frequency band enabling the detection of low-frequency sources such as massive black hole (MBH) binaries and intemediate/extreme mass-ratio inspirals (I/EMRIs), offering unprecedented insights into the astrophysics of galaxy evolution and fundamental physics. Realizing the full potential of these observatories requires significant advances in computational modeling. This thesis develops and applies advanced numerical techniques, particularly fast multipole methods (FMM) in N-body simulations, to study the merger dynamics of massive black hole (MBH) binaries in nuclear star clusters (NSCs). I investigate how collisional relaxation and mass segregation within NSCs influence binary hardening timescales, and how NSCs affect the orbital evolution of MBH seeds in post-merger galaxies, where dynamical friction may be inefficient, leading to long delays in binary formation and merger. I also explore whether GWs can probe dark matter distributions by analyzing dephasing effects caused by dark matter spikes around intermediate-mass black holes. Finally, I present proof-of-concept improvements to Hamiltonian splitting integrators, assessing the performance of a fully GPU-resident N-body implementation as a step toward future high-performance FMM-based GW source modeling frameworks.

Acknowledgments

It was about a decade ago that I decided that I wanted to pursue a PhD in astronomy and astrophysics. The journey from college to completing my PhD has been everything I dreamt of and more, truly the "greatest sidequest" of my life so far. The most meaningful aspect of this decade-long journey has been the people I've met along the way. They have enriched my life immeasurably, and I cannot imagine how different my path would be without their presence. Pursuing astronomy has been the best decision I've ever made precisely because of these remarkable individuals I've encountered over the last ten years.

I begin by thanking my advisor, Hy Trac. Hy has been a pillar of support throughout these six years, and I deeply appreciate the freedom he granted me to pursue my own research interests. His mentorship guided me to where I stand today. Hy introduced me to Qirong Zhu, who brought me into my first scientific project at CMU. Without Qirong's guidance, Taichi, a significant focus of my thesis, would never have been developed. I am grateful to Qirong for teaching me about numerical methods and computational simulations. I also extend my thanks to the other members of my committee—Tiziana Di Matteo, Katie Breivik, and Carl Rodriguez for their invaluable advice and mentorship throughout these years.

I want to acknowledge the profound impact of my undergraduate advisor at Allegheny College, Jamie Lombardi. Jamie formally introduced me to astronomy when I began working on SPH simulations of stellar mergers. He has had one of the greatest influences on my development as a scientist. Jamie taught me the art of blending computer science with astrophysics and introduced me to GPU computing, something I fell in love with that eventually helped me discover my next career path.

Allegheny College also brought some of my closest friends into my life. Through the astronomy club, I met Nicholas Lewis, Nicholas Reynolds, Kyle O'Sullivan, Keith Irvin, and Hannah McAvoy, who have become my best friends. We have remained steadfast in supporting each other through life's challenges. During particularly difficult times throughout graduate school and the job hunting process, these five individuals believed in me even when I didn't believe in myself.

I want to thank two of my closest friends in Pittsburgh, whom I also met at Allegheny College - Travis Court and Liam Dugan. Our friendship has spanned a decade now, and we lived together for several years. Travis has been a source of support and an attentive listener whenever I needed someone to talk to. Co-hosting Astronomy on Tap with Travis has been one of my favorite activities over the past six years. Without his commitment and effort, Astronomy on Tap, Pittsburgh wouldn't be where it is today. During the three-year period of COVID when I couldn't return home, Liam and his family went above and beyond to ensure I never felt alone. They treated me like their own son—a kindness I will forever cherish.

I extend my gratitude to my closest friend in the astronomy community, Ugo Niccolo Di Carlo. I am thankful for his friendship and mentorship, especially regarding

star cluster simulations and N-body methods. I have created countless memories with Ugo, both in and out of the office, experiences I will forever treasure.

My favorite aspect of graduate school at CMU was undoubtedly the people I met throughout the process. They enriched my life in numerous ways—from collaborating on graduate coursework to holding "yap" sessions in the office and enjoying Pittsburgh's nightlife together. I want to thank my friends from both CMU and the University of Pittsburgh—Logan, Wes, Omar, Murman, Eka, Hannah, Sena, Joy, Chris, Vahid, Nate, Daniel, Brendan, Rapha, Ryan, and Nick—for making my time in Pittsburgh truly special. Those fun-filled nights, whether spent tackling E&M problem sets or exploring Walnut Street and Lawrenceville, will forever remain in my memories. I also acknowledge the support from other members of my research group, Yizhou He, Nianyi Chen, and Matt Ho, who welcomed me and provided guidance that helped shape me into the researcher I am today.

I reserve special appreciation for my colleague, officemate, and girlfriend, Sofia Splawska. She enriches my life not only through our discussions about cosmology and physics but also by continuously introducing me to new art, music, and travel ideas, adding color and spontaneity to my world. Her unwavering emotional support helped me navigate the challenges of job hunting, defending my thesis, and relocating. She is one of the kindest and most supportive people I know, and I am truly lucky to have met her at CMU.

Finally, I acknowledge the three most important people in my life. My brother, Dibyo, the finest computer programmer I know and my lifelong inspiration. He gave me the dream of studying in America and has supported me unfailingly over the last decade. I learned coding by imitating his work in BASIC. Whenever I face a conundrum, I seek his advice, and there has never been a moment when he hasn't supported me. In times of uncertainty, I ask myself, "What would Dibyo do?" Lastly, I acknowledge my parents, without whose love, support, and sacrifice none of this would have been possible. They have prioritized my brother's and my happiness and goals above everything else, supporting us through every challenge. They taught me intellectual curiosity, the value of hard work, and the importance of remaining humble and grounded. They provided me with the greatest gift possible—a good education—and all the privileges that accompany it. It is to them that I dedicate this thesis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

We stand on the brink of the most significant revolution in astronomy since the invention of the telescope. While humanity has long possessed the ability to see the cosmos, gravitational waves (GWs) now offer us the unprecedented ability to listen to the universe and uncover its deepest mysteries. These waves hold the potential to answer profound questions about the universe's expansion, the elusive nature of dark matter (DM), and the formation and growth of supermassive black holes (SMBHs). The advent of the James Webb Space Telescope (JWST) has opened an observational window into the high-redshift universe, revealing early SMBH populations whose existence challenges our theoretical understanding of black hole formation and growth. Disentangling this mystery requires the next frontier of GW astronomy, where space-based detectors like the Laser Interferometer Space Antenna (LISA) will play a pivotal role in probing the dynamical environments of merging massive black hole binaries (MBHBs). However, maximizing the scientific return of LISA and future detectors demands significant improvements in theoretical modeling, particularly in understanding MBHB coalescence within their complex astrophysical environments.

This thesis explores two key aspects of this problem: the role of nuclear star clusters (NSCs) in shaping MBH binary merger timescales and the potential imprint of DM spikes on GW signals. By leveraging fast N-body methods such as the fast multipole method, this work enables high-resolution simulations that capture the intricate gravitational interactions governing MBHB evolution. These computational advances, coupled with zoom-in cosmological simulations, provide a powerful framework for constraining merger rates and enhancing the science output of next-generation GW observatories. Ultimately, this research seeks to bridge the gap between theory and observation, using GWs as a tool to unravel the nature of DM and the pathways that drive the growth of SMBHs.

1.1 Collisional N-body simulations

The collisional N-body problem, which models the gravitational interaction of N particles over time, is one of the most computationally demanding challenges in modern physics. Unlike collisionless N-body simulations—where close encounters between particles are neglected and the system is treated in a mean-field approximation—collisional simulations must explicitly account for two-body interactions. These interactions drive the long-term evolution of dense stellar systems, including open clusters, globular clusters, and nuclear star clusters (e.g., Spitzer, 1987). Collisional relaxation plays a pivotal role in numerous astrophysical phenomena, such gravo-thermal instability leading to core-collapse, and black hole mergers (Heggie and Hut, 2003). Given that this thesis focuses on nuclear star clusters, which are inherently collisional systems, a significant portion of this introduction is dedicated to outlining the fundamental aspects of collisional N-body simulations.

Collisional N-body methods can be categorized based on how the gravitational force acting on each particle is computed (e.g., Aarseth, 2003; Heggie and Hut, 2003; Binney and Tremaine, 2011; Dehnen and Read, 2011). Direct N-body methods evaluate the total force on each particle by summing contributions from all other particles in the system. The acceleration $\mathbf{a}_i$ of particle i is given by

$$\mathbf{a}_i = G \sum_{\substack{j=1 \\ j \neq i}}^{N} m_j \frac{\mathbf{r}_j - \mathbf{r}_i}{|\mathbf{r}_j - \mathbf{r}_i|^3}, \tag{1.1}$$

where G is the gravitational constant, m_j is the mass of particle j, and $\mathbf{r}_i$ and $\mathbf{r}_j$ denote the positions of particles i and j, respectively. Although direct methods provide high accuracy, their $\mathcal{O}(N^2)$ computational complexity poses a significant bottleneck for simulations with $N \sim 10^6$—a regime necessary for modeling nuclear star clusters and galactic nuclei. Running such large-scale simulations often requires months on state-of-the-art computing hardware. Consequently, substantial effort has been dedicated to developing more efficient approximate methods for gravitational force calculations.

1.1.1 Fast force calculation algorithms

To overcome the steep cost of direct summation, one widely used approach is to treat long-range and short-range interactions differently, updating only the more rapidly varying short-range forces at every time step. Such techniques have been used in direct-summation based N-body codes such as `NBODY6` (Aarseth, 1999) using the Ahmad-Cohen (Ahmad and Cohen, 1973) scheme.

Beyond partial treatments of direct summation, approximate force solvers can dramatically cut computational expense. One prominent example is the Barnes-Hut (BH) tree (Barnes and Hut, 1986), which employs a hierarchical decomposition of the simulation volume and uses multipole expansions to approximate forces. The

algorithmic complexity of BH tree codes is typically $\mathcal{O}(N \log N)$, representing a substantial improvement over direct summation. It is challenging to utilize tree-codes for collisional simulations primarily for two reasons. The lower force accuracy in tree-codes can lead to artificial effects leading to non-physical evolution. Additionally, the integration with an individual timestepping scheme creates a bottleneck where the tree has to be updated at every timestep. Nonetheless, particle-particle particle-tree (P^3T) codes that combine a partial direct force calculation (for nearby particles) with a tree-based approximation (for distant particles) have recently emerged, demonstrating performance that can rival fully direct methods (Wang et al., 2020).

The fast multipole method (FMM)(Greengard and Rokhlin, 1987; Cheng et al., 1999) offers another promising route for accelerating collisional N-body calculations. Building on earlier applications of FMM to collisionless systems (e.g., Dehnen, 2000, 2002), Dehnen (2014) presented an optimized version specifically designed to handle high-accuracy requirements, highlighting its $\mathcal{O}(N)$ or even sublinear scaling in special cases. These efficiency gains, along with well-controlled error properties, make FMM an attractive alternative to standard tree codes. However, fully adapting FMM for collisional dynamics, which demands high accuracy and robust handling of individual stellar interactions, remains an active area of research. As larger N-body problems move beyond the reach of direct summation, advancements in methods like FMM are crucial for simulating dense stellar systems such as nuclear star clusters.

1.1.2 Hierarchical Hamiltonian splitting integrators

While sophisticated force-calculation algorithms, as discussed in the previous section, are crucial for efficient N-body simulations, the other key component is the choice of time integration scheme. For large N, adaptive, individual timesteps are typically introduced to maximize computational efficiency, ensuring that particles moving on very short dynamical timescales are updated more frequently than those evolving slowly.

Broadly, N-body integrators can be divided into symplectic and non-symplectic schemes. Symplectic integrators preserve the volume of position-momentum phase space, which translates into superior long-term energy conservation. Unfortunately, when individual timesteps are adopted, this strict symplecticity is typically lost. For instance, although the popular fourth-order Hermite (e.g., Aarseth, 2003) scheme is widely used in many N-body codes, its incorporation of adaptive timesteps breaks strict symplecticity, resulting in large long term energy errors. In particular, one tries to numerically integrate the trajectory $(\mathbf{r}(t), \mathbf{v}(t))$ of a particle using a timestep function $\tau(t)$. After integration, the particle reaches a new state $(\mathbf{r}(t+\tau), \mathbf{v}(t+\tau))$. In this state, the particle has a new timestep τ'. Under ideal circumstances, to preserve the time-reversible properties of the Hamiltonian, integrating the particle backwards in time using τ' *should* produce the original state of the particle. However, due to numerical errors, $\tau' \neq \tau$ in general and the time reversible properties of the

Hamiltonian are lost, leading to long term energy drift.

One approach to alleviate this loss of symplecticity is to instead preserve the time-symmetric properties of the Hamiltonian (Hut et al., 1995; Dehnen, 2017). This requires ensuring that the forward and backward timesteps remain identical at each integration step. Using the previous notation, this condition is expressed as:

$$\tau' = \tau \tag{1.2}$$

$$\Rightarrow \tau(\tau(t) + t) = \tau(t). \tag{1.3}$$

Following Pelupessy et al. (2012a), we note that $\tau(\tau(t) + t)$ represents a timestep in the negative direction, whereas $\tau(t)$ represents a timestep in the positive direction. Using the notation introduced in Pelupessy et al. (2012a), we rewrite the equation as:

$$\tau^{-}(\tau(t) + t) = \tau^{+}(t). \tag{1.4}$$

A symmetrized timestep function satisfying the above condition can be introduced as follows:

$$\tau^{-}_{\rm sym} = \tau(t)/2 + \tau(t - \tau^{-}_{\rm sym})/2 \tag{1.5}$$

$$\tau^{+}_{\rm sym} = \tau(t)/2 + \tau(t + \tau^{+}_{\rm sym})/2 \tag{1.6}$$

. Taylor expanding the second equation to first order in $\tau^{+}_{\rm sym}$ and rearranging gives us the timestep function

$$\tau_{\rm sym} = \frac{\tau(t)}{1 - \frac{1}{2}\frac{d\tau}{dt}} \tag{1.7}$$

We split the Hamiltonian into a *fast* subsystem, consisting of particles that timesteps shorter than a threshold timestep, and a *slow* subsystem, containing the remaining particles. By adopting symmetrized timesteps using the formulation derived above based on freefall and flyby timescales, the algorithm achieves long term integration accuracy. An example of a second-order decomposition built on this principle is the `HOLD` integrator (Pelupessy et al., 2012a), in which both the fast subsystem and its interaction with the slow subsystem are integrated via a leapfrog scheme and is presented in Listing 1.1. In the HOLD decomposition the kicks between the slow and fast subsystems are symmetric leading to an exact conservation of total linear momentum.

```
func HOLD(system, timestep, calculate_timestep):

    if(calculate_timestep)
        get_timesteps(system)

    slow, fast = split_system(system)
```

```
if(fast.N = 0)
    simtime = simtime + timestep

if(fast.N > 0)
    HOLD(fast, timestep / 2, false)

if(slow.N > 0)
    drift(slow, timestep / 2)

if(slow.N > 0)
    kick_slow(slow, timestep)

    if(fast.N > 0)
        kick_slowfast(slow, fast, timestep)

if(slow.N > 0)
    drift(slow, timestep / 2)

if(fast.N > 0)
    HOLD(fast, timestep / 2, true)
```

Listing 1.1: Psedocode for the second order `HOLD` decomposition scheme using a drift-kick-drift leapfrog integrator. This leads to exact conservation of momentum.

More recently, fourth-order decomposition schemes such as `HHS-FSI` (e.g., Chin and Chen, 2005; Rantala et al., 2021a) have gained popularity due to their use of strictly positive timesteps. This contrasts with Yoshida-type schemes (Yoshida, 1990), which require negative timesteps for fourth- or higher-order symplectic integrators. A major challenge in higher-order schemes is the accurate evaluation of the force gradient, which is necessary for higher-order corrections. The `Taichi` N-body code (introduced in Chapters 2 and 3) implements both the second-order `HOLD` and the fourth-order `HHS-FSI` integrators. In particular, our work employs `HHS-FSI` with an additional modification inspired by Omelyan (2006) and Farr and Bertschinger (2007), where the force gradient is computed using a position-extrapolation method. This approach preserves the benefits of fourth-order accuracy while simplifying computations, eliminating the need for additional steps to evaluate the force gradient.

1.1.3 Few-body regularization

Even with higher-order integration schemes, accurately capturing the long-term dynamics of very tight or highly eccentric binaries can remain challenging (e.g., Wang et al., 2020). Such hard binaries, which play a central role in this thesis, require

specialized treatment to avoid significant energy errors over many dynamical times. A fundamental problem in collisional N-body simulations is that the timestep tends to zero as the separation between two strongly interacting particles approaches zero, often leading to large energy errors and numerical instability (e.g., Aarseth, 2003, and references within).

One strategy to address this involves introducing a fictitious time parameter s through a coordinate transformation. In the well-known Kustaanheimo-Stiefel (KS) regularization (Kustaanheimo et al., 1965), a Levi-Civita transformation is applied to the system's coordinates, removing the singularity and converting the system into a four dimensional harmonic oscillator. KS regularization can be complicated to implement and computationally expensive when generalized beyond a few particles.

An alternative approach, pioneered by Mikkola and Tanikawa (1999a) and Preto and Tremaine (1999), is to use a time transformation function inversely proportional to the gravitational potential. Combined with a leapfrog integrator in drift-kick-drift (DKD) form, this method can achieve exact integration of a two-body Kepler orbit. Consider the Hamiltonian H of a system:

$$H = T(\mathbf{p}) - U(\mathbf{r}, t), \tag{1.8}$$

where $T(\mathbf{p})$ is the kinetic energy and $U(\mathbf{r}, t)$ is the potential energy. Assuming $U > 0$, we apply the time-transformation

$$ds = U\, dt, \tag{1.9}$$

where s is the new regularized time for the system. The time-transformed Hamiltonian Γ can be written as

$$\Gamma = \frac{T - U + B}{U}, \tag{1.10}$$

with $B = -E_0$, the negative of the system's initial total energy. Here, B is the conjugate momentum of time, which now becomes a coordinate itself. While this Hamiltonian is non-separable, one notes that $\Gamma = 0$ in real systems (due to energy conservation). Mikkola and Tanikawa (1999a) instead propose the following "log" Hamiltonian:

$$\tilde{\Gamma} = \log\big(1 + \Gamma\big). \tag{1.11}$$

This version is separable, and the time evolution of the coordinates is:

$$\mathbf{p}' = -\frac{\partial \tilde{\Gamma}}{\partial \mathbf{r}} = -\frac{1}{1+\Gamma}\frac{\partial \Gamma}{\partial \mathbf{r}}, \tag{1.12}$$

$$B' = -\frac{\partial \tilde{\Gamma}}{\partial t} = -\frac{1}{1+\Gamma}\frac{\partial \Gamma}{\partial t}, \tag{1.13}$$

$$\mathbf{r}' = \frac{\partial \tilde{\Gamma}}{\partial \mathbf{p}} = \frac{1}{1+\Gamma}\frac{\partial \Gamma}{\partial \mathbf{p}}, \tag{1.14}$$

$$t' = \frac{\partial \tilde{\Gamma}}{\partial B} = \frac{1}{1+\Gamma}\frac{\partial \Gamma}{\partial B}. \tag{1.15}$$

For an N-body system, using a DKD integrator, the DKD steps are modified as:

$$\mathbf{r}_{1/2} = \mathbf{r}_0 + \frac{(dt/2)\big(\mathbf{p}_0/m\big)}{T_0 + B}, \tag{1.16}$$

$$t_{1/2} = t_0 + \frac{dt/2}{T_0 + B}, \tag{1.17}$$

$$\mathbf{p}_1 = \mathbf{p}_0 + \frac{dt}{U(\mathbf{r}_{1/2})} \frac{\partial U(\mathbf{r}_{1/2})}{\partial \mathbf{r}_{1/2}}, \tag{1.18}$$

$$\mathbf{r}_1 = \mathbf{r}_{1/2} + \frac{(dt/2)\big(\mathbf{p}_1/m_k\big)}{T_1 + B}, \tag{1.19}$$

$$t_1 = t_{1/2} + \frac{dt/2}{T_1 + B}, \tag{1.20}$$

where values with the subscript $1/2$ indicate the half step, $\mathbf{r}$ & $\mathbf{p}$ indicate the positions and momenta, and t represents the time. In the above equations, we utilize the fact that U does not depend on time in the general Newtonian N-body equation.

However, for systems with $N > 2$, additional techniques are required to maintain high accuracy. By combining a Bulirsch-Stoer (BS) integrator (e.g., Gragg, 1965; Press, 1986) with chain coordinate transformations (Mikkola and Aarseth, 1990) for close encounters, one can achieve near machine-precision accuracy for multi-body systems-a technique often referred to as algorithmic regularization.

In our work, we combine algorithmic regularization with `HOLD` and `HHS-FSI` integrators through the `SpaceHub` (Wang et al., 2021a) library, integrating it directly into the `Taichi` N-body code to handle the evolution of highly eccentric massive black hole (MBH) binaries. This approach ensures stable, precise long-term integration for few-body subsystems.

1.1.4 Monte-Carlo methods and other alternatives

Because direct collisional N-body simulations scale as $\mathcal{O}(N^2)$ and quickly become prohibitive for realistic globular clusters, researchers have long exploited orbit-averaged Monte-Carlo techniques that assume an underlying spherical potential. Pioneered independently by Hénon (1971a) and by Spitzer and Hart (1971), these algorithms advance every star through a smooth cluster potential for one dynamical time, then apply a stochastic velocity "kick" drawn from local diffusion coefficients before rebuilding the global potential. All orbits can be evolved in parallel, yielding orders-of-magnitude speed-ups. In the original Hénon implementation the diffusion step is estimated from the exact two-body deflection with the nearest neighbor, a strategy adopted by widely used cluster codes such as `MOCCA` (e.g., Hypki and Giersz, 2013) and `CMC` (e.g., Joshi et al., 2000; Pattabiraman et al., 2013); by contrast, the Spitzer–Hart formulation, employed in `RAGA` (Vasiliev, 2015), derives the kicks from an orbit-averaged local diffusion coefficients and couples them to a self-consistent-field expansion, allowing the code to simulate non-spherical geometries. Complementary approaches

integrate the Fokker–Planck equation directly on an energy–angular-momentum grid with finite-difference schemes.

1.2 Fast multipole method

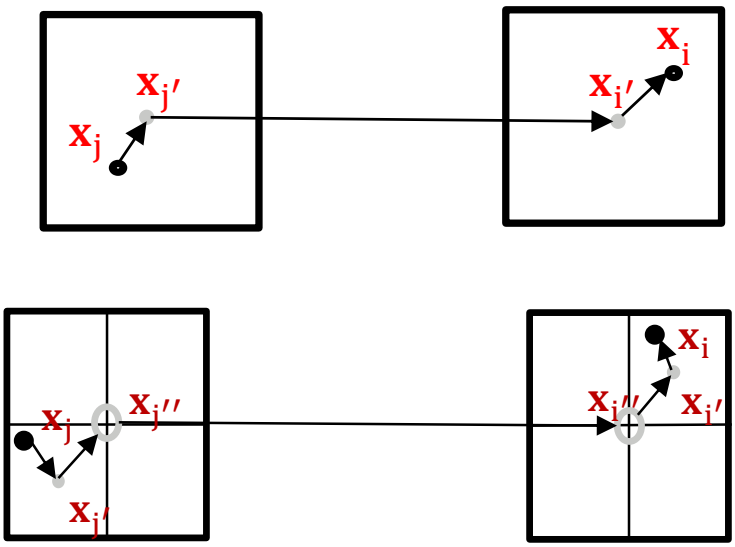

Figure 1.1: Visualization of the cell-cell interaction in the FMM algorithm. Top: interaction between two cells with no sub-cells. In this case, the position vector $\mathbf{x}_{ij}$ can be decomposed as the sum of the position vectors $\mathbf{x}_{jj'}$ (source particle-cell-center vector), $\mathbf{x}_{ii'}$ (sink particle-cell-center vector), and $\mathbf{x}_{ij}$ (source cell-center – sink cell-center vector). Bottom: in this case, the source and sink cells have to be subdivided into four sub-cells. This leads to two additional shifts $\mathbf{x}_{j''j'}$ (source sub-cell center - source cell center vector) and $\mathbf{x}_{i''i'}$ (sink sub-cell center - sink cell center vector).

As introduced in the previous section, FMM (e.g., Greengard and Rokhlin, 1987; Cheng et al., 1999; Dehnen, 2014) is an approximate force solver that relies on multipole expansions to evaluate gravitational interactions efficiently. Unlike the BH tree algorithm, which approximates distant groups of particles using a particle-to-multipole approach, FMM exploits a cell-to-cell interaction scheme based on *local expansions* of distant multipoles. This distinction allows FMM to capture long-range forces with high accuracy while remaining scalable to very large numbers of particles.

Similar to BH tree codes, FMM begins with an octtree decomposition of the simulation volume. Each node (or cell) in the tree encloses a subset of particles, and an opening angle criterion governed by an accuracy parameter determines whether a node's interaction with another node is approximated by multipole expansion or

computed via direct summation. In practice, particles within the same or neighboring cells are handled directly, ensuring high accuracy for close encounters. Meanwhile, interactions between well-separated cells are evaluated through multipole and local expansions, and a dual-tree walk efficiently identifies which cells require approximations.

The chief advantage of FMM over simpler tree algorithms is its optimal $\mathcal{O}(N)$ or near-linear scaling. For large N, such an approach dramatically reduces the computational expense compared to direct summation or even classic BH tree schemes. Demonstrating that an FMM-based code remains viable for collisional N-body simulations, where accuracy requirements are stringent and individual time steps are common, is a key goal of this thesis.

In the remainder of this chapter, we present the core components of our FMM implementation. We begin by outlining how the multipole and local expansions are defined and combined to compute the gravitational potential and force in a linear-time algorithm. Although we introduce these expansions in terms of Cartesian coordinates for clarity, our actual simulations employ solid spherical harmonics for enhanced accuracy and computational efficiency. This foundational understanding of the FMM kernels sets the stage for the more detailed discussion of our collisional N-body implementation in the following chapters.

1.2.1 FMM in Cartesian coordinates

We begin by considering the Green's function for gravitational interactions,

$$G(\mathbf{x}_{ij}) = \frac{1}{|\mathbf{x}_i - \mathbf{x}_j|}, \tag{1.21}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the positions of particles (or cells). In Figure 1.1 (top), suppose we have two well-separated cells, and we write

$$\mathbf{x}_{ij} = \mathbf{x}_{jj'} + \mathbf{x}_{i'j'} + \mathbf{x}_{ii'}. \tag{1.22}$$

Under the assumption

$$\mathbf{x}_{i'j'} \gg \big(\mathbf{x}_{jj'} + \mathbf{x}_{ii'}\big),$$

we can Taylor-expand $G(\mathbf{x}_{ij})$ around $\mathbf{x}_{i'j'}$ to obtain

$$G(\mathbf{x}_{ij}) = \sum_{n=0}^{\infty} \frac{1}{n!} \big(\mathbf{x}_{ii'} + \mathbf{x}_{jj'}\big)^n \nabla^{(n)} G(\mathbf{x}_{i'j'}). \tag{1.23}$$

Truncating this infinite sum at a finite order p and applying the binomial theorem yields

$$G(\mathbf{x}_{ij}) = \sum_{n=0}^{p} \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{(n-k)!\,k!}\, \mathbf{x}_{ii'}^{k}\, \mathbf{x}_{jj'}^{n-k}\, \nabla^{(n)} G(\mathbf{x}_{i'j'}), \tag{1.24}$$

$$G(\mathbf{x}_{ij}) = \sum_{k=0}^{p} \frac{1}{k!}\, \mathbf{x}_{ii'}^{k} \sum_{n=0}^{p-k} \nabla^{(n+k)} G(\mathbf{x}_{i'j'})\, \frac{1}{n!}\, \mathbf{x}_{jj'}^{n}. \tag{1.25}$$

Within this framework, we identify three key steps associated with FMM:

Particle-to-Multipole (P2M). For a source cell, we define the multipole moments by

$$M^{n}(\mathbf{x}_{j'}) = \sum_{j=1}^{N} \frac{1}{n!}\, \mathbf{x}_{jj'}^{n}\, q_j, \tag{1.26}$$

where $\mathbf{x}_{jj'}$ is the displacement of particle j relative to the cell center, and q_j is the mass (or charge, in a generic $1/r$ interaction).

Multipole-to-Local (M2L). The local expansion at the sink cell is given by

$$L^{k}(\mathbf{x}_{i'}) = \sum_{n=0}^{p-k} \nabla^{(n+k)} G(\mathbf{x}_{i'j'})\, M^{n}(\mathbf{x}_{j'}), \tag{1.27}$$

where $\mathbf{x}_{i'j'}$ is the distance between the source and sink cell centers.

Local-to-Particle (L2P). Finally, each particle within the sink cell evaluates its potential (or field) via

$$u_i = \sum_{k=0}^{p} \frac{1}{k!}\, \mathbf{x}_{ii'}^{k}\, L^{k}(\mathbf{x}_{i'}). \tag{1.28}$$

In a hierarchical decomposition scenario, as demonstrated by the bottom figure in Figure 1.1, we follow two further steps. First, we note the relations

$$\mathbf{x}_{j''j} = \mathbf{x}_{j''j} + \mathbf{x}_{j'j}, \quad \mathbf{x}_{ii''} = \mathbf{x}_{ii'} + \mathbf{x}_{i'i''}, \tag{1.29}$$

which indicate that two shifts are required instead of one. This naturally leads to two additional FMM kernels:

Multipole-to-Multipole (M2M).

$$M^n(\mathbf{x}_{j''}) = \sum_{j'=1}^{4} \sum_{k=0}^{n} \frac{1}{(n-k)!} \mathbf{x}_{j''j'}^{n-k} M^k(\mathbf{x}_{j'}), \tag{1.30}$$

where contributions from child cells at $\mathbf{x}_{j'}$ are aggregated into the parent cell at $\mathbf{x}_{j''}$.

Local-to-Local (L2L).

$$L^n(\mathbf{x}_{i'}) = \sum_{i'=1}^{4} \sum_{k=n}^{p} \frac{1}{(k-n)!} \mathbf{x}_{i'i''}^{k-n} L^k(\mathbf{x}_{i''}), \tag{1.31}$$

which transfers local expansions from parent cells to their children.

1.2.2 Dual-tree walk

The FMM algorithm typically works in tandem with a dual-tree walk algorithm (Dehnen, 2000, 2002) to achieve its characteristic $\mathcal{O}(N)$ scaling. Unlike the single-tree traversal in the BH approach, the dual-tree walk recursively descends into both the *source* and *sink* octtrees. As illustrated in Listing 1.2, if the cells in question satisfy a well-separation criterion, the algorithm performs cell-cell interactions for all pairs of child nodes in the two cells. Otherwise, it subdivides one or both cells further, continuing to recurse until the required separation criterion is met.

Within this framework, the particles are hierarchically arranged in an oct-tree structure. Each leaf cell first gathers its multipole expansion via the P2M kernel. These multipoles are then aggregated up the tree, from child nodes to their parent nodes, using the M2M kernel. At the root, the algorithm has a complete multipole description of all cells for computing far-field gravitational interactions. For well-separated cell pairs, the approximate gravitational force is evaluated with a M2L kernel. The resulting local expansions then propagate down the tree to the leaf cells via the L2L kernel. Finally, each particle's potential and force are determined from the local expansion in its leaf cell by the L2P kernel. For particles not satisfying the well-separation criteria, direct force summation is used instead.

```
func Interact(A,B):
    try to perform the interaction between nodes A and B

    if(it cannot be performed)
        if (A = B)
            for(all pairs {a, b} of child nodes of A)
                Interact(a, b);
        else if (rmax(A) > rmax(B))
            for(all child nodes a of A)
```

```
            Interact(a, B);
    else
        for(all child nodes b of B)
            Interact(A, b);
```

Listing 1.2: Psuedocode (adopted from Dehnen 2002) that demonstrates the dual tree walk algorithm for a set of two cells A and B

1.3 Massive black hole binary evolution

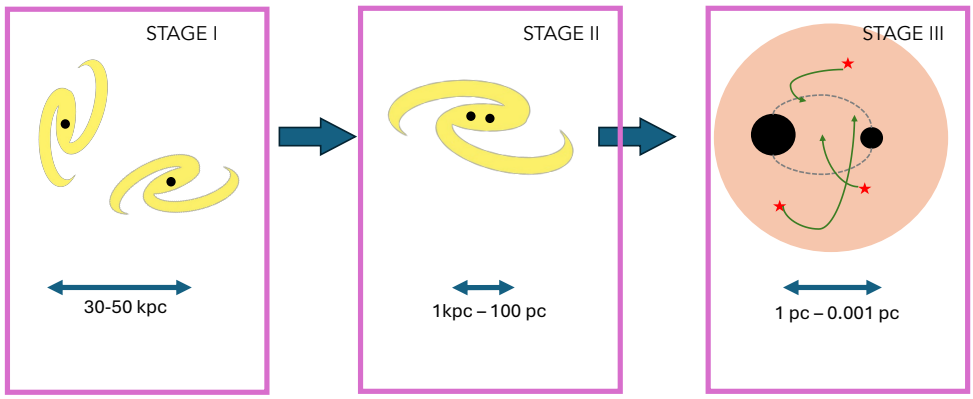

Figure 1.2: A visualization of the three-step process for MBHB coalescence described in Section 1.3. Tidal interactions between the galaxies and dynamical friction from the dark matter halo is key to sinking the MBHs in the first stage. In the second stage, after the galaxies have merged, dynamical friction from gas, stars, and dark matter plays a major role. In the third stage, after the MBHs have reached separations of $\lesssim 0.1$ pc, three-body interactions with nearby stars play a more dominant role in driving the merger process.

MBH binaries are among the most promising sources of millihertz GWs that will be detectable by LISA. MBH binaries with masses in the range $10^4 M_\odot - 10^7 M_\odot$ are primary targets for LISA (Amaro-Seoane et al., 2017) and TianQin (Luo et al., 2016), which are expected to detect MBH mergers out to redshifts $z > 20$. Depending on different seeding mechanisms, accretion models, and subgrid dynamical prescriptions, large-volume cosmological simulations predict merger rates of $0.1 - 2.0\,\mathrm{yr}^{-1}$ at $z = 2$ (see Figure 20 from Di Matteo et al., 2023, and references therein) . Observing GWs from the final inspiral phase of these mergers not only provides insight into the assembly history of galaxies but also constrains key dynamical properties of galactic nuclei surrounding MBH binaries. Consequently, modeling the dynamical evolution of MBH binaries within galactic nuclei is crucial for understanding both the galactic environment and its influence on black hole growth and merger rates.

Due to the wide range of conditions in which MBH binaries may form, their evolutionary paths remain a topic of active debate. Broadly, the evolution of an MBH binary can be split into three distinct stages (Begelman et al., 1980) before the final, GW-driven coalescence:

1. **Sinking Phase:** When MBHs are separated by more than ~ 100 pc, they experience dynamical friction as they move through a medium composed of stars, dark matter, and gas. This motion induces an overdensity in their wake, which exerts a gravitational drag force (Chandrasekhar, 1943), causing the MBHs to lose energy and angular momentum, gradually sinking toward the center of the merger remnant. While accurate estimation of the dynamical friction timescale $T_{\rm df}$ requires N-body simulations, following Binney and Tremaine (2011), we obtain:

$$T_{\rm df} \propto \frac{M}{M_{\rm BH}} \frac{1}{\ln \Lambda}, \tag{1.32}$$

   where M is the total mass of the merged galaxy, $M_{\rm BH}$ is the MBH mass, and $\ln \Lambda$ is the Coulomb logarithm, typically ~ 10. Smaller MBHs experience longer dynamical friction timescales and may "wander" within the merged galaxy without efficiently sinking to the center (e.g., Ma et al., 2021).

2. **Bound Binary Formation:** Once the MBHs reach separations below ~ 100 pc, they can form a gravitationally bound binary by ejecting mass from the central region of the merger remnant. Using equation (8.78a) from Merritt (2013), the mass that must be ejected, $M_{\rm ej}$, is given by:

$$M_{\rm ej} \approx 2 M_{12} \sigma^2, \tag{1.33}$$

   where M_{12} is the total mass of the binary, and σ is the velocity dispersion of the galaxy. Thus, an amount of mass comparable to the binary's total mass must be removed from the system for the MBHs to become bound. This mass ejection occurs primarily through a combination of dynamical friction and three-body scattering. Initially, dynamical friction dominates, but as the MBH separation decreases, three-body scattering becomes increasingly significant.

3. **Hard Binary Stage:** When the binary reaches separations below ~ 0.1 pc, it enters the hard binary phase, where further orbital decay is driven primarily by three-body interactions. Incoming stars that pass close to the binary can undergo strong encounters, either being ejected from the vicinity or from the galaxy entirely. These ejections extract energy from the binary, leading to further orbital decay. Under the assumption of an abundant supply of such stars, the change in the inverse semi-major axis $1/a$ of the binary follows:

$$\frac{d}{dt}\left(\frac{1}{a}\right) = \frac{GH\rho}{\sigma}, \tag{1.34}$$

where H is a dimensionless number called the hardening rate and ρ is the stellar density (Quinlan, 1996; Sesana et al., 2006). Typically $H \approx 1-10$ Thus, the binary loses energy at a constant rate. If the supply of scattering stars is insufficient, the binary may stall at parsec-scale separations, leading to the so-called "final parsec problem"—a barrier to further evolution via GW emission (e.g., Milosavljević and Merritt, 2003).

Within the hard binary stage, the strong interaction between the MBH binary and stars in its loss cone heavily influences the binary's orbital decay. This loss cone contains low-angular-momentum stars that can approach the binary closely enough to be ejected in slingshot events, transferring energy and angular momentum to the MBHs. In a spherical system, the loss-cone angular momentum L_{lc} is given by

$$L_{lc} \approx \sqrt{2\,G\,M_{12}\,K\,a}, \tag{1.35}$$

where M_{12} is the total mass of the MBH binary, $K \sim 1$–2 is a dimensionless constant, and a is the binary's semi-major axis. The timescale for binary hardening can be extremely sensitive to the supply of these stars, and in spherical, gas-poor galaxies, the lack of efficient loss-cone refilling can lead to orbital decay times exceeding the Hubble time. In spherical galaxies loss cone filled by collisional processes like two-body relaxation leading to a long refilling timescale in systems with large number of stars, like galactic nuclei. However, a wealth of numerical and analytical studies over the past two decades indicate that non-sphericity (e.g., triaxial or clumpy structures) can introduce torques that replenish the loss cone on much shorter timescales, thereby circumventing the final parsec problem (e.g., Berczik et al., 2006; Khan et al., 2013; Vasiliev et al., 2015; Vasiliev, 2017).

Another critical factor in binary evolution is the numerical resolution of simulations (e.g., Sesana et al., 2011; Nasim et al., 2020; Khan and Holley-Bockelmann, 2021; Rawlings et al., 2023). If the number of particles is too low, collisional relaxation effects become artificially enhanced, leading to unphysically short GW merger timescales (e.g., Merritt et al., 2007a). Faithfully resolving the relevant dynamical processes requires $N \gtrsim 10^6$. While low particle numbers impact both the evolution of the semi-major axis and eccentricity, the latter is more strongly affected, as it is a second-order effect in angular momentum. Low-resolution simulations can artificially produce more eccentric binaries, resulting in unrealistically short GW timescales. Examining Figure 1.3, where we calculate the GW timescale of an eccentric binary compared to a circular binary with the same mass ratio using the Peters (1964b) formula, we find that at high eccentricities ($e = 0.9$), GW effects are significantly enhanced, reducing the GW timescale by up to a factor of 100 compared to $e = 0$. Detailed N-body simulations suggest that a mass ratio ≥ 50 (see Chapter 8 of Merritt, 2013a) between the incoming stellar particles and the smaller MBH is required for statistically robust eccentricity evolution; however, stochastic effects make it difficult to precisely constrain extreme eccentricities ($e \gtrsim 0.99$) (Nasim et al., 2020; Rawlings et al., 2023).

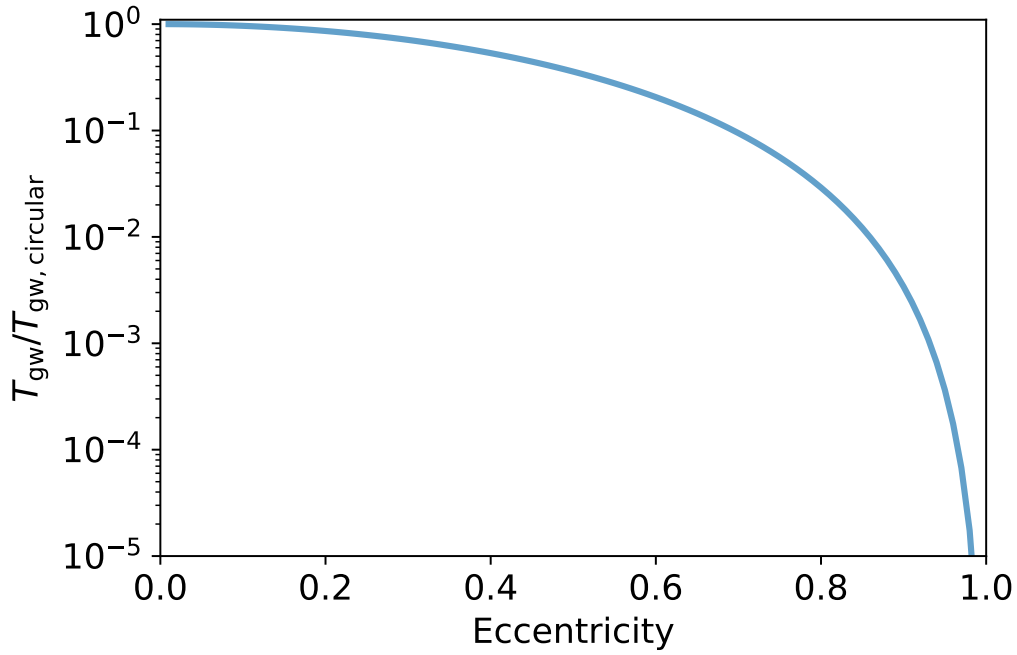

Figure 1.3: The ratio of GW timescale of an eccentric binary to that of a circular binary with the same mass ratio. We use Peters (1964a) equations to calculate the GW merger timescale. We notice that the timescale is very sensitive to the eccentricity. At $e = 0.9$, the GW timescale is $100\times$ shorter than at $e = 0$.

In large-volume cosmological and zoom-in simulations, subgrid models for dynamical friction (e.g., Tremmel et al., 2015a; Chen et al., 2022a; Ma et al., 2023), three-body interactions, and accretion introduce additional uncertainties into merger rate predictions. Consequently, high-resolution N-body simulations remain indispensable for constraining MBH coalescence timescales and accurately modeling the structure of galactic nuclei at parsec and sub-parsec scales. These simulations provide critical insights into the stellar and dark matter distributions around MBH binaries—details that remain beyond the resolution limits of large-scale cosmological simulations. A wealth of literature is available on methods to resolve such substructures in galaxies derived from large cosmological volumes. The `GRIFFIN` (e.g., Partmann et al., 2025) project uses the `KETJU` code to perform zoom-in simulations to study the co-evolution of the MBH with the surrounding stellar cluster. Additionally, multi-scale techniques using a combination of N-body or Monte-carlo simulations in conjunction with cosmological simulations such as MOSAICS (e.g., Pfeffer et al., 2018) and FIRE (e.g., Rodriguez et al., 2023) to study the evolution of globular clusters in galaxies

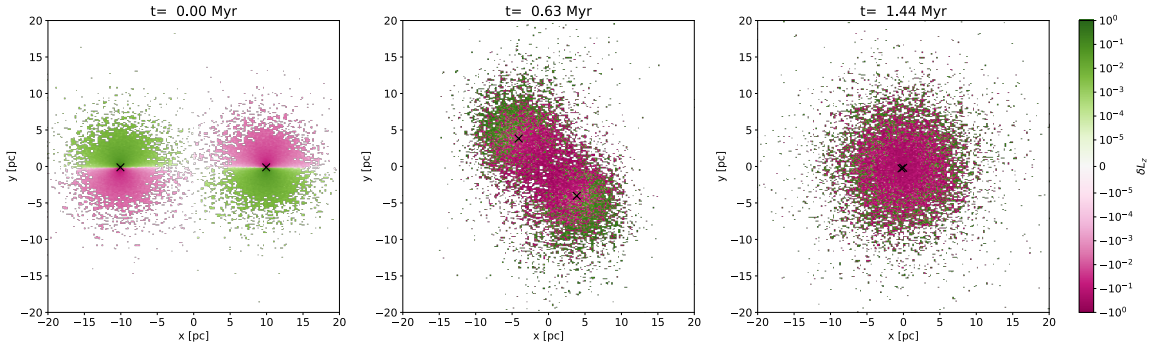

Figure 1.4: Relative change in angular momentum of stars in two NSCs during a merger process. Tidally stripped stars gaining angular momentum (green) move outward, while stars losing angular momentum (magenta) move inwards towards the merger. The stripped stars steal angular momentum from the NSC system helping the MBHs embedded in them shrink from 20 pc to ~ 0.01 pc within ~ 1 Myr. This process is substantially faster than the DF timescale leading to efficient mergers (using data from Mukherjee et al. (2023)).

1.4 Nuclear star clusters

Nuclear star clusters (NSCs) rank among the densest stellar systems in the universe, with mass densities reaching up to $\sim 10^6\,M_\odot\,\mathrm{pc}^{-3}$. Surveys indicate that NSCs are both common and diverse across galactic environments: Georgiev et al. (2016) find that up to $\sim 90\%$ of galaxies with stellar masses $M_* = 10^9\,M_\odot$ and $\sim 20\%$ of those with $M_* = 10^7\,M_\odot$ host NSCs. Current formation models suggest that these clusters arise from a combination of *in situ* star formation, where gas in the galactic center collapses to form new stars (e.g., Loose et al., 1982; Hopkins and Quataert, 2010, 2011; Guillard et al., 2016; Brown et al., 2018), and *ex situ* processes, where massive clusters (e.g., globular clusters) undergo orbital decay and migrate inward (e.g., Tremaine et al., 1975; Hartmann et al., 2011; Antonini and Perets, 2012). This dual-origin scenario is further supported by the presence of multiple stellar populations observed in many NSCs.

Owing to their high density, NSCs can host runaway collisions capable of forming large black hole (BH) seeds. Particularly in low-mass, or "dwarf", galaxies, seed BHs of up to $10^6\,M_\odot$ may form in merely $\sim 100\,\mathrm{Myr}$ (Kritos et al., 2023). Additionally, numerical simulations of supermassive black hole (SMBH) binary formation through galaxy mergers point to heightened gas compression and bursts of star formation at pericentric passages, ultimately driving the formation of dense NSCs. In these scenarios, newly merged SMBHs often become embedded in the central star cluster during the last merger stages (Van Wassenhove et al., 2014).

Recent studies emphasize the critical role of NSCs in accelerating massive black hole (MBH) mergers. For instance, Ogiya et al. (2020b) demonstrate that tidal in-

teractions between NSCs can strip stars and facilitate rapid MBH sinking, reducing separations from tens of parsecs to milliparsec scales. As NSCs undergo tidal stripping, the orbits of stripped stars expand, leading to a corresponding decrease in the angular momentum of the MBH binary and accelerating its orbital decay. The stripped material exerts a drag force on the binary, further enhancing the decay process as visualized in Figure 1.4. This mechanism is particularly relevant for lower-mass MBHs (e.g., $10^4\,M_\odot$), where dynamical friction alone may require gigayear timescales to bring MBHs from separations of over 100 pc to the galactic center.

Beyond isolated NSC mergers, multiphysics simulations indicate that protoclusters in high-redshift galaxies, which surround seed BHs, can undergo hierarchical merging, further expediting MBH binary formation (Shi et al., 2024). These processes are particularly important for understanding early-universe BH growth. Consequently, high-resolution N-body simulations of NSCs provide crucial insights into how stellar density, mass spectra, and orbital dynamics influence MBH binary formation and coalescence.

This thesis focuses on two key aspects of MBH dynamics in NSCs. First, we examine the impact of a realistic *mass spectrum* on MBH merger timescales. Many previous studies have neglected detailed mass distributions within NSCs, despite the fact that these clusters are collisional and subject to mass segregation. Second, we explore the role of NSCs in assisting seed MBHs that must sink to galactic centers for eventual binary formation. To investigate these processes at parsec-scale resolutions, we employ high-resolution N-body simulations (Chapters 3 & 4), offering deeper insights into the role NSCs play in shaping the fate of MBH binaries, particularly in high-redshift environments.

1.5 Dark matter spikes

Beyond massive black hole binaries, intermediate or extreme mass ratio inspirals (IMRIs/EMRIs) also rank among the most promising gravitational-wave (GW) sources for LISA. In many scenarios, a relatively small compact object (e.g., a stellar-mass black hole) gradually inspirals into a more massive black hole, radiating GWs over months to years in the LISA band. These long-lived signals are exquisitely sensitive to the mass distribution near the central MBH.

Recent theoretical work suggests that MBHs growing adiabatically through dark matter (DM) accretion can develop dense DM spikes (Gondolo and Silk, 1999). Such spikes are expected to leave a discernible imprint on the GW signal emitted by IMRIs/EMRIs. The heightened gravitational potential of a spike effectively modifies the orbital evolution of the inspiraling secondary, leading to additional dephasing that can amount to $\sim 10^3$–10^7 GW cycles (e.g., Eda et al., 2013; Eda et al., 2015; Kavanagh et al., 2020; Becker et al., 2022; Becker and Sagunski, 2023). Over the course of a year-long LISA observation, these cycles accumulate, making the DM spike's effect potentially observable.

Recent literature proposes that the key mechanism behind the spike-driven orbital modifications is dynamical friction (Chandrasekhar, 1943), arising from scattering events between the inspiraling body and the DM particles constituting the spike. While earlier studies (e.g., Eda et al., 2015) treated the spike as a static background, (Kavanagh et al., 2020) demonstrated that the inspiral itself can substantially heat and alter the spike, sometimes injecting energy comparable to its total binding energy. Consequently, to accurately capture the feedback of the inspiral on the spike (and vice versa), more realistic dynamical modeling is required.

In this thesis, we focus on improving such modeling by considering eccentric inspirals rather than purely circular orbits, given that eccentric orbits are common in astrophysical environments. We employ N-body methods, which are often regarded as the gold standard for dynamical simulations, to avoid relying on simplified assumptions about the feedback mechanism. By directly calculating forces between the inspiraling object and the spike particles, and incorporating post-Newtonian corrections, we aim to provide a more self-consistent picture of how DM spikes affect IMRIs and their observable GW signals.

1.6 Outline

This thesis focuses on the dynamical modeling of massive black hole (MBH) binaries in nuclear star clusters (NSCs) and intermediate-mass ratio inspirals (IMRIs) in dark matter (DM) spikes using N-body simulations. In Chapter 2, I introduce `Taichi` and present optimizations to FMM FMM for accurately simulating collisional systems. Chapters 3 and 4 explore the impact of NSCs on the evolution of embedded MBH binaries. Chapter 3 focuses on improving the modeling of NSCs by incorporating a two-component mass function. Using this framework, I investigate the differences in MBH binary evolution between mass-segregated and non-segregated NSCs, making GW timescale predictions. Chapter 4 extends this analysis to nucleated environments in high-redshift dwarf galaxies. Specifically, I examine how NSCs influence the merger of seed MBHs, potentially circumventing the seed sinking problem. These studies leverage realistic environments extracted from cosmological simulations, which include the stellar bulge, dark matter halo, and gas dynamics.

In Chapter 5, I introduce a fast N-body code designed to simulate eccentric IMRIs in DM spikes. Using this code, I identify the dynamical processes responsible for the dephasing effect in gravitational wave signals and predict the number of detectable dephasing cycles that LISA can observe over a five-year observation period. Additionally, I use these simulations to infer the conditions under which DM spikes can form and persist in realistic astrophysical environments.

In Chapter 6, I introduce a version of the HOLD integrator that runs entirely on NVIDIA general-purpose GPUs (GPGPUs). This proof-of-concept study aims to explore the performance of N-body codes that operate exclusively on GPUs. While further optimizations to the code are possible, this work demonstrates a fully GPU-

based solution that can be integrated with FMM codes running entirely on GPUs, paving the way for an efficient next-generation N-body code that runs directly on GPGPUs.

The works in Chapters 2, 3, 4, and 5 have been published in peer-reviewed journals (Mukherjee et al., 2021a, 2023, 2024, 2025).

Chapter 2

Fast multipole methods for N-body simulations of collisional star systems

Diptajyoti Mukherjee[1], Qirong Zhu[1], Hy Trac[1,2], Carl L. Rodriguez[1]

[1]McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213
[2]NSF AI Planning Institute for Physics of the Future, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Abstract

Direct N-body simulations of star clusters are accurate but expensive, largely due to the numerous $\mathcal{O}(N^2)$ pairwise force calculations. To solve the post-million-body problem, it will be necessary to use approximate force solvers, such as tree codes. In this work, we adapt a tree-based, optimized Fast Multipole Method (FMM) to the collisional N-body problem. The use of a rotation-accelerated translation operator and an error-controlled cell opening criterion leads to a code that can be tuned to arbitrary accuracy. We demonstrate that our code, `Taichi`, can be as accurate as direct summation when $N > 10^4$. This opens up the possibility of performing large-N, star-by-star simulations of massive stellar clusters, and would permit large parameter space studies that would require years with the current generation of direct summation codes. Using a series of tests and idealized models, we show that `Taichi` can accurately model collisional effects, such as dynamical friction and the core-collapse time of idealized clusters, producing results in strong agreement with benchmarks from other collisional codes such as `NBODY6++GPU` or `PeTar`. Parallelized using `OpenMP` and `AVX`, `Taichi` is demonstrated to be more efficient than other CPU-based direct N-body codes for simulating large systems. With future improvements to the handling of close encounters and binary evolution, we clearly demonstrate the

potential of an optimized FMM for the modelling of collisional stellar systems, opening the door to accurate simulations of massive globular clusters, super star clusters, and even galactic nuclei.

2.1 Introduction

The collisional N-body problem, in which the gravitational dynamics of N particles in a system are modeled over time, is one of the most challenging problems in modern computational physics. The stellar environments represented by such models, such as open, globular, and nuclear star clusters, contain some of the highest known densities of stars and compact objects and can produce many interesting astrophysical systems and transients. Systems such as X-ray binaries (e.g., Clark, 1975; Davies and Hansen, 1998; Ivanova et al., 2008; Hailey et al., 2018), recycled millisecond pulsars (e.g., Rappaport et al., 1989; Kulkarni et al., 1990; Sigurdsson and Phinney, 1995; Ye et al., 2019), cataclysmic variables (e.g., Ivanova et al., 2006; Pooley and Hut, 2006), and merging binary black holes (e.g., Portegies Zwart and McMillan, 2000; Rodriguez et al., 2015) can be produced with orders of magnitude more efficiency through dynamical encounters in dense star clusters than through typical stellar evolutionary processes. The compact objects within globular clusters are believed to be the sources of gravitational waves detected by LIGO (e.g., Abbott et al., 2020a,b,c, 2017). Collisions of stars and compact objects in the central region are thought to be responsible for the formation of intermediate-mass black holes (e.g., Freitag et al., 2006; Gürkan et al., 2006; Giersz et al., 2015), and possibly even the seeds of supermassive black holes at high redshift (e.g., Ebisuzaki et al., 2001). These black hole seeds grow along with their host galaxies and will be the targets of incoming space detectors such as LISA (Amaro-Seoane et al., 2017) and Tianqin (Luo et al., 2016).

To circumvent the difficulties associated with a direct integration approach to the N-body problem, approximate techniques such as Hénon-style Monte Carlo approaches (e.g., Hénon, 1971a,b; Giersz and Spurzem, 2000; Joshi et al., 2000; Pattabiraman et al., 2013; Rodriguez et al., 2016; Hypki and Giersz, 2017), or approximate solvers of the collisional Fokker-Planck (FP) equation (e.g., Vasiliev, 2017), are often used to follow the dynamics. The foundation of both approaches is a statistical treatment of uncorrelated two-body encounters over long time (Chandrasekhar, 1942). For these methods to work, the classic Chandrasekhar's formulae for dynamical friction (first-order) and diffusion (second-order) coefficients are derived under somewhat strong assumptions, which neglects coherent motions of each individual particles, i.e., resonances (Meiron and Kocsis, 2019) or self-gravity (Lau and Binney, 2019). Moreover, the uncertainties in the Coulomb logarithm need to be calibrated against the numerical method (Merritt, 2013). Therefore, the most versatile and adaptive method is direct integration of N-body system. However, direct N-body modeling is notoriously expensive to run and difficult to interpret (e.g., Miller, 1964).

With the increase in computational power, direct N-body methods have become

more accessible for performing larger simulations with $N \sim 10^6$. This makes them one of the preferred means to simulate star clusters today. Although accurate, the simulations are computationally expensive owing to the $\mathcal{O}(N^2)$ complexity of the direct forces calculation algorithm. For example, the DRAGON simulations (Wang et al., 2016) took $\sim$ 8600 hours using 160 `Xeon-2560` cores and 16 `K20m` GPUs to simulate a globular cluster containing 10^6 stars. Thus, solving the post-million-body problem using direct N-body codes presents a considerable challenge and is of interest to the astrophysical community.

One method that has been in widespread use is the differential treatment of long-range and short-range interactions. The Ahmad-Cohen scheme (Ahmad and Cohen, 1973), which has been adopted in `NBODY6` (Aarseth, 2003), allows the long-range interactions to be updated less frequently compared to the short-range interactions, thereby requiring fewer force calculations per average time. The alternatives to the expensive direct summation method are approximate force solvers. Among those, the Barnes-Hut (BH) tree (Barnes and Hut, 1986) is a natural candidate. Tree codes use hierarchical decomposition and multipole expansions to calculate gravitational forces between the particles. The latter results in an algorithmic complexity of $\mathcal{O}(N \log(N))$, which is a major improvement from that of direct summation methods. Although tree codes have found widespread usage in collisionless dynamics, they have found limited usage for collisional dynamics simulations so far (e.g., McMillan and Aarseth, 1993; Aarseth, 1999), the primary reason being the concern of force accuracy. Iwasawa et al. (2015) indicate that one of the other reasons may lie in the the fact that collisional simulations adopt individual or block time stepping. This would decrease efficiency since the particle tree would have to be reconstructed every time step. Despite this issue, P^3T (particle-particle particle-tree) codes which combine the force splitting with the Barnes-Hut tree have come out recently (e.g., Iwasawa et al., 2015; Wang et al., 2020). Wang et al. (2020) have demonstrated that their code is highly competitive compared to the direct summation code.

BH tree codes are not the only option here. Dehnen (2014) describes a more efficient approach compared to the traditional tree code by adapting the Fast Multipole Method ((FMM:) Greengard and Rokhlin, 1987; Cheng et al., 1999) to collisional dynamics. Dehnen (2014) presents various optimizations to the traditional FMM algorithm, which makes it suitable for adoption into star cluster simulations. An algorithmic complexity of $\mathcal{O}(N)$ and even sub-N complexity for special cases is demonstrated. The efficiency and the well-behaved error properties of this approach make it an attractive alternative to the tree codes. However, adaptation of FMM for collisional dynamics is missing. FMM has been used for collisionless N-body simulations before (e.g. Dehnen, 2000, 2002). Collisional simulations demanding higher accuracies and proper treatment of individual particles complicated the adoption of FMM.

An interesting question arises about whether a full-fledged N-body code using FMM could be as accurate and efficient as direct summation codes. The issue with

accuracy echoes same concern voiced by Press (1986). In this study, we adopt the FMM algorithm presented in Dehnen (2014) into our code, `Taichi` (Zhu, 2021), to perform various tests in collisional dynamics.

In Section 2.2 we describe briefly the modifications that were made to Zhu (2021) to adapt `Taichi` to collisional simulations. In Section 2.3 we describe the tests performed with the code. Section 2.4 goes into the results of the various tests and discusses the significance of the results. It is followed by discussion in Section 2.5, future work in Section 2.6, and conclusions in Section 2.7.

2.2 Taichi for Collisional Dynamics

As emphasized in the introduction, an efficient FMM algorithm is the core of our study. Zhu (2021) implements a collisionless N-body with FMM and individual time steps into a code called `Taichi`. In this work, we build on Zhu (2021) and make important modifications to enable it to be used for collisional simulations. The modifications are made in three different areas: usage of solid spherical harmonics instead of Cartesian coordinates for the multipole expansion, inclusion of an explicit error control algorithm, and, lastly, usage of rotation operators to speed up multipole expansions. In the following subsections we describe the overall structure of FMM and the modifications over Zhu (2021) in more detail.

2.2.1 Dual-tree Walk

Figure 2.1 illustrates the flow of FMM adopted as a force solver in `Taichi`. The particles are hierarchically arrange in an oct-tree. Each leaf cell first collects the multipole expansion with a Particle-to-Multipole (P2M) kernel. The multipoles are then passed recursively upward by the parent nodes until the root using a Multipole-to-Multipole (M2M) kernel. At this point, we have a complete description of all the cells for their far-field gravity if necessary.

Next, we pass the root to a dual-tree walk (Dehnen, 2000) and determine which interactions can be approximated. Our implementation of FMM uses a task parallel version of the dual-tree walk. Approximated gravitational force between the cells is calculated with a Multipole-to-Local (M2L) kernel for well-separated pairs of cells. The local expansions are then passed recursively down the nodes to the leaf cells. This step is achieved with a Local-to-Local (L2L) kernel. Finally, the force and the potential energy of each particle are determined based on the local expansions of the leaf cells they reside in using a Local-to-Particle (L2P) kernel. To calculate the near-field contribution to the leaf cells, which cannot be handled using multipole expansions, a direct summation is carried out instead.

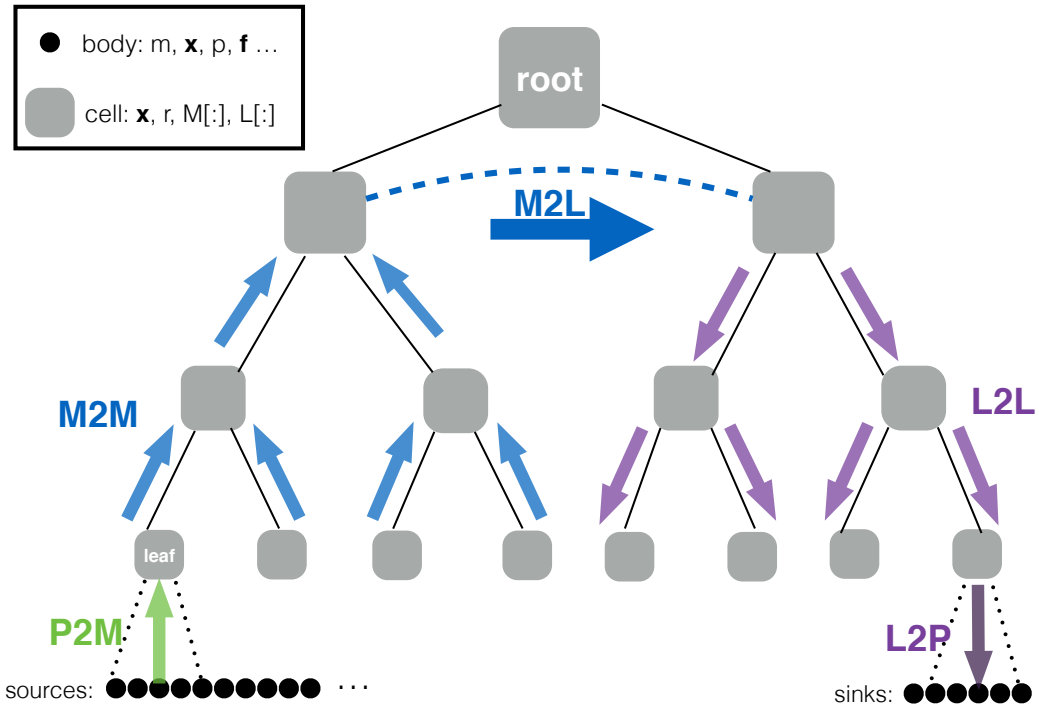

Figure 2.1: Flowchart (adapted from Yokota and Barba, 2012) illustrating various working parts of the FMM algorithm. For information on the abbreviations, check out section 2.2.3. The particles are arranged in an oct-tree, which provides a complete and hierarchical description of a set of particles, and will be ready for calculating the far-field of any cell consisting of a group of particles. The individual bodies contain information regarding the mass, position, velocity, and force experienced. The cells contain information regarding the position, radius, and multipole and local expansions. For clarity, we show the left half of the tree, consisting of only sources, and the right half, consisting of only sinks. The interaction list is obtained by a dual-tree walk (Dehnen, 2002) to determine which pair of cells can use approximations.

2.2.2 FMM with Solid Spherical Harmonics

We use the same architecture in Zhu (2021) for tree building, dual-tree walk, and direct summation. The FMM implementation in Zhu (2021) is based on a Taylor expansion in Cartesian coordinates (Dehnen, 2000), which is sufficient for collisionless systems where a modest force accuracy is needed. In collisional systems, a more stringent force accuracies are presumably necessary to follow changes to the particle's orbit over many dynamical times. While previous studies based on the BH tree method indicate that quadrupole expansion with opening angle $\theta = 0.5$ is sufficient, a systematic study of force accuracies on the collisional dynamics is desired. To be conservative, we consider that the approximate force aiming at the round-off level of single-precision floating point arithmetic should be safe. To that end, we adopt the approach using *solid harmonics* outlined in Dehnen (2014).

For $1/r$ interaction, the solid harmonics essentially form a complete and orthogonal basis for all the multipole moments and their far-field potentials. Up to order $p = 2$, the explicit expressions of these functions (table 3 in Dehnen, 2014) read

$$1, x, y, z, 3(x^2 - y^2), 6xy, 3xz, 3yz, 3z^2 - r^2; \tag{2.1}$$

One can see that these are the traceless part of the the Cartesian expressions (see also Coles and Bieri, 2020). We follow Equations (52), (53), and (54) in Dehnen (2014) to generate the regular (Υ_n^m) and irregular (Θ_n^m) parts of solid spherical harmonics. The former is used in the multipole moments

$$\mathcal{M}_n^m = \sum_i m_i \Upsilon_n^m(\mathbf{r}_i), \tag{2.2}$$

where the summation is taken over each particle i with mass m_i and displacement vector $\mathbf{r}_i \equiv (x_i, y_i, z_i)$. The latter is used in the expansion and translation operations. For more detailed mathematical properties of Υ_n^m and Θ_n^m and their translations, we refer the reader to Dehnen (2014). All the factorials present in the normalization coefficients are precomputed in a look-up table. The adoption of the solid harmonics enables us to truncate the expansion to a very high order, $p \gg 10$, which would otherwise become cumbersome for Cartesian expansions.

2.2.3 Error-controlled multipole-acceptance criteria

The second important modification is an error-controlled multipole-acceptance criterion (MAC) over the conventional varying θ as in Zhu (2021). We aim at some fractional force error for the force calculation: therefore, we adopt a scalar force as a quick but reasonable approximation of the actual accelerations according to

$$f_i = \sum_{i \neq j} \frac{G m_j}{|\mathbf{x}_i - \mathbf{x}_j|^2}. \tag{2.3}$$

Before the actual dual-tree walk, we first estimate the total scalar force f with $\theta < 1$. Each cell collects the contribution from far-field cells as $\sum \frac{M_c}{r^2}$, where M_c is the cell mass and r the separation between the cells. Next, the actual force calculation proceeds with the following MAC:

$$\theta < 1 \ \wedge E_{A\to B}\frac{M_A}{r^2} < \epsilon f_B \wedge E_{B\to A}\frac{M_B}{r^2} < \epsilon f_A \tag{2.4}$$

where f_A and f_B are the minimum of scalar force f for those particles in cells A and B respectively. The tolerance parameter ϵ directly controls the final force accuracies by FMM. We note that the above criterion which slightly differs from Dehnen (2014), additionally ensures that the forces among particles are symmetric at some extra cost.

The error coefficients $E_{A\to B}$ and $E_{B\to A}$ are entirely determined by the multipole moments of A, B themselves. With $\mathcal{P}_n$ defined as the sum of all the $(2n+1)$ multipole terms on order n as

$$\mathcal{P}_n^2 = \sum_{m=-n}^{n} (n-m)!(n+m)!\left|\mathcal{M}_n^m\right|^2, \tag{2.5}$$

where M_n^m are the multipole moments of the cell as in Eq. (2.2), and

$$E_{A\to B} = \frac{8\max\{r_A, r_B\}}{r_A + r_B}\frac{1}{M_A}\sum_{k=0}^{p}\frac{p!}{k!(p-k)!}\frac{\mathcal{P}_{k,A}r_B^{p-k}}{r^p}, \tag{2.6}$$

where M_A is the total mass of source cell A, i.e., its M_0^0 moment, r_A and r_B are the sizes of each cell, and r is their separation.

2.2.4 Rotation-accelerated M2L Operations

The last, but not the least, improvement is a fast M2L kernel. To speed up these expensive operations, we adopt the rotation-accelerated $\mathcal{O}(p^3)$ approach (Cheng et al., 1999; Dehnen, 2014). Additionally, we generate and save the swapping matrices for expansion order $p \leq 30$. These swapping matrices are essential for M2L translation operations, where the new z-axis is aligned with the interaction direction as follows: the multipole moments of the source cell are rotated in the z-direction, then with its x and z coordinates swapped, rotated in z-direction again, and with its x and z swapped again. The translation in the new coordinates features$\mathcal{O}(p^3)$ complexity instead of $\mathcal{O}(p^4)$.

We use `OpenMP` to speed up the dual-tree walk using task model with atomic clause used to update of multipole moments (Fortin and Touche, 2019). The near-field contribution is handled by a direct summation kernel, which is vectorized using `AVX` intrinsics as in Zhu (2021). Time integration closely follows the `HUAYNO` code (Pelupessy et al., 2012a) with little changes. A brief overview of the integrator and the time-symmetric time-stepping method used in `Taichi` is provided in the Appendix.

2.3 Tests

To measure how effective FMM is at simulating star clusters, we compare models of idealized clusters to the results of other codes and theoretical predictions. We use a homogeneous Plummer model (Plummer, 1911) to generate our initial conditions. The initial conditions are generated using the tool `MCLUSTER` (Küpper et al., 2011). Each test involves a number of independent realizations, and the results are derived after taking statistical averages over these realizations. For comparison, `NBODY6++GPU` (Wang et al., 2015) without GPU acceleration enabled is used to perform the direct N-body simulations. The same tests are also performed using the direct version of `Taichi` to ensure that the presence of the second-order `HOLD` integrator did not bias the results in any manner. In addition, Fokker-Planck simulations are performed using `Phaseflow` (Vasiliev, 2017) to compare density profiles over the course of evolution until core collapse. For the dynamical friction tests, `PeTar` (Wang et al., 2020) is used as a benchmark along with the other codes mentioned. In all of the tests Hénon (Heggie and Mathieu, 1986) units are used. In these units, $G = M = 1$ and $E_0 = -0.25$, where M is the total mass and E_0 is the initial total energy. A summary of the input parameters used for the simulations has been provided in Table 1.

All tests are done on 28-core `Intel Xeon E5-2635 v3` nodes. `Taichi`, `NBODY6++GPU`, and `PeTar` are run with only `OpenMP` and `AVX` enabled.

2.3.1 Tests Performed

Three different tests are performed in order to examine how FMM compares to direct N-body codes. In the first set of tests, we compare how accurate the FMM algorithm is compared to the direct summation method by examining force discrepancies. In the second set of tests, we evolve idealized Plummer models until core collapse to measure global properties, including conservation of energy, evolution of Lagrangian radii, core radii, and density function. Finally, we compare dynamical friction effects via the inspiral of a massive particle in a field of smaller mass stars. We also perform scaling tests to examine how `Taichi` scales with the number of cores within a node. The tests are presented in more detail in the upcoming sections.

Force Accuracies

We compare how the forces on individual particles vary between the direct and FMM versions of `Taichi` after a single time step. We construct a Plummer sphere with 10^5 particles using `MCLUSTER` and integrate it using both direct and FMM versions of `Taichi` with different input accuracies and multipole expansions. The relative force

accuracies are then computed using the L2 norm as follows:

$$\delta \mathbf{f} = \mathbf{f}_{\text{direct}} - \mathbf{f}_{\text{fmm}} \tag{2.7}$$

$$\frac{\delta f}{f} = \frac{|\delta \mathbf{f}|}{|\mathbf{f}_{\text{direct}}|} \tag{2.8}$$

We look at how the distribution of relative force accuracies varies with changing input parameters. Using the grid of simulations performed, we construct a heat map plotting the median and 99.99th percentile fractional force accuracies and histograms plotting the distribution of relative force accuracies. In addition, we look at the time taken by the FMM simulations with different input accuracy and multipole parameters to construct heat maps showing the variation of the integration time and the Poisson step time with the change in input parameters.

Core Collapse of a Plummer sphere

`MCLUSTER` is used to generate three sets of 16 simulations each containing $N = 1024$, 2048, and 4096 particles. The clusters are evolved until core collapse ($\sim 15 t_{\text{rh}}$) where t_{rh} is the half-mass relaxation time, which for the Plummer model is defined as

$$t_{\text{rh}} = 0.206 \frac{a^{3/2} N}{\log(\gamma N)} \tag{2.9}$$

where a is the characteristic scale or Plummer radius and $\log(\gamma N)$ refers to the Coulomb logarithm. All of the models start in virial equilibrium and do not contain any primordial binaries.

Post-collapse treatment is unfeasible since `Taichi` does not include regularization treatment for hard binaries, and as such the simulations take a long time to finish post-collapse. `Phaseflow` is used to simulate Fokker-Planck models of the clusters. The scale radius of the cluster is set to 0.59 which corresponds to Hénon units. The Coulomb logarithm ($\log \Lambda \equiv \gamma N$) is calculated by setting $\gamma = 0.11$. This value is representative of clusters with a uniform mass function (Giersz and Heggie, 1994).

Dynamical Friction

In order to measure whether FMM can accurately model dynamical friction, we seek to reproduce the black hole inspiral test performed by Rodriguez et al. (2018). In this test, a massive particle several times the mass of the stars in the cluster is introduced on a circular orbit at the viral radius of the cluster. Its position relative to the center of mass of the cluster is tracked. The time taken by the massive particle to inspiral to the center of the cluster can be modeled analytically. For a massive particle with mass m starting at a radius r with a circular orbit in a Plummer model, Rodriguez et al. (2018) provide the rate of change of r as

$$\frac{dr}{dt} = \frac{-8\pi G^2 \log \Lambda \chi m r}{V_c^3 \left[1 + 3(1 + \frac{r^2}{a^2})^{-1}\right]} \tag{2.10}$$

where $\log\Lambda$ is the Coulomb logarithm, m is the mass of the massive object, V_c is the circular velocity of the massive object at a distance r and a is the scale radius of the cluster. $\chi \equiv \mathrm{erf}(X) - 2X\exp(-X^2)/\sqrt{\pi}$, $X \equiv V_c/(\sqrt{2}\sigma(r))$, where $\sigma(r)$ is the velocity dispersion at a radius r has been used in equation 2.10. In Hénon units, $G = 1$ and $a = 0.590$. γ is set to be 0.01 in this equation (Rodriguez et al., 2018). These tests can also be used to determine γ since the analytic solution is very sensitive to the value of γ. More details on the derivation of this equation can be found in Rodriguez et al. (2018) and (Binney and Tremaine, 2011, chapter 8).

Using `MCLUSTER`, 30 independent realizations containing 10^4 stars are generated. The last star in the initial conditions is then replaced with either an object either 10 times more massive or 20 times more massive depending on the simulation. The massive object is placed one virial radius from the center of mass of the cluster on a circular orbit with velocity

$$V_c(r_{\mathrm{vir}}) = \sqrt{\frac{GM(r_{\mathrm{vir}})}{r_{\mathrm{vir}}}}, \tag{2.11}$$

where $M(r_{\mathrm{vir}})$ is the mass contained within one virial radius. Note that $V_c(r_{\mathrm{vir}}) \approx 0.799$ when $r_{\mathrm{vir}} = 1$.

Code	Input Parameters
`Taichi` FMM	$\epsilon = 10^{-7}$ $p = 20$ $\eta = 0.025$
`NBODY6`	`NNBOPT`$= 500$ $\eta_{\mathrm{I}} = \eta_{\mathrm{R}} = 0.01$
`PeTar`	$\theta = 0.3$ $\eta_{\mathrm{Hermite}} = 0.1$ $\Delta t_{\mathrm{tree}} =$ Default

Table 2.1: Summary of the input parameters used for the different codes for the tests mentioned above. Note that for the direct version of `Taichi` the same η was used. The input parameters for `NBODY6` were chosen in order to maximize the accuracy. In case of `PeTar` the tree timestep parameter is calculated automatically by the code from the changeover radius. Please note that these input parameters were also used for the simulations performed in the Appendix.

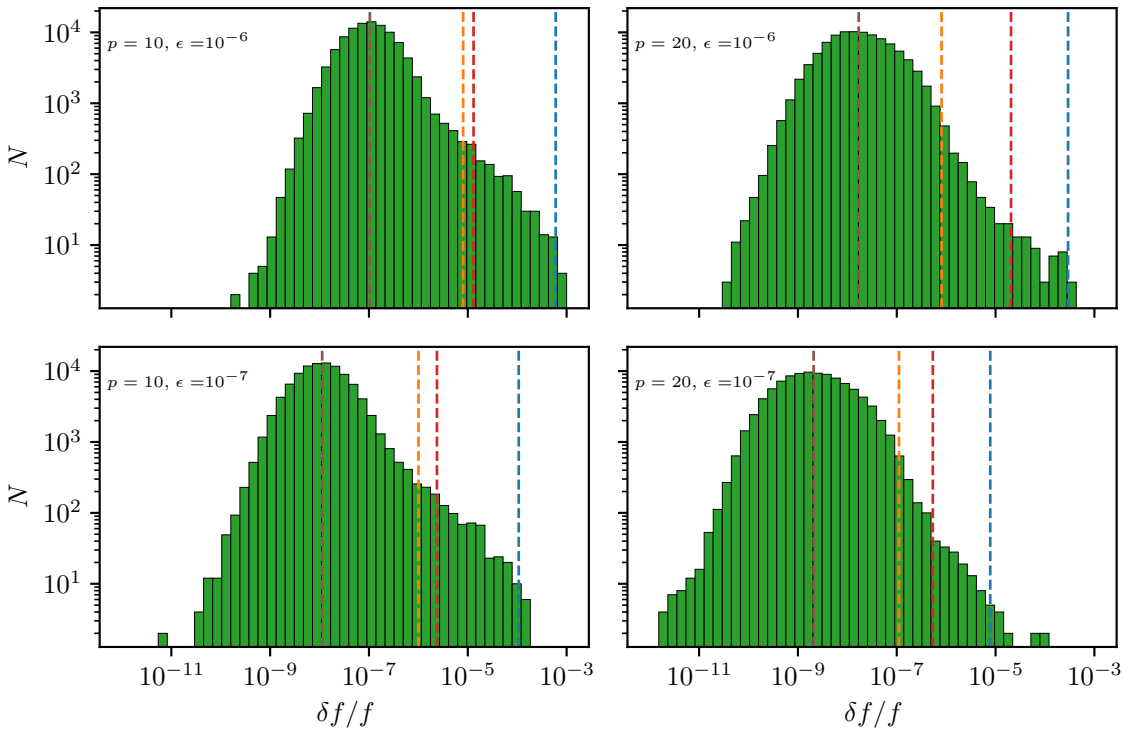

Figure 2.2: A histogram of the relative force error ($\delta f/f$). The particles are binned by their relative force errors. The brown line represents the median value, the yellow line represents the 99th percentile value, and the red line represents the rms value, and the blue line represents the 99.99th percentile values of relative force error. Increasing the multipole order shifts the distribution to the left and reduces outliers.

2.4 Results

2.4.1 Accuracy

The global energy error does not provide a full picture of the validity of the simulations. As Dehnen (2014) mentions, even though in FMM the energy error is indicative of the average force errors, it can cloak individual force errors that might be large enough to question the validity of the simulations. Therefore, it is imperative that we use another means to measure the validity of the simulations.

To determine the quality of force calculations, we look at the distribution of relative force errors for individual particles. In figure 2.2, we compare the distribution of relative force errors after one time step while varying the input accuracy and the multipole expansion parameters. We find that the median of the distribution is always better than the input accuracy and improves as we increase the multipole expansion parameter. We also notice that the 99th percentile values of the distribution improve after increasing the multipole expansion parameter. In fact, the 99th percentile value

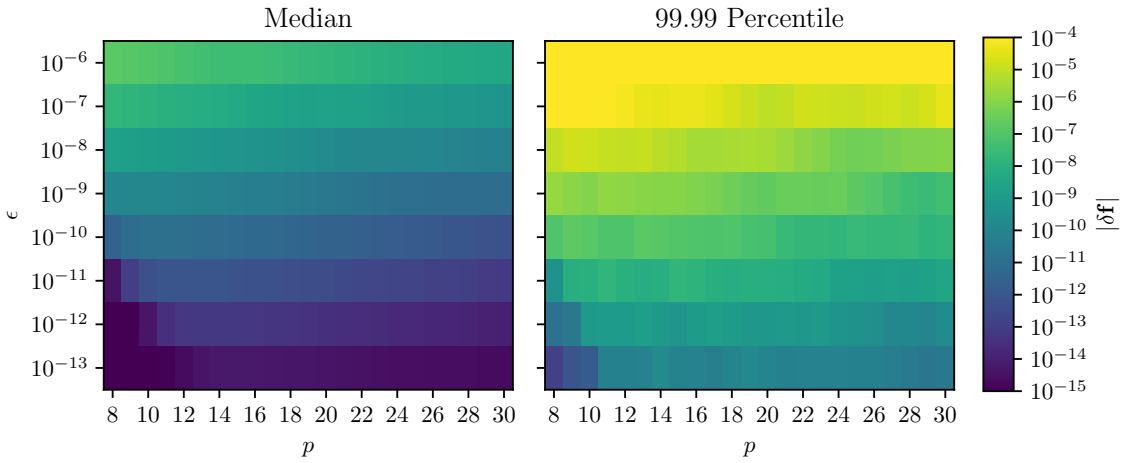

Figure 2.3: The relative force error distributions as functions of input accuracy (ϵ) and mulipole parameter (p). Left: the median relative force error is presented in this heat map. It is evident that the median fractional force error is extremely tightly controlled. In fact, for a given input accuracy, the median error is several orders of magnitude lower than it. Right: the 99.99th percentile fractional force error heat map is presented here. Since this value is representative of the number of outliers, we notice that the brighter patches indicate that the distributions contain more outliers than the darker portions. Within each row, there appears to be a fixed p for which the 99.99th percentile values are lowest.

is almost exactly equal to 10^{-7} when $\epsilon = 10^{-7}$ and $p = 20$. The right tails of the the distribution plots show that the force errors can sometimes go as high up as 10^{-3}. However, further analysis shows that the particles at the high-error tail of the distribution are located predominantly at large radii. For the particles at large radii, the magnitudes of the forces are small, which can additionally lead to misleading large fractional force errors (Dehnen, 2014). Also noted by Dehnen (2014), few of the particles at the high-error tail could also lie at the center of the cluster, where the forces mostly cancel out, leading to a small force that contributes to a large fractional error.

We conclude that the input parameters can be tuned in order to reduce the force discrepancy between the direct summation and the approximate values. To better understand how the parameters affect the distribution, we construct heat maps showing the distribution of the median and 99.99th percentile force error values since this gives us information about the distribution itself. Discrepancies between the two values indicate the length of the right tail of the error distributions, giving us an idea about the outliers. From the median value heat map in Figure 5.2 we find that force error is well controlled by the the input accuracy parameter itself. Within any particular row, we see that increasing the multipole expansion parameter increases the overall force accuracy. The exception to this rule is seen in the lower left corner, where we

have a combination of lower values of ϵ and p. For very low values of $\epsilon \leq 10^{-13}$, more cells are opened at low values of p rather than at relatively larger values of p because of the error estimation algorithm. In such cases, a number of cells can contain at most one particle, essentially reducing FMM to direct summation. This increases the relative force accuracy compared to direct summation but also results in a lot more pairwise force calculations, reducing efficiency. Thus, we find that the overall relative force accuracy decreases when we move from lower to higher p values.

2.4.2 Core Collapse of a Plummer Sphere

As a preliminary check, we analyze the growth of energy errors over the long-term evolution of the system as shown in Figure 2.4. We notice that the growth of energy errors is relatively slow toward the beginning of the simulation and increases more rapidly as the cluster approaches core collapse. For `Taichi` this is more evident since the usage of a symmetrized time steps ensures that the energy error grows more slowly in the beginning (Pelupessy et al., 2012a). However, due to the lack of regularization, close encounters or few body interactions can lead to jumps in the energy errors, especially close to core collapse. We present an analysis of energy growth with softening enabled in the Appendix .

It should be noted that in the case unsoftened version of `NBODY6++GPU`, the energy jumps are caused due to improper KS regularization switching as observed and noted by Wang et al. (2020). The cumulative energy conservation in the long term for the unsoftened version is typically $\mathcal{O}(10^{-4})$ for both `Taichi` and `NBODY6++GPU`. We notice that this is true for both direct summation and FMM versions of `Taichi`. For the softened version, we find that the energy conservation in the `Taichi` simulations is an order of magnitude or two better than that of the `NBODY6` simulations. However, we should not only use the overall energy conservation as a measurement of the accuracy or quality of the simulations (Dehnen, 2014; Wang et al., 2020). In fact, for tree and FMM codes, it is a better idea to study the distribution of force errors (section 2.4.1) along with the evolution of energy to get a better picture (Dehnen, 2014).

The first set of tests using the uniform mass Plummer model clusters reveal that `Taichi` models the long-term evolution of the clusters properly. This is evident visually from Figure 5.1 where we see the formation of the core and from the overlap of the Lagrangian radii curves in Figure 2.6. Here we have utilized `AMUSE` (Zwart et al., 2009; Pelupessy et al., 2013; Zwart and McMillan, 2018) to calculate the Lagrangian radius. For the $N = 1024$ model simulations, we find that the maximum relative difference between `NBODY6` and `Taichi` FMM among all 16 realizations is about 4% for the 1% mass fraction Lagrangian radius, but the average relative difference is about 1%. For the half-mass radius, the average relative difference is about 0.3%. The agreement between `Taichi` direct and FMM versions is even better, with the maximum relative difference across all mass fractions being close to 10^{-7}. As we increase N the agreement between the two methods improves considerably. For example, for

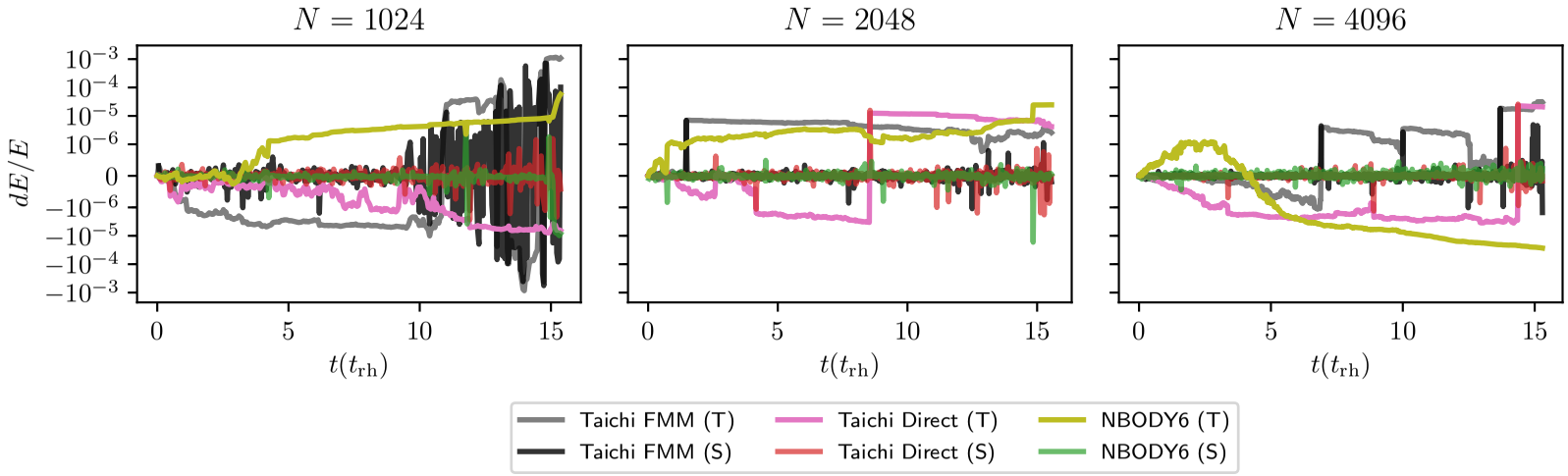

Figure 2.4: The relative energy errors represented as a function of the relaxation time. The curves with (T) denote the cumulative relative energy errors,m whereas the curves with (S) in them denote the relative energy error over one N-body time unit. The cumulative energy error starts growing more rapidly toward the end owing to the formation of hard binaries as the simulations approach core collapse.

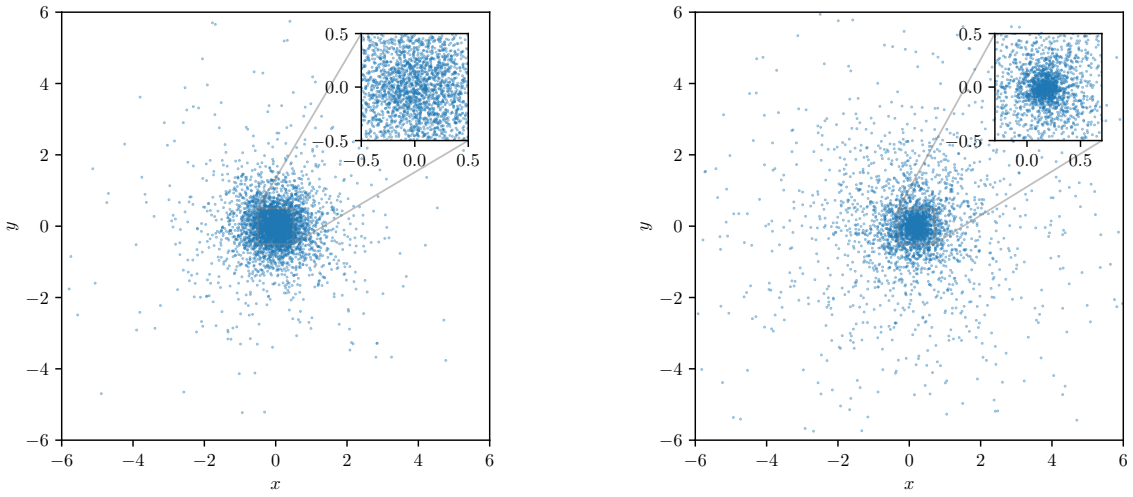

Figure 2.5: A cross-sectional scatter plot of an $N = 4096$ particle simulation run with `Taichi` FMM. Left: the particles at the initial timestep. The zoomed in area shows the region near the center of the cluster. Right: the particles right before core-collapse. One can clearly see the core that has been formed.

the $N = 4096$ particle simulations, the maximum relative difference between `NBODY6` and FMM in the Lagrangian radii across all four mass fractions is 0.9%, while the averages range between 0.05% and 0.1%.

The core radius is calculated using the definition provided by Casertano and Hut (1985). The core radius is defined as the density-square-weighted sum of the distance from the density center to the particle. Then, the core radius becomes

$$r_c = \sqrt{\frac{\sum_{i=1}^{N} \rho_i^2 |\mathbf{r_i} - \mathbf{r_d}|^2}{\sum_{i=1}^{N} \rho_i^2}}. \tag{2.12}$$

The density center is defined as

$$\mathbf{r_d} = \frac{\sum_{i=1}^{N} \rho_i \mathbf{r_i}}{\sum_{i=1}^{N} \rho_i} \tag{2.13}$$

where ρ_i is the density and is calculated by using a cubic spline kernel over the 32 nearest neighbors from the particle. The density is calculated using `HOPInterface` (Eisenstein and Hut, 1998) present inside `AMUSE`.

The core-radius curves in Figure 2.6 show agreement between all three codes. For example, in the $N = 1024$ particle simulations, the maximum relative difference in core radius is about 0.7%. With larger N the agreement becomes stronger with smaller deviation between individual simulations. We find that for the $N = 4096$ simulations, the maximum relative difference decreases to 0.1%.

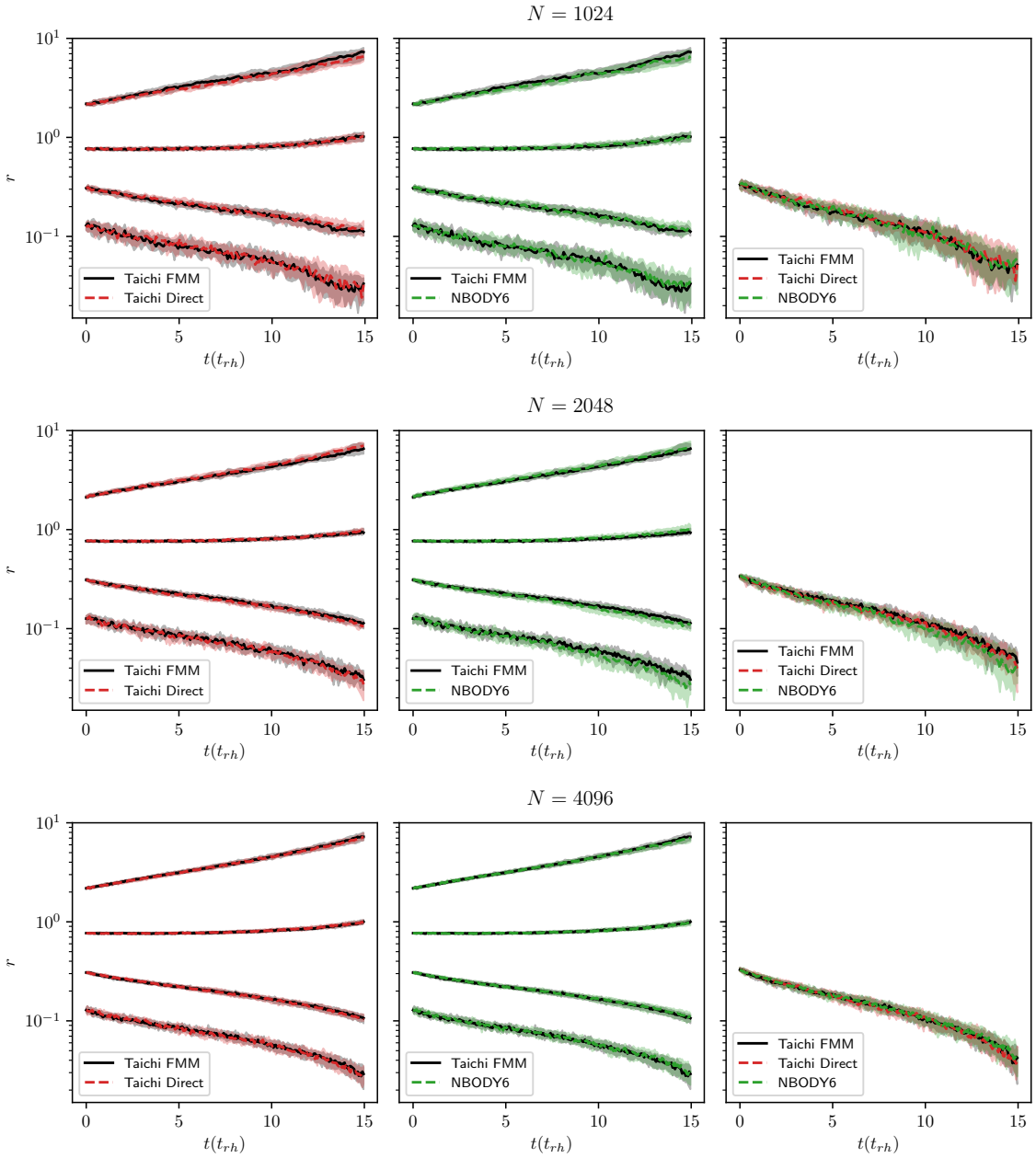

Figure 2.6: Evolution of Lagrangian radius using `NBODY6++GPU` and `Taichi` direct and FMM modes. From the bottom to the top, the curves represent 1%, 10%, 50%, and 90% mass fractions, respectively. The curves have been produced by taking the median of 16 independent realizations in each case. The shaded regions represent the 90th percentile values in each case. The rightmost plot shows the evolution of the core radius. It can be seen that as N increases, the agreement between FMM and the direct codes gets better.

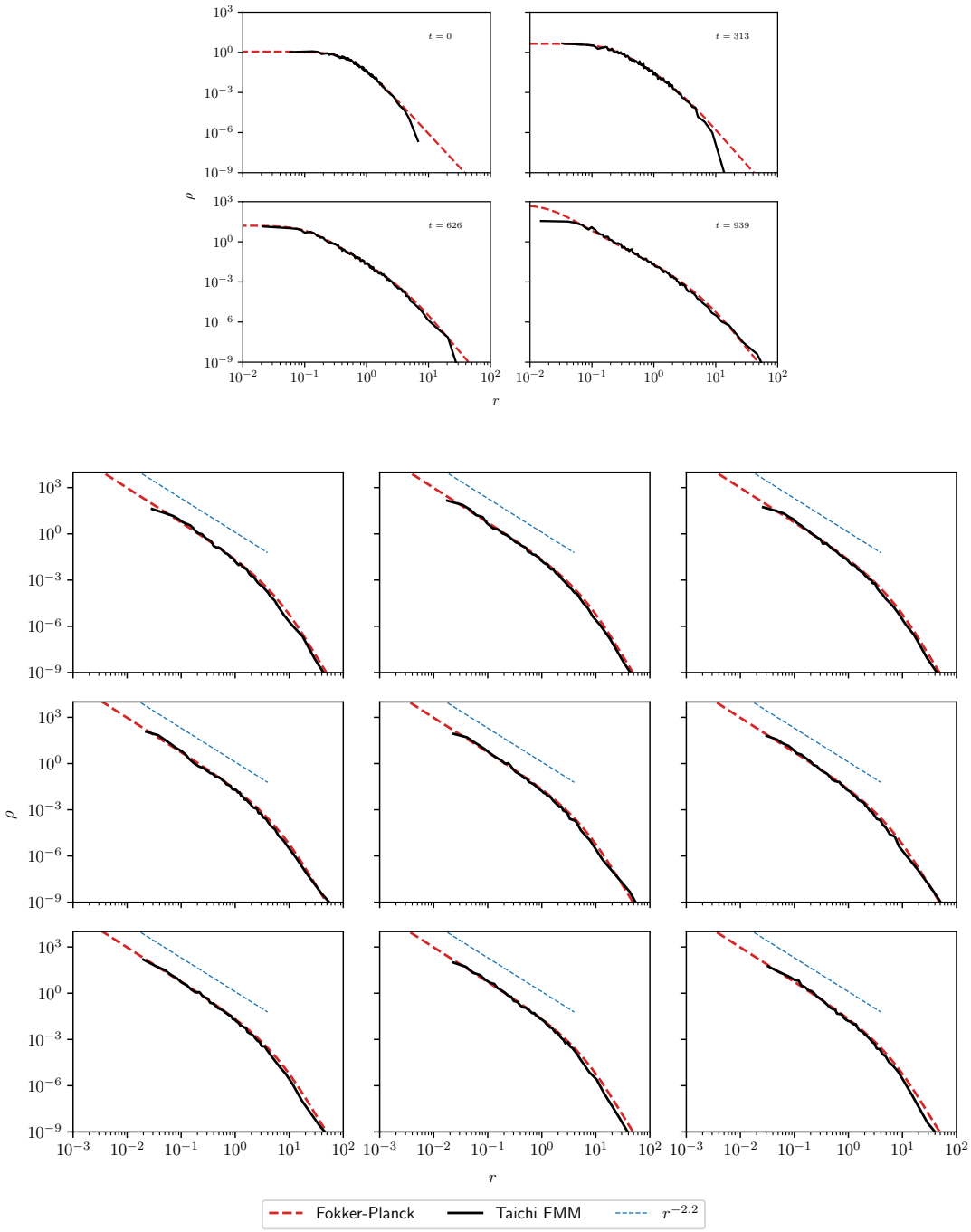

Figure 2.7: The density of the cluster ρ plotted as a function of the radius r. **Top:** The density function of a single 4096-particle realization simulated using FMM is compared to that produced by the Fokker–Planck code at different times during evolution until core collapse. The divergence between the codes at larger radii is caused by the dearth of particles present at larger radii initially. **Bottom:** The density functions of nine independent 4096-particle realizations compared to the density function produced using the Fokker-Planck code at the time of core collapse. The results show significant agreement between the two codes. This also indicates that the density function agrees with the theoretical power law of the density profile $r^{-2.2}$.

The agreement in the long-term evolution of the Lagrangian and core radii suggests that the evolution of the cluster density should be in agreement. We show this in Figure 2.7. Comparing the time evolution of the one realization of a 4096-particle model, we find that the density as a function of the radius produced by `Taichi` using FMM matches that of `Phaseflow` at different points during the evolution. Although not presented here, we found a similar picture for simulations with smaller particle numbers.

What becomes of considerable interest is the behavior of the density function at core collapse. In order to pinpoint the moment of core collpase, we simulate the evolution of an idealized Plummer model until core collapse using `Phaseflow` and compare the density functions at the time indicated by `Phaseflow` as the core collapse time. We compare nine independent realizations of the 4096 particle model to the idealized density function and find that there is a considerable amount of agreement between them. Some simulations could not be simulated to the core-collapse time owing to the formation of hard binaries (discussed further in section 2.5.2). The idealized density profile follows the theoretical density profile $\rho \propto r^{-2.2}$ (e.g. Joshi et al., 2000) and thus the density profiles produced by simulating the clusters using FMM also follow the theoretical density profile. This is significant since this phenomenon is purely driven by two-body effects, indicating that `Taichi` with FMM can model the two-body relaxation properly, a fundamental aspect of collisional N-body simulations.

In order to examine whether there are significant changes with accuracy parameters, we run `Taichi` with $\epsilon = 10^{-3}$ and $p = 10, 20$. We run a set of five $N = 1024$ Plummer model simulations until core collapse to examine the evolution of the Lagrangian radius and compare it to the previous `Taichi` FMM results. We also note the energy drift of the simulations using both the softened and unsoftened versions of the code.

From Figure 2.8, we find that there is very little change in the overall evolution of the Lagrangian radius compared to that of the more accurate simulations. This suggests that even with lower ϵ and p, `Taichi` is properly able to reproduce second-order relaxation effects. This is significant since simulations using lower accuracy parameters are faster. Although not presented here, we find that for the $p = 20$ scenario, even though the overall relative energy error does not exceed 10^{-6} at any point in time, the relative energy error between any two subsequent snapshots is larger compared to the more accurate simulations. For the $p = 10$ scenario, the overall relative energy errors are about 10x larger compared to the original simulations.

2.4.3 Dynamical Friction

Another important component of collisional simulations is dynamical friction, which is a purely two-body effect. We can examine it in our tests via the inspiral of a massive object in the field of less massive stars. The rate of inspiral can provide us a direct

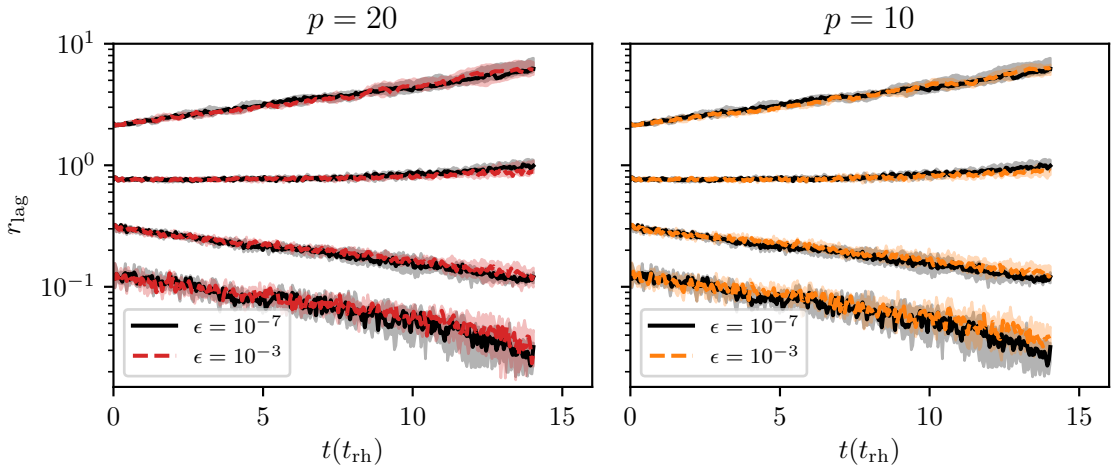

Figure 2.8: The evolution of the Lagrangian radius for five $N = 1024$ simulations is presented here, similar to Figure 2.6. We notice that, even while using lower accuracy parameters, we arrive at a similar evolution of Lagrangian radii of different mass fractions. Compared to the original results, we find that the maximum relative difference in the Lagrangian radii of individual simulations is of the order of 0.001%.

idea of the ability of a method to reproduce two-body dynamical effects properly. In these tests, we emphasize that we are using the `NBODY6` and `PeTar` results as a benchmark rather than the analytic results. This is because of the inability of the Chandrasekhar model to reproduce the position of the massive particle near the core. This issue is discussed in detail later.

We notice from figure 2.9 that `Taichi` with FMM is statistically able to reproduce the inspiral rates for both of the tests, agreeing with both the direct N-body results and the analytic results. The agreement between individual simulations is, however, not guaranteed. Individual simulations, even though they may agree at the beginning, can vary significantly. For example, for the $M_{\rm BH}/M_{\rm star} = 10.0$ case, the average relative difference between `NBODY6` and `Taichi` with FMM for the positions of the black hole at the end of the simulation was about 20%. Even individual simulations performed can vary considerably over multiple runs. The same initial conditions can produce different inspiral rates if simulated multiple times. It is an artifact of the nonassociativity of floating point operations for multithreaded programs. Even machine-precision errors ($\sim 10^{-16}$) can grow exponentially over time, and results may diverge after a couple of dynamical times. This is a result of Miller's instability, and this issue has been discussed in more length in a later section. It is thus imperative to perform multiple simulations and use the statistical average of the results rather than results from a single simulation.

The discrepancy between the N-body and the analytic results are toward later

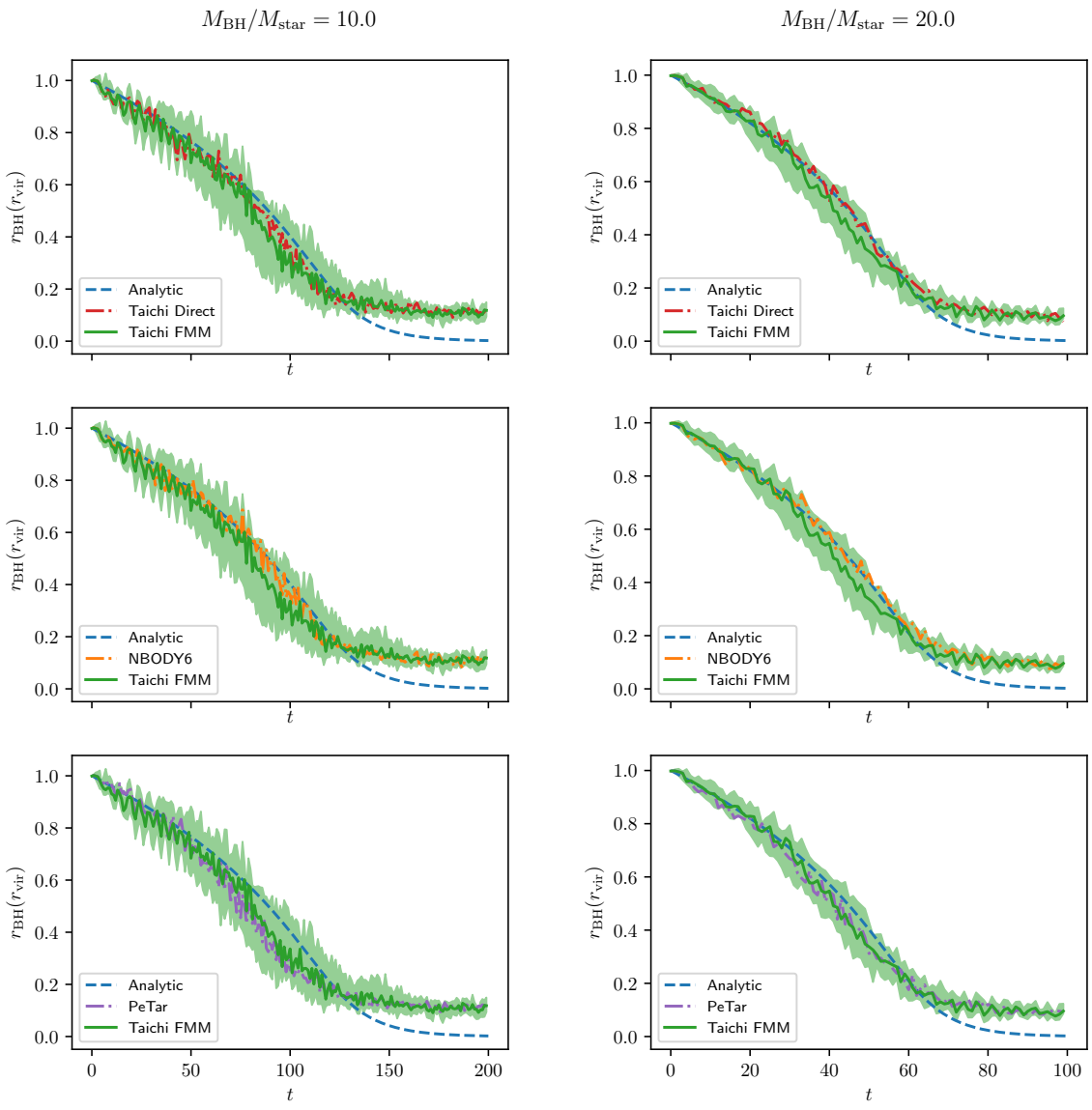

Figure 2.9: The distance of the massive object ($r_{\rm BH}$) is presented as a function of time (in Henon units) and the virial radius of the cluster. The curves show the inspiral of massive objects of two different masses due to dynamical friction. The solid and dashed curves indicate the median distance of the massive object from the center of mass that was produced after running 30 independent realizations. The shaded regions indicate the 95% confidence interval values of the median distance for the FMM simulations. All values are binned over one N-body time step. We notice that as we increase the mass of the massive object, the agreement between the different methods improves significantly.

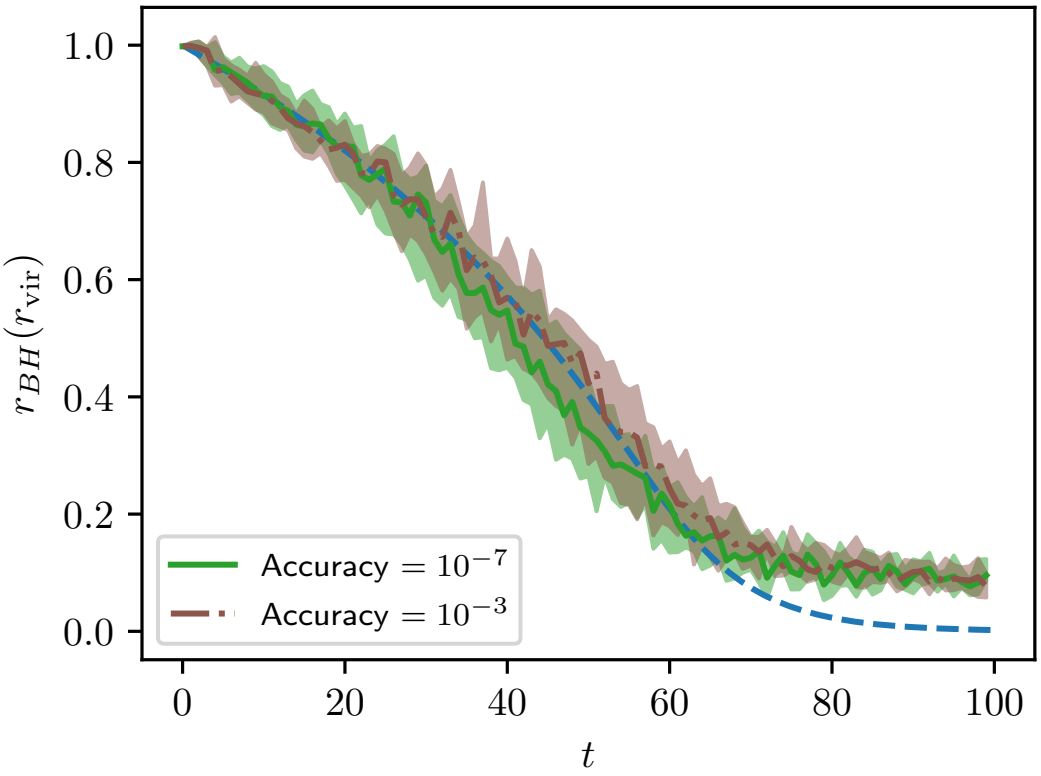

Figure 2.10: Same as the $M_{\rm BH}/M_{\rm star} = 20.0$ case from Figure 2.9 but with two different input force accuracies: $\epsilon = 10^{-3}$ and $\epsilon = 10^{-7}$. Even with an input accuracy four times lower in magnitude compared to the original FMM simulations, the massive particle inspiral time is reproduced very well.

time steps is caused by the "core stalling problem." The issue has been noted by Goerdt et al. (2006) and others (e.g. Inoue, 2009; Goerdt et al., 2010) performing N-body simulations involving the inspiral of objects in gravitational systems with cores. The stalling represents a flaw in the Chandrasekhar model of relaxation that assumes a Maxwellian distribution of velocity and spherical symmetry that is not perfectly reproduced in discrete models. According to Goerdt et al. (2006), the stalling is caused by an orbit-scattering resonance in which the perturber and the background reach a stable state. Semianalytic models can be used to correct for the stalling effect (e.g., Petts et al., 2016; Silva et al., 2016).

The question whether `Taichi` FMM using a lower input accuracy and multipole expansion order can reproduce similar inspiral times to that using higher accuracies is interesting. For example, if we used $\epsilon = 10^{-3}$ and $p = 10$ instead of $\epsilon = 10^{-7}$ and $p = 20$, should we expect results that agree with those from earlier? Figure 2.10 suggests that we should in fact find that the results to be in agreement. This

is a very important result that suggests that lower-order FMM can be used in cases where we want to model dynamical friction on a few specific particles. Switching to a lower order can save time. In fact, in our simulations, switching to a lower order sped up the simulations by $\sim 2-3$ times. This can have applications in modeling supermassive black hole (SMBH) binaries in a field of smaller stars. We expect the agreement to be good as long as the mass ratio of the massive object to that of the field stars is high enough. Whether low-order FMM can model dynamical friction in cases where the mass ratio is closer to 1 needs to be tested. However, results do suggest that FMM can be used to simulate SMBH binaries and intermediate-mass black hole binaries safely.

2.4.4 Scaling

We perform scaling comparisons and strong scaling tests for `Taichi` on a single 28-core `Intel Xeon E5-2635 v3` node. All of the wall-clock times have been averaged over wall-clock times from five individual simulations of each realization. The `OpenMP` parallelization in Taichi allows us to scale the code over multiple cores in a single node. We run `Taichi` FMM with input accuracies of 10^{-7} and 10^{-3} and a multipole expansion $p = 20$ and compare it to `Taichi` direct, `NBODY6` and `PeTar`. They are then used to perform simulations for $10^3, 10^4, 10^5$, and 10^6 particles. The systems are evolved until $t_{\rm final} = 1/8192$ N-body units. This was used to obtain quick results for large-N systems. A summary of the input parameters has been provided in Table 2.2.

Code	Input Parameters
`Taichi` FMM	$\epsilon = 10^{-7}, 10^{-3}$ $p = 20$ $\eta = 0.025$
`NBODY6`	`NNBOPT`$= 200$ $\eta_{\rm I} = \eta_{\rm R} = 0.01$
`PeTar`	$\theta = 0.3$ $\eta_{\rm Hermite} = 0.1$ $\Delta t_{\rm tree} = 1/8192$

Table 2.2: Input parameters used for the scaling tests. Please note that in this case the input parameters for both `NBODY6` and `PeTar` have been changed slightly compared to Table 2.1.

To find optimal input parameters, we construct integration and Poisson step time heat maps as shown in Figure 2.11. We find that our choice of input parameters ($\epsilon = 10^{-7}, p = 20$) is optimal. Looking at Figure 2.12, we notice that FMM is inefficient for simulations with fewer than $\sim 10^4$ particles. This is in part due to the

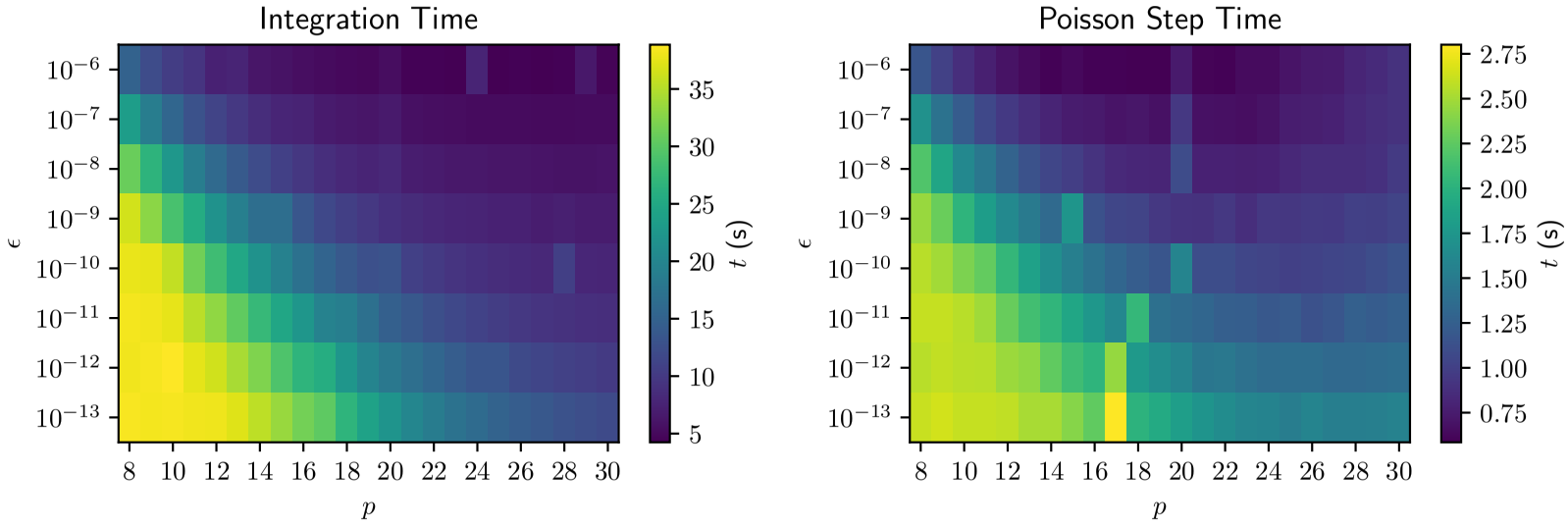

Figure 2.11: Heatmaps showing the distribution of wall-clock time as a function of both p and ϵ. Both integration and Poisson step times are determined for evolving a 10^5 star cluster to 1 timestep. *Left*: The total integration time. It essentially represents how long it takes for `Taichi` in total. *Right*: This heatmap only reprsents the amount of time spent computing the forces.

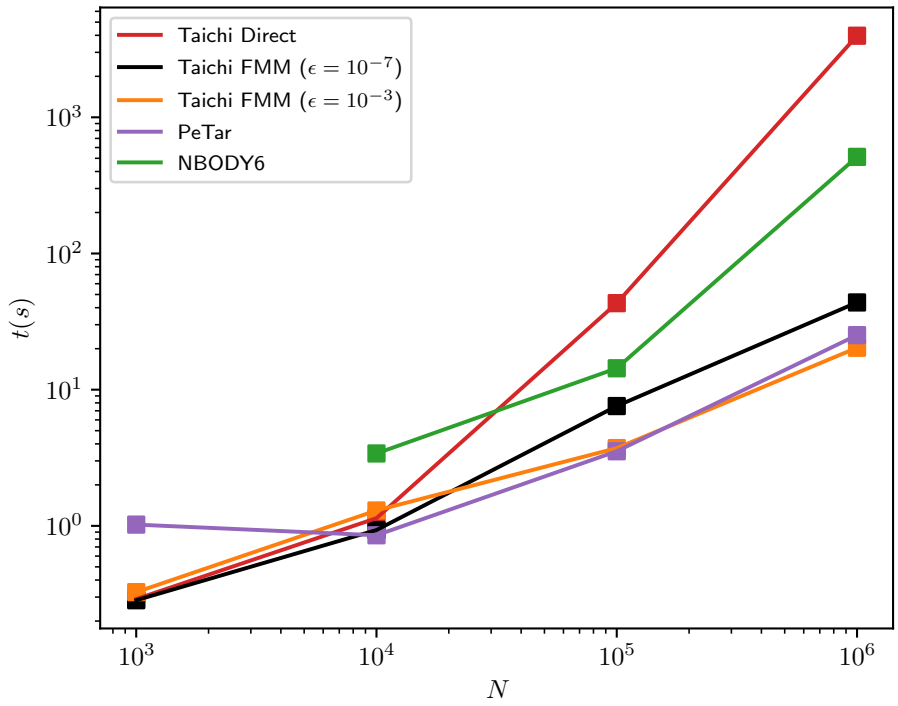

Figure 2.12: The wall-clock time for one integration step presented as a function of the problem size. For $N < 10^4$ direct summation is more efficient. However, owing to the $\mathcal{O}(N)$ scaling, for large N, FMM becomes highly efficient.

tree building process, which proves to be inefficient compared to the direct algorithm for a smaller number of particles. However, past that threshold, it becomes more efficient. For instance, for the million-particle simulation, FMM using $\epsilon = 10^{-7}$ is more than 10x faster than `NBODY6`. We also find that although FMM using $\epsilon = 10^{-7}$ is 1.7x as slow as `PeTar` for a million particles, the version using $\epsilon = 10^{-3}$ is 1.25x faster than `PeTar`. We also find that FMM using smaller ϵ conserves energy to 1 part in 10^{12} whereas the version using higher ϵ conserves energy to about 1 part in 10^9. This is about an order of magnitude smaller than that of `PeTar`.

For the strong scaling test, we simulate a cluster containing 10^6 stars using $1, 2, 4, 8, 16$, and 28 cores. An input accuracy of 10^{-7} and a multipole expansion of $p = 20$ are used again. The speedup is computed as the ratio of the wall-clock time of the single threaded simulation to that of the multithreaded simulation. As is evident from Figure 2.13, `Taichi` FMM scales as dictated by Amdahl's law. The graph also indicates that the maximum speedup is not reached on 28 cores and is therefore limited by the number of cores available to us.

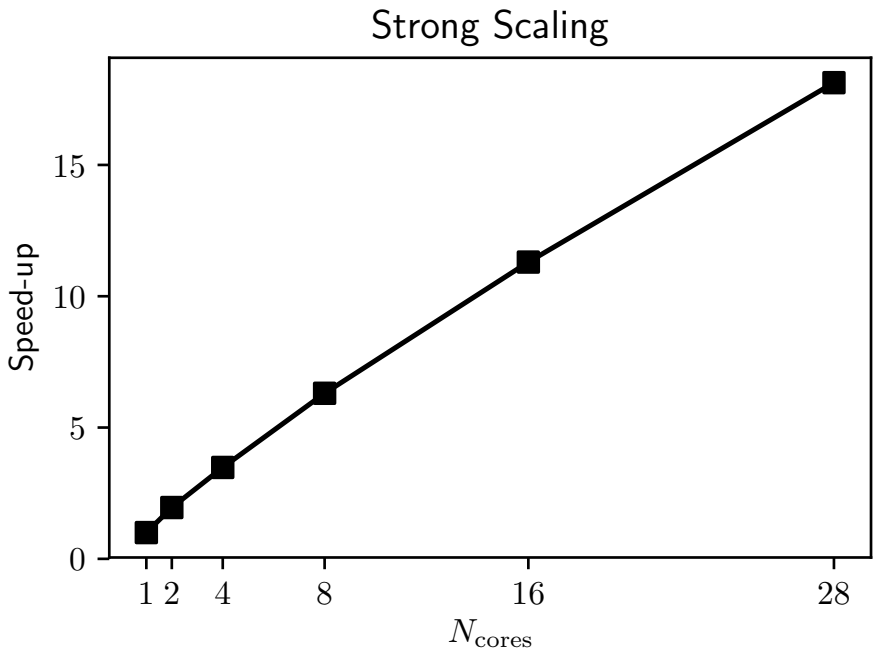

Figure 2.13: The overall speedup presented as a function of the number of physical cores used. This determines the intranode scaling of the FMM force determination algorithm. The Poisson step time has been used to determine the scaling. The overall scaling follows the same pattern.

2.5 Discussion

2.5.1 Parallelization and Miller's instability

As briefly discussed in section 2.4.3 floating point arithmetic can play an important part in the outcome of individual simulations. Floating point arithmetic is inherently nonassociative in nature (e.g. Villa et al., 2009). This is particularly exacerbated in the case of multithreaded floating point operations. For example, reduction operations can lead to different round-off errors during different runs of the same program. In iterative solvers, the results are propagated through various iterations and at the end can produce different round-off errors (e.g., Villa et al., 2009). Force calculation relies on iterating over particles and cells and at each time step. As such, any dynamical change in the ordering of threads between two different runs of the same program can lead to discrepancies in the results between two simulations. This is not a feature of the FMM algorithm. This artifact is present in direct summation as well. In our case, analysis reveals that for a simulation containing 1024 particles, the maximum discrepancy between the forces calculated on individual particles between two runs is $\mathcal{O}(10^{-16})$. Out of caution, the serial version of the code was also run multiple times,

but no discrepancies were found. This is consistent with round-off errors resulting from dynamical ordering of threads. Any single run of the force algorithm over all particles results in errors of this order. However, even differences of such small order can lead to major discrepancies between the positions and velocities of particles at later times. This is triggered owing to Miller's instability. Over the course of a few dynamical times, the difference between the position and velocity of a particular particle grows exponentially. Although not presented here, we noted that between $t = 1$ to $t = 10$ for an $N = 1024$ particle simulation the maximum difference in the position over all particles grows exponentially from 10^{-16} to 10^{-2}. This is consistent with Miller's instability. This discrepancy only presents itself explicitly when we are looking interested in tracking properties of individual particles. Global properties like energy conservation, evolution of half-mass radius, etc., remain consistent over simulations. This further reiterates the importance of performing multiple simulations and drawing statistical averages rather than relying on single simulations.

2.5.2 Integration Issues

Due to a lack of a dedicated regularization scheme, the integrator is sometimes forced to spend a lot of time integrating hard binaries in our test. This "slowing down" of the simulation becomes more apparent as the simulation approaches core collapse or if they contain primordial binaries. In the process of evolving some of our simulations to core collapse, we noticed that the formation of even one hard binary significantly increased the time required to evolve the system further. For example, in a particular realization containing 1024 stars, we noticed that the simulation basically halted after 297 time units. Further analysis showed that a binary, with stars having time steps several orders of magnitude smaller than the average time step, was the culprit. One way to alleviate this issue could be to include special treatment of isolated and perturbed binaries such as regularization.

2.6 Future Work

In a cluster of a million bodies, many primordial binaries are weakly perturbed for most of the time (or regarded as entirely isolated as in Monte Carlo codes), therefore permitting an efficient treatment of their orbits. For an optimal treatment of binary and few-body systems, we seek to integrate a regularization scheme with a future version of `Taichi`. One of the potential regularization schemes includes Slow Down Algorithmic Regularization (SDAR: Wang et al., 2020) which has been included in `PeTar`.

Integration efficiency can also be improved by increasing the integration order, which would allow the usage of larger time steps. A higher-order scheme could also allow the usage of more optimized time step calculation schemes like the Aarseth

scheme (Aarseth, 2003). Higher-order integration schemes would require the calculation of jerks which is nontrivial with the FMM algorithm.

We have implemented the approach by Dehnen (2014) to calculate the jerks, i.e. the time derivative of forces. Therefore, FMM can be incorporated into a traditional fourth-order Hermite codes tat updates the positions and velocities using the information up to jerks. Alternatively, a hierarchical version of the force-gradient integrator recently proposed by (e.g., Rantala et al., 2021a) is also promising.

As the Aarseth step function is widely used in the Hermite integrator, the adaptive stepping is not time symmetric such that a secular energy drift is present (e.g. Hut et al., 1995; Dehnen, 2017). This energy drift is present even if the time symmetric version of Hut et al. (1995) is used. An implicit scheme by Makino et al. (2006a) is proposed, but requires many iterations and is therefore unpractical. We adopted an approximate time-symmetric method introduced in Pelupessy et al. (2012a), taking the derivative of time steps into account. This approach can be generalized with the recent method based on the tidal tensor by Grudić and Hopkins (2020) as the tidal tensor can be easily calculated by FMM, as well as its time derivatives.

2.7 Conclusions

In this work, we have described a collisional N-body code, `Taichi`, which incorporates a novel method of calculating forces using FMM. In our implementation, we split up the forces into short range and long range. The former is calculated via direct summation, whereas the latter is calculated using FMM. This results in an algorithmic complexity of $\mathcal{O}(N)$ rather than the expensive $\mathcal{O}(N^2)$. This makes post-million-body simulations viable.

Through various tests, we demonstrate that `Taichi` can be used to perform collisional stellar system simulations. In the first set of tests, we show that by tuning two input parameters, the mutipole expansion order (p) and the input accuracy parameter, (ϵ), we can tightly control the force errors. The median and 99th percentile values are constrained by the input accuracy. The RMS error values are more weighted toward outliers and can be reduced by increasing p for a given ϵ.

The second set of tests was used to compare long-term behavior of `Taichi` with that of `NBODY6++GPU`. The relative energy error remained below 10^{-4} for `Taichi` and only grew sharply as the simulations approached core collapse. The evolution of Lagrangian radii for different mass fractions and core radius shows agreement between both of the codes. This shows that using an approximate force solver like FMM is as good as direct summation for reproducing global properties. Comparison of the density profile with a Fokker-Planck code also shows agreement. At core collapse, the agreement of the density profiles indicates that the realizations simulated with FMM follow the theoretical density power law. This indicates that the approximate force solver is able to reproduce two-body relaxation effects since the theoretical power law is a result of the two-body effects.

Dynamical friction tests allow us to arrive at the same conclusion. The median inspiral times of objects several times the mass of the stars in the clusters closely follow the analytic results and are in agreement with those of `NBODY6++GPU`. Furthermore, we demonstrate that we can reproduce proper dynamical friction effects even with a considerably lower input accuracy and multipole expansion order.

Compared to the direct version of `Taichi`, the FMM version speeds up the integration over 100 times for a simulation containing a million stars. We also find that `Taichi` FMM is more than 10x faster than `NBODY6++GPU` on a 28-core machine. However, in our current implementation, FMM becomes effective only for simulations containing more than 10^4 stars. Several bottlenecks are also present in the code. The lack of a proper regularization scheme makes simulations with binaries virtually impossible. The lack of a higher-order scheme also implies that the code takes smaller time steps which hinders the efficiency. While close binaries or small-N subsystems indeed require extra care, we have shown that approximate force solvers are sufficiently accurate to simulate collective effects due to the uncorrelated two-body encounters in the sense of Chandrasekhar. It is foreseeable that FMM can be combined with Ahmad-Cohen scheme for the regular force calculations.

2.8 Appendix

2.8.1 Time step symmetrization

We review the time integration based on the hierarchical Hamiltonian splitting here. The Hamiltonian of the N-body system consists of the total potential term

$$\mathcal{V} = \sum_{i<j} \sum_{j} \frac{m_i m_j}{r_{ij}}, \tag{2.14}$$

and the total kinetic term

$$\mathcal{T} = \frac{1}{2} \sum_{i} \frac{\mathbf{p}_i^2}{m_i}. \tag{2.15}$$

For most collisional systems, there exists a wide range of dynamical timescales, defined by both the smooth orbit of a particle in the mean field potential and interactions between individual particles. To speed up the calculations, individual time stepping has been used since Aarseth (2003). In Pelupessy et al. (2012a), this was implemented by splitting the Hamiltonian

$$\mathcal{H} = \mathcal{H}_F + \mathcal{H}_S, \tag{2.16}$$

where the original Hamiltonian $\mathcal{H} = \mathcal{V} + \mathcal{T}$ is decomposed into a fast and slow subsystem based on the step size assigned to each particle. We adopt the step size criteria from Pelupessy et al. (2012a), which is constrained by both the freefall time,

$$\tau_{\text{freefall}} = \eta \sqrt{\frac{r_{ij}}{a_{ij}}} \tag{2.17}$$

and a flyby time

$$\tau_{\rm flyby} = \eta \frac{r_{ij}}{v_{ij}}. \tag{2.18}$$

The slow system contains the contributions from both slow particles and the cross interaction between slow and fast particles as

$$\mathcal{H}_S = \mathcal{T}_S + \mathcal{V}_{SS} + \mathcal{V}_{FS}, \tag{2.19}$$

The fast subsystem now only consists of fast particles

$$\mathcal{H}_F = \mathcal{T}_F + \mathcal{V}_{FF}, \tag{2.20}$$

where $\mathcal{T}_F$ is the kinetic energy of fast particles and $\mathcal{V}_{FF}$ consists of potential energy solely from fast particles. Now, $\mathcal{H}_F$ can be integrated separately from the slow system, where the forces between the fast and slow systems need to be calculated at the pace of the slow system. The integration then proceeds recursively to $\mathcal{H}_F$. The above procedure leads to a second-order accurate and momentum-conserving scheme. One subtle point is that the use of individual time steps breaks the time symmetry of the leap-frog integrator, leading to a drift in the total energy. To counter this, we adopt an approximate time-symmetric stepping function that is introduced by Pelupessy et al. (2012a) which removes the iterations required by an implicit time-symmetric scheme (Hut et al., 1995; Makino et al., 2006a). The idea is to incorporate the time derivatives of Eq. (2.17) and (2.18) to construct a first-order-accurate estimate of step size into the future according to

$$\tau_{\rm sym} = \tau (1 - \frac{1}{2} \frac{d\tau}{dt})^{-1} \tag{2.21}$$

As a result, this treatment removes a secular energy drift often associated with individual time steps in long-term evolution of N-body systems. An alternative form to Eq. 6.4 is given by Rantala et al. (2021a). In Figure 2.14, we simulate 10 different $N = 100$ Plummer models to 50 N-body units and find that the time-symmetrized version of the `HOLD` integrator conserves energy better than the unsymmetrized version.

2.8.2 Energy error associated with hard binaries

`NBODY6` has sophisticated treatment of binaries and multiples, which are absent in `Taichi`. To access the impact of hard binaries, we modify the Newtonian gravity for both `NBODY6` and `Taichi` to incorporate a Plummer softening $\phi(r) = (r^2 + \epsilon_{\rm soft}^2)^{-1/2}$, where $\epsilon_{\rm soft}$ is the softening length. We then simulate Plummer models of $N = 1024, 2048$, and 4096 with $\epsilon_{\rm soft} = \frac{1}{N}$. This estimate of close encounter distance follows from Dehnen and Read (2011), where $\epsilon_{\rm 2body} = \frac{2R}{N}$ where R is a characteristic radius of the system. In N-body units, $2R \sim 1$, so $\epsilon \sim \frac{1}{N}$.

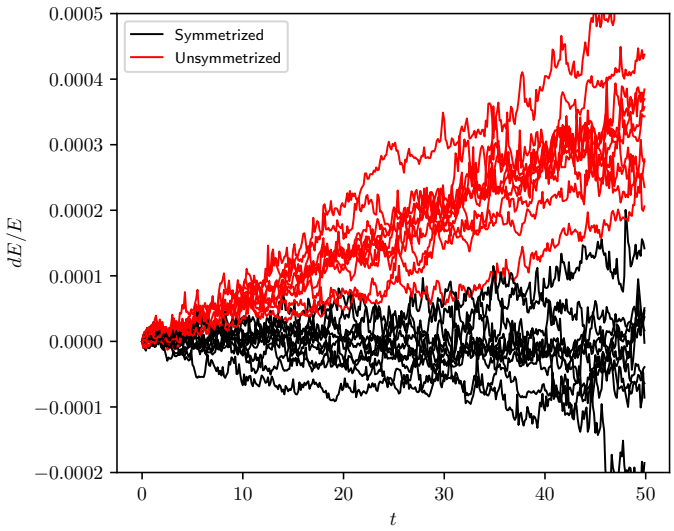

Figure 2.14: The relative energy error as a function of the N-body time here for 10 Plummer model realizations. For this simulation we used the `HOLD` integrator and the same softening length and step size as that of Makino et al. (2006a). We find that the symmetrization scheme helps remove the energy drift, identical to Makino et al. (2006a) and Pelupessy et al. (2012a).

We find that when softening is enabled, the drift in energy is smooth over time and is virtually similar for both `Taichi` direct and FMM. We notice this for all simulations as shown in Figures 2.15 & 2.16. This signals that the jumps in energy observed before are primarily related to close encounters. In the absence of close encounters, FMM does not lead to any systemic bias in the energy drift compared to its direct counterpart. In case of `NBODY6` the relative energy error for the softened version is an order of magnitude or two times larger than that of `Taichi`. We speculate that this could be because of the lack of symmetrized time steps or because the code is not optimized to run with softening enabled.

We also run a separate $N = 1024$ particle simulation with $\epsilon_{\rm soft} = \frac{0.01}{N}$. A smaller softening would allow closer encounters: hence, we should expect jumps in energy similar to the unsoftened versions of the codes. This is indeed what we have observed. The impact of different softening is is also reported in Maureira-Fredes and Amaro-Seoane (2018), in a fourth-order Hermite N-body code with Plummer softening. Our findings are consistent with Maureira-Fredes and Amaro-Seoane (2018). Nitadori and Makino (2008) have conducted a comparison between fourth-, sixth- and eighth-order Hermite schemes. The energy error jumps are absent in the eigth-order run, primarily due to shortened step size in the *outer* region. Therefore, we suspect that the cause of energy error jumps is a mismatch between those high-speed particles ejected after close encounters and those particles assigned with large step size (i.e., in the outer region).

This mirrors a well-known issue with SPH using individual time steps (Saitoh and Makino, 2009). Interestingly, Löckmann and Baumgardt (2008) adapted a strategy of detecting fast approaches and close encounters to prevent sudden changes in their Hermite integrator. More studies are needed to verify this assertion.

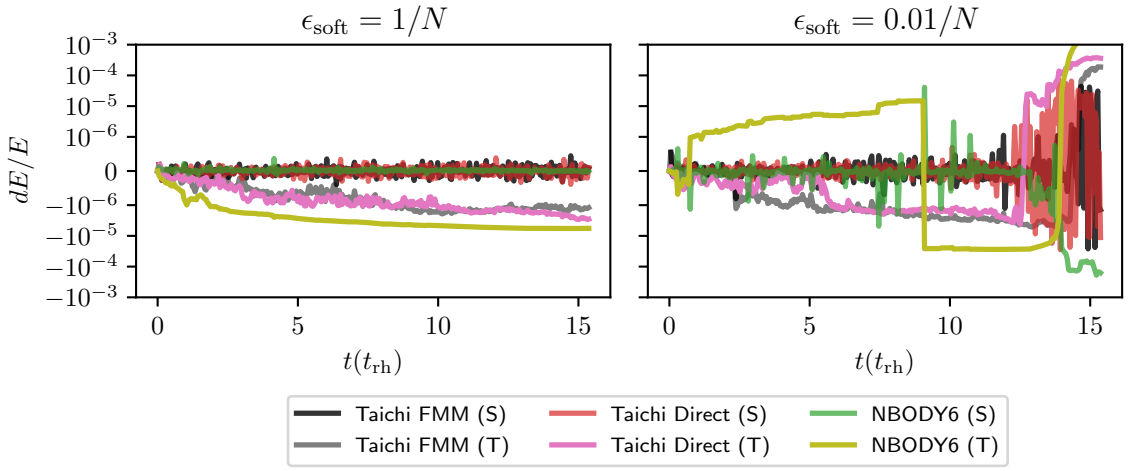

Figure 2.15: The relative energy error is presented as a function of the relaxation time for the evolution of a single $N = 1024$ Plummer model realization. The softening lengths used in this simulation are $1/N$ and $0.01/N$. The energy drift in the former case is smooth, which is what is expected when force softening is used. We find that the energy drift is virtually similar for both `Taichi` direct and FMM and is about an order of magnitude or two better than that of `NBODY6`. We find that for the latter simulation, which uses a lower softening, there are jumps in energy that are caused by close encounters.

2.8.3 Dynamical Friction

One important point that we noted from the dynamical friction test was that there was a lot of stochasticity involved with the position of the black hole particle. This is especially prominent in the case of the less massive black hole. The stochastic nature of the inspiral has been observed and noted in Rodriguez et al. (2018) as well. To present an idea of how the spread of the positions of the massive particle vary with different masses and methods, we present Figure 2.17. We notice that as the mass increases, the spread of the positions becomes smaller. This indicates that the $M_{\rm BH}/M_{\rm star} = 10.0$ case is more sensitive to force discrepancies and round-off errors. Indeed, we noticed that the discrepancy between direct-summation-based N-body codes and codes using approximate solvers is more noticeable for that case. Increasing the mass, however, reduces the difference.

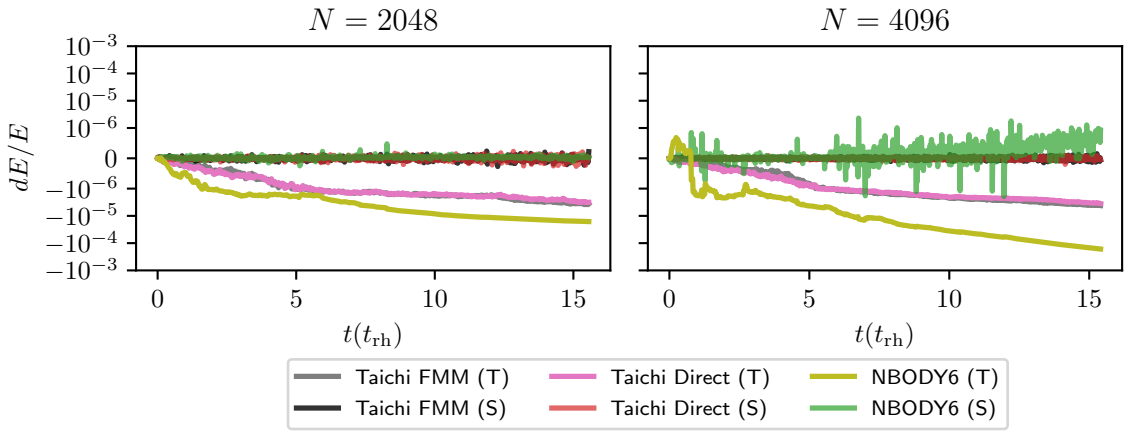

Figure 2.16: The relative energy error is presented as a function of the relaxation time for the evolution of single $N = 2048$ and 4096 Plummer models. The softening length used in this simulation is $1/N$. The energy drift is smooth, which is what is expected when force softening is used. This is similar to what we found in Figure 2.15.

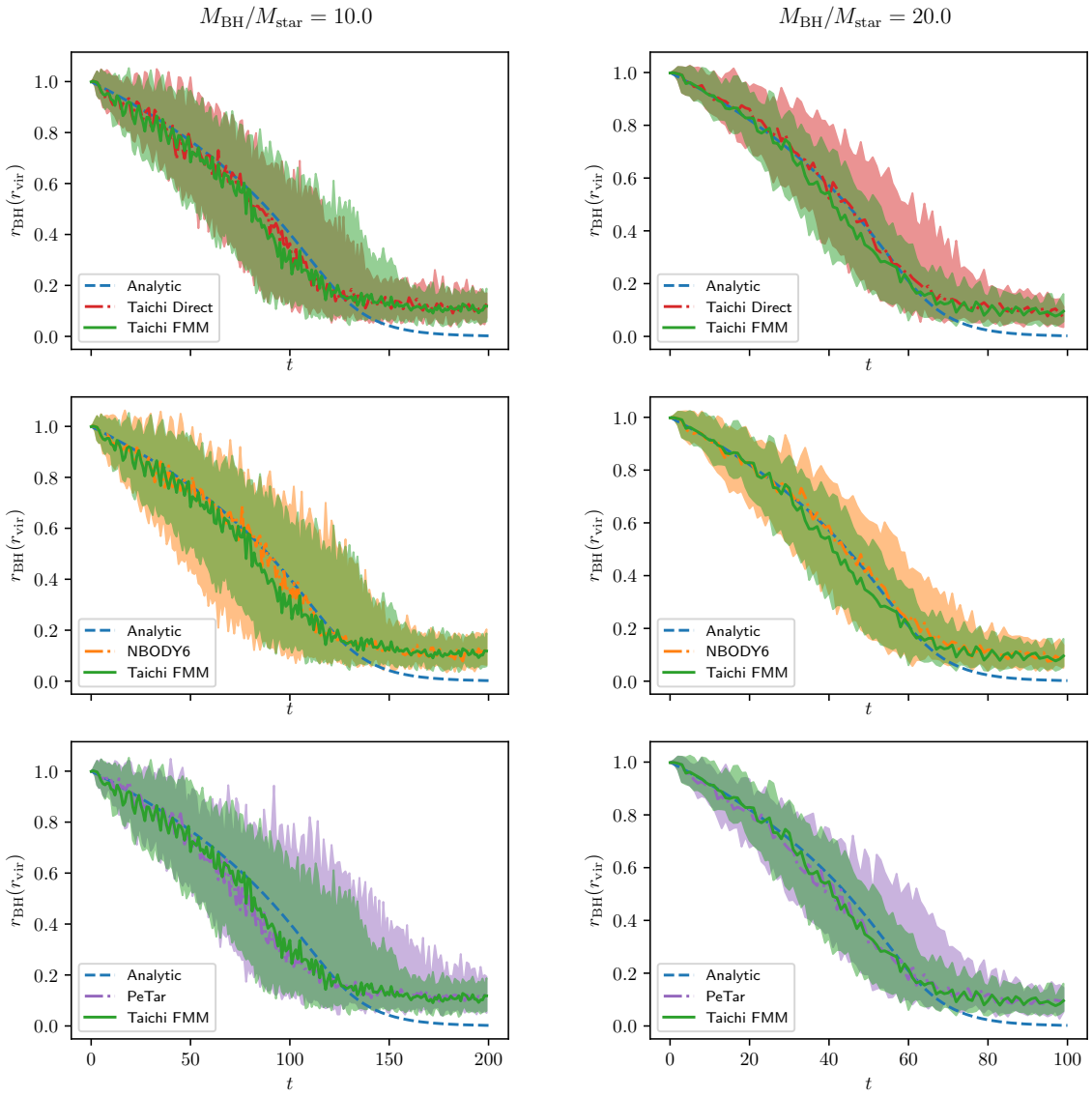

Figure 2.17: The median distance of the massive object is presented as a function of the time (in Henon units) and the virial radius of the cluster. Unlike Figure 2.9, the shaded regions in this figure indicate the spread of radius of the black hole particle from the center of mass of the cluster. Presented here are the 90th percentile values of the distance. All values are binned over one N-body time step. One can see the large spread of radii, indicating the inherent stochasticity present in the simulation.

Chapter 3

Evolution of massive black hole binaries in collisionally relaxed nuclear star clusters

Diptajyoti Mukherjee[1], Qirong Zhu[1], Go Ogiya[2,3,4], Carl L. Rodriguez[1], Hy Trac[1,5]

[1]McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213
[2] Institute for Astronomy, School of Physics, Zhejiang University, Hangzhou 310027, China
[3] Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[4] Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
[5]NSF AI Planning Institute for Physics of the Future, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Abstract

Massive Black Hole (MBH) binaries are considered to be one of the most important sources of Gravitational Waves (GW) that can be detected by GW detectors like LISA. However, there are a lot of uncertainties in the dynamics of MBH binaries in the stages leading up to the GW-emission phase. It has been recently suggested that Nuclear Star Clusters (NSCs) could provide a viable route to overcome the final parsec problem for MBH binaries at the center of galaxies. NSCs are collisional systems where the dynamics would be altered by the presence of a mass spectrum. In this study, we use a suite of high-resolution N-body simulations with over 1 million particles to understand how collisional relaxation under the presence of a mass spectrum of NSC particles affects the dynamics of the MBH binary under the merger of two NSCs. We consider MBH binaries with different mass ratios and additional

non-relaxed models. We find that mass-segregation driven by collisional relaxation can lead to accelerated hardening in lower mass ratio binaries but has the opposite effect in higher mass ratio binaries. Crucially, the relaxed models also demonstrate much lower eccentricities at binary formation and negligible growth during hardening stages leading to longer merger timescales. The results are robust and highlight the importance of collisional relaxation on changing the dynamics of the binary. Our models are state-of-the-art, use zero softening, and high enough particle numbers to model NSCs realistically.

3.1 Introduction

Observations in the past twenty years have demonstrated that Massive Black Holes (MBHs), despite being point sources in the center of galaxies, play a vital role in galaxy evolution and growth (e.g., Kormendy and Gebhardt, 2001; Kormendy and Ho, 2013a). In the hierarchical growth of structures, galaxies frequently merge and form larger systems (e.g., Rodriguez-Gomez et al., 2016). Upon the merger of galaxies, two MBHs get the opportunity to come close to one another to form a bound pair. Recent observation searches have revealed the presence of well-separated accreting MBHs seen as multiple Active Galactic Nuclei (AGN) in a single galaxy, as well as circumstantial evidence for bound Keplerian binaries (e.g., Komossa et al., 2003; Komossa, 2006; Bogdanović, 2015).

MBH binaries are purported to be one of the strongest sources of Gravitational Waves (GW) in the universe. Observations of GW during the final inspiral phase of the binary would reveal information not only regarding the merger history of the galaxy over time but also constrain dynamical properties of galactic nuclei surrounding the MBH binary. Therefore, modeling the dynamical evolution of MBH binaries inside galactic nuclei is central to the astrophysical interpretation of galactic environment and dynamics. MBH binaries are expected to be sources of millihertz (mHz) GW that will be detectable by future space based GW detectors like LISA (Amaro-Seoane et al., 2017) or Tianqin (Luo et al., 2016).

Due to variances in galactic environments in which they are embedded in, the dynamics of MBH coalescence is up for much debate. It is theorized that the merger is a three step process before the the final GW-emission step (Begelman et al., 1980; Merritt, 2013b, chapter 8). In the first step, the dynamical friction of stars and dark matter and that of the interstellar gas play a role in reducing the angular momentum of the black holes which then sink towards the center of the merged galaxy. When the black holes get close enough, they form a bound binary which signals the beginning of the second stage. This stage proceeds rapidly and the separation of the binary decreases due to dynamical friction and three-body scattering events. The third stage prior to GW-driven coalescence begins as the black holes form a hard binary. Once that happens further orbital decay occurs via three-body scattering. If not enough MBH binary-star scattering occurs, the orbital decay of the binary stalls

before it can reach the GW-emission phase. This is called the final parsec problem (e.g., Milosavljević and Merritt, 2003).

The timescale associated with the shrinkage of the binary in the hard binary stage is unclear and heavily depends on the environment. For example, in spherical gas-poor galaxies, simulations have shown that the orbital decay of the MBH binary essentially halts due to the lack of stars in the loss-cone and the GW merger timescales often exceed the Hubble time (e.g., Vasiliev et al., 2015). Additionally, in cosmological simulations MBH seeds have been found to be inefficient at sinking to the center of the nuclei leading to longer merger timescales (e.g., Ma et al., 2021). In case of merging galaxies, however, the non-sphericity of the merger product introduces global torques which can populate the loss-cone more effectively leading to a continued orbital decay of the MBH binary system. Simulations have also shown that merger timescales in such cases are less than the Hubble time (e.g., Berczik et al., 2006; Khan et al., 2013; Vasiliev et al., 2015; Vasiliev, 2017).

One way to overcome the final parsec problem is by embedding the MBHs in Nuclear Star Clusters (NSCs). Nuclear Star Clusters (NSCs) represent some of the densest stellar systems in the universe. They can have mass densities of $\rho \geq 10^6 M_\odot \mathrm{pc}^{-3}$ (e.g., Neumayer et al., 2020b). As the name suggests, NSCs are found in galactic nuclei. The masses and presence of NSCs correlate with the mass of the host galaxy. Sánchez-Janssen et al. (2019) showed that their presence in galaxies depends on the stellar mass of the host galaxies, with a peak of 90% at $M_{\mathrm{stellar}} \sim 10^9 M_\odot$. NSCs and MBHs can coexist in many cases. In fact, the Milky Way galaxy contains an NSC at the Galactic Center that has an MBH embedded in it (e.g., Ghez et al., 2008). Using data from Sánchez-Janssen et al. (2019), Ogiya et al. (2020a) speculate that under the assumption that all NSCs contain an MBH at the center, 50% of all Milky Way sized galaxies should have both an NSC and and MBH present in their nuclei.

In Ogiya et al. (2020a), the authors show that if MBHs are embedded in NSCs prior to merger, tidal effects from the merging NSCs accelerate the orbital evolution timescale before and around the time the binary is formed, thus circumventing the final parsec problem. In the presence of NSCs the formation of a hard binary occurs faster and the whole process of decay into the GW regime is accelerated. The authors found that the mergers were extremely efficient with lower mass ratio binaries merging in ~ 100 Myr while binaries with mass ratio of unity merging in 5 Gyr, which is still less than the Hubble time.

Since NSCs are collisional stellar systems they undergo collisional relaxation even under the presence of an MBH at the center. A collisionally relaxed state implies the presence of a Bahcall-Wolf cusp (Bahcall and Wolf, 1976). If there is a mass spectrum present, the more-massive objects form a steeper cusp than the less-massive objects (Bahcall and Wolf, 1977; Alexander and Hopman, 2009). During the merger of two galaxies containing NSCs, in the absence of MBH binaries, the cusp is expected to be retained (Dehnen, 2005). The presence of MBH binaries leads to a partial or complete disruption of the cusp (e.g., Dehnen, 2005).

In Ogiya et al. (2020a), the authors considered a one-component mass function to study the effect of NSCs on MBH binaries. Realistic NSCs, however, are comprised of a spectrum of masses (e.g., Preto and Amaro-Seoane, 2010; Gualandris and Merritt, 2012a). The effects of a mass spectrum have been explored previously by Gualandris and Merritt (2012a) and Khan et al. (2018) using a fixed binary mass ratio. However, a systematic comparison of unsegregated versus segregated models as a function of the binary mass ratio is missing. A segregated Bahcall-Wolf cusp could lead to enhanced hardening rates and could potentially accelerate the evolution to GW driven coalescence stage. In addition, the effects of the relaxed cusp on the hardening rates of the binary would be an interesting investigation. We, therefore, are motivated to understand how collisionally relaxed NSCs, under the presence of a mass spectrum, affect the dynamics of the binary in the stages leading up to GW driven coalescence.

In this work, we extend the models presented in Ogiya et al. (2020a) to understand the effects of mass-segregation in NSCs on the dynamics of the MBH binary. With the usage of higher mass resolution compared to previous studies and a two-component mass function, our models are able to better represent realistic NSCs. We use the Fast Multipole Method (FMM) (Greengard and Rokhlin, 1987; Cheng et al., 1999) based N-body code `Taichi` (Zhu, 2021a; Mukherjee et al., 2021b) which has been shown to reproduce collisional effects as accurately as direct-summation based N-body codes while using a fraction of the computational time. To understand the effects of mass segregation we use a two-component mass function: one consisting of objects roughly a solar mass or less and the other consisting of heavier objects like stellar mass black holes. We systematically study the effect of relaxed, segregated cusps on the dynamics of MBH binaries with different mass ratios using a suite of N-body simulations and discuss the role of the relaxed cusp in the merger process.

The paper is organized as follows: in section 3.2 we describe the numerical methods and improvements made to `Taichi` to handle dense systems more accurately. In section 3.3, we describe the models of the mergers. In section 3.4 we provide the results and in section 3.5 we discuss the impact of stochasticity and compare our results to those of previous studies. This is followed conclusions in section 3.6.

3.2 Numerical Methods

We perform a suite of N-body simulations using $N \sim 1.32 \times 10^6$ particles to study the formation and evolution of the MBHs embedded in NSCs and their evolution after the formation of the MBH binary. The simulation setup is similar to that presented in Ogiya et al. (2020a) with changes to improve the resolution and the modeling of the clusters to include a two-component mass species. The detailed description of the models is provided in the next section.

To simulate the system, we utilize the FMM based code `Taichi` (Zhu, 2021a; Mukherjee et al., 2021b). Mukherjee et al. (2021b) showed that `Taichi` can simulate systems as accurately as direct-summation based collisional N-body codes while

scaling as $\mathcal{O}(N)$. The accuracy of the force calculation in `Taichi` can be tuned via the usage of an input accuracy parameter (ϵ) which controls the opening angle and a multipole expansion parameter (p) which controls the number of expansion terms used in the force calculation. Using `Taichi` we can simulate large-N systems without the usage of specialized hardware within a physically reasonable amount of time. In this work, we extend `Taichi` to include a fourth-order force-gradient integrator and regularization using the `AR-Chain` scheme (Mikkola and Tanikawa, 1999b). Additionally, we improve the accuracy of the force solver in `Taichi` in this work. We briefly detail the improvements below.

3.2.1 Updated Integration Scheme

Contemporary direct-summation based N-body codes use a fourth order time integration scheme like the Hermite method (e.g., Makino and Aarseth, 1992). In our previous work (Mukherjee et al., 2021b) we adopted an integrator based on hierarchical Hamiltonian splitting which is only second-order accurate (Pelupessy et al., 2012b). In this work, we extend `Taichi` to include a novel fourth-order force-gradient integrator (Rantala et al., 2021b) which decomposes the system into slow and fast subsystems based on the interaction timesteps of the particles. The fast subsystem is then hierarchically split until the slow-fast split results in no particles in the fast subsystem. We refer to this scheme as the `HHS-FSI` scheme hereafter. Hamiltonian splitting integrators are suitable due to the large dynamical range present in our simulations. Unlike conventional composition symplectic integrators presented by Yoshida (Yoshida, 1990), `HHS-FSI` utilizes strictly positive sub-steps made possible by computing an additional gradient term along with the Newtonian accelerations (Chin, 1997; Chin and Chen, 2005) which is only possible because the potential term does not depend on momentum.

Unlike Rantala et al. (2021b) where the gradient term is calculated by direct summation, we utilize an extrapolation method described in Omelyan (2006) which uses a fictitious middle step to approximate the gradient-force term. Tests by Omelyan (2006) have shown this approach is indistinguishable from the Chin & Chen (Chin and Chen, 2005) method. To do this, we follow the method (see also Farr and

Bertschinger, 2007, section 3.1) where

$$\mathbf{p} \leftarrow \mathbf{p} - \frac{h}{6}\mathbf{D}V(\mathbf{q}) \tag{3.1}$$

$$\mathbf{q} \leftarrow \mathbf{q} + \frac{h}{2}\frac{\mathbf{p}}{m} \tag{3.2}$$

$$\mathbf{p} \leftarrow \mathbf{p} - \frac{2h}{3}\mathbf{D}V\left(\mathbf{q} - \frac{h^2}{24m}\mathbf{D}V(\mathbf{q})\right) \tag{3.3}$$

$$\mathbf{q} \leftarrow \mathbf{q} + \frac{h}{2}\frac{\mathbf{p}}{m} \tag{3.4}$$

$$\mathbf{p} \leftarrow \mathbf{p} - \frac{h}{6}\mathbf{D}V(\mathbf{q}) \tag{3.5}$$

where h is the step-size, $\mathbf{q}$ is the position, $\mathbf{p}$ is the momentum and $\mathbf{D}V$ represents the gradient of the potential V evaluated at the position given in the parenthesis. In total, four calls to the Poisson solver are needed in one step.

The forward symplectic nature ensures more accurate and efficient integration compared to that of the Yoshida scheme (e.g., Chin, 2007). The integrator is manifestly momentum conserving and includes an individual symmetrized timestepping scheme similar to that described in Mukherjee et al. (2021b). Despite the loss of symplecticity due to the usage of individual timesteps, the usage of time symmetrization ensures that there is no secular drift in the energy leading to much better energy conservation (e.g., Makino et al., 2006b).

3.2.2 Algorithmic regularization

Even with the inclusion of a fourth order scheme, treatment of close encounters with the MBH binary can prove to be computationally challenging. However, hierarchical Hamiltonian splitting integrators are easy to modify to include regularization as a result of clean separation of fast system from the slow system. This enables accurate handling of close binaries and/or addition of post-Newtonian terms as one can plug-in any accurate few-body solvers to evolve the Hamiltonian of the fast system. Therefore, in order to handle the dynamics of the MBH binary and its interactions with scattering particles more accurately, we include regularization in `Taichi`.

We have utilized the `SpaceHub` API (Wang et al., 2021b) which includes multiple regularization algorithms for accurate few-body integration. `Taichi` can be used along with any of the integration schemes present inside `SpaceHub`. For this work, we found that the `AR-Chain-Sym6+` regularization scheme is the most optimal. `AR-Chain-Sym6+` is an updated `AR-Chain` scheme which is more accurate than Mikkola's (Mikkola and Tanikawa, 1999b) original implementation.

The improvements in `AR-Chain-Sym6+` include an updated chain coordinate transformation, which improves on the CPU time taken to perform the coordinate transformation, active round-off error compensation, and the usage of a sixth order symplectic

integration scheme instead of the traditional GBS extrapolation scheme used in the original `AR-Chain` (Mikkola and Tanikawa, 1999b) method. This method is extremely efficient at handling highly eccentric systems. The usage of a fixed timestep maintains the time symmetry and as such helps achieve higher precision in round-off error dominated regime. For the same relative tolerance parameter, Wang et al. (2021b) found that the `AR-Chain-Sym6+` is at least 1-2 orders of magnitude better at conserving energy. For more information, we refer the interested reader to Wang et al. (2021b). `Taichi` can be configured to allow an arbitrary number of particles to be treated by the regularization scheme. We performed tests and found that treating up to 20 particles with the regularization scheme was optimal in terms of performance and accuracy.

3.2.3 Updates to FMM based force solver

The multipole-to-local (M2L) kernel plays a crucial role in FMM by translating the multipole moments to local expansions for approximated force calculations. In the previous version of `Taichi`, the order of expansions in the M2L kernel is kept at p, for both multipole moments and the derivatives of $1/r$. After various optimizations, it is found that M2L kernel is evidently memory-bound instead of compute-bound. To increase the efficiency of this kernel, we increase the expansion order for $1/r$ derivatives from p to $2p$. This is called the *double height* M2L kernel (Coulaud et al., 2008) as opposed to the *single height* formulation in our previous version. We found that this modifications improves the force error by a factor of $\sim 10\times$ for the same settings compared to the *single height* version. As a result, we relax the force accuracy parameter ϵ by the same factor if double-height M2L is used. For more information, we refer the interested reader to Coulaud et al. (2008).

All of the improvements presented above enhance the accuracy and capability of `Taichi`. We tested `Taichi` with the initial conditions from Ogiya et al. (2020a) and compared the results from `NBODY6++GPU` (Wang et al., 2015) to ensure correspondence between the two codes. The results are briefly presented in the Appendix. We found that `Taichi` was able to simulate the systems as accurately as `NBODY6++GPU`. For the purposes of our simulations, we found that an FMM input relative force accuracy parameter $\epsilon = 2 \times 10^{-5}$ and a multipole expansion parameter $p = 12$ was most optimal. For more information on these parameters, we refer the reader to Mukherjee et al. (2021b).

We ran tests with different values of ϵ and p and found no difference in the final results. For our simulations we used a timestep parameter $\eta_{\rm T} = 0.3$ unless more accuracy was demanded. For most simulations this results in a relative energy conservation of the order of $\sim$0.01%. Under the presence of a dense segregated cusp, we found that the total relative energy conservation was $\sim$ 0.1%-1%. We note that no softening was used in the simulations. All of the simulations presented in this study were performed using only 32 threads on an `AMD Epyc 7742` node. The simulations

took $\sim 14-18$ days for $N \sim 1.32 \times 10^6$ to run to completion. The simulations with highly eccentric binaries or extremely dense cusps took much longer due to the formation of stable multiple systems.

3.3 Models

As mentioned in the previous section our choice of models is motivated by the work described in Ogiya et al. (2020a). We are interested in MBHs whose coalescence will be detectable by LISA and Tianqin. We set the mass of the primary $M_1 = 10^6 M_\odot$. The masses of the secondaries are generated such that we have mass-ratios $q = 1.0, 0.1, 0.01$ where $q \equiv \frac{M_2}{M_1}$ and M_2 is the mass of the secondary. This enables us to systematically study the effect of the secondary on the collisonally relaxed cusp and vice-versa.

To generate the two-component models we assumed that the stars, white dwarfs, and neutron stars can be represented by a population of 1 $M_\odot$ particles, which are termed as the MS particles, while the heavier objects are represented using a population of 10 $M_\odot$ particles, which are called the BH particles. This is similar to the values used in contemporary studies (Preto and Amaro-Seoane, 2010; Gualandris and Merritt, 2012a). To determine the fraction of MS particles to that of BH particles we consult Gualandris and Merritt (2012a) and Antonini (2014). Kroupa IMF (Kroupa, 2001) predicts that

$$N_{\rm MS} : N_{\rm BH} \sim 1 : 0.001 \tag{3.6}$$

However, we used values that are consistent with a top-heavy IMF (e.g., Maness et al., 2007; Bartko et al., 2010). Spectroscopic data from late-type giants in inner parsec of Galactic Center provides evidence of continuous star formation consistent with that of a top-heavy IMF (Maness et al., 2007). In our simulations

$$N_{\rm MS} : N_{\rm BH} = 1 : 0.005, \tag{3.7}$$

similar to the value used in Gualandris and Merritt (2012a). Some IMFs predict an even higher fraction of BH particles (e.g., Chabrier, 2005). The effect of changing the ratio of MS particles to BH particles would require further studies and is beyond the scope of this work.

Each NSC in our simulation weighs $10^7 M_\odot$ in total, comprised of both MS and BH particles. Using the IMF number ratio from equation 3.7, we find that this implies that $M_{\rm BH} = 4.75 \times 10^5 M_\odot$ and $M_{\rm MS} = 9.525 \times 10^6 M_\odot$. In each NSC there are $N_{\rm MS} = 655360$ and $N_{\rm BH} = 3276$ particles. Thus, we model the system using a total of $N = 1317272$ particles. Two additional particles are used to model the MBHs. To generate the N-body representations of the models, we use the self-consistent galaxy modeling toolkit `Agama` (Vasiliev, 2019a). In addition, `Agama` includes a Fokker-Planck code called `Phaseflow` (Vasiliev, 2017) which we utilize to generate collisionally relaxed density profiles near MBHs. We describe the models in more details below. A brief summary of the initial conditions is provided in Tables 3.1 and 3.2.

Parameter	**Value**
$N_{\rm MS}$	655360
$N_{\rm BH}$	3276
N	1317272
$M_{\rm MS}$	$9.525 \times 10^6 M_\odot$
$M_{\rm BH}$	$4.75 \times 10^5 M_\odot$
$M_{\rm part;MS}$	$14.5 M_\odot$
$M_{\rm part;BH}$	$145 M_\odot$

Table 3.1: Summary of the initial parameters used in the generation of the N-body models.

3.3.1 Non-relaxed NSC Models

To generate both the non-relaxed and the relaxed models, we start off with the Dehnen density profile (Dehnen, 1993) for both the MS and BH particles. The density profile for each component i is given as:

$$\rho_i(r) = \rho_0 \left(\frac{r}{r_0}\right)^{-\gamma_i} \left(1 + \frac{r}{r_0}\right)^{\gamma_i - 4} \tag{3.8}$$

where ρ_0 is the normalizing factor, r_0 is the scale radius and γ_i determines the inner slope of the component i. We set $r_0 = 1.4$ pc following Ogiya et al. (2020a). We use $\gamma_{\rm MS} = \gamma_{\rm BH} = 0.5$ which is the lowest value of γ that can support an MBH at the center with an isotropic velocity distribution (Baes et al., 2005). We truncate the density profile at 1000 pc using an exponential truncation function. The distribution functions of both the MS and BH particles are generated by using the density function of each individual component and the combined potential of all components including the MBH, with an isotropic velocity distribution. This is all done under the self consistent framework of `Agama`.

3.3.2 Relaxed NSC Models

To generate the collisionally relaxed models, we have to follow a few more steps. We input the non-relaxed density profiles of both the MS and BH particles along with the mass of the MBH at the center. Then, we evolve the system using the Fokker-Planck code `Phaseflow` until the system has evolved to a collisionally relaxed state. For the model with a $10^6 M_\odot$ MBH at the center, the relaxed state is achieved in ~ 0.5 Gyr. This occurs when the inner density profile of the BH particles falls off as $\sim r^{-2}$ and that of the MS particles falls off as $\sim r^{-1.5}$ (Bahcall and Wolf, 1977; Hopman and Alexander, 2006; Alexander and Hopman, 2009). The output from `Phaseflow` can be easily used to generate isotropic models using `Agama` in a fashion similar to the one described above. We present a comparison of the analytic density and mass profiles in Figure 3.1 for the non-relaxed and the relaxed cases when we have a $10^6 M_\odot$ SMBH at the center. In both the non-relaxed and the relaxed cases, we verified the N-body models accurately reproduced the analytic density and mass profiles.

3.3.3 Generating the Merger Models

To initialize the merger between two NSCs, we follow the steps outlined in Ogiya et al. (2020a). The two NSCs and their corresponding MBHs are initially unbound and are allowed to become bound over the course of the simulation. The initial separation between the two MBHs is denoted as $d_{\rm in}$. In our simulations we set $d_{\rm in} = 20$ pc. We verified that this is less than the effective radius of the NSCs. To generate the initial relative velocity of the two NSCs, we use a free parameter ξ similar to the parameter η described in Ogiya et al. (2020a) equation (9). ξ quantifies the initial

angular momentum of the orbit. Smaller values of ξ imply a more eccentric orbit. In our models, the relative velocity $v_{\rm in}$ is defined as

$$v_{\rm in} = \xi \sqrt{\frac{GM_*(d_{\rm in})}{d_{\rm in}}} \tag{3.9}$$

where $M_*(d_{\rm in})$ accounts for the total mass (excluding the MBH masses) within a distance of $d_{\rm in}$ from the center of each NSC. Once $v_{\rm in}$ has been obtained, the NSC of the secondary along with its MBH is placed at a position centered around $\mathbf{r}_{\rm in} = (d_{\rm in}, 0, 0)$ with a velocity $\mathbf{v}_{\rm in} = (0, v_{\rm in}, 0)$, while the other NSC is place at the origin with zero bulk velocity. In each simulation, we verified that the initial relative velocity was the same to maintain consistency. We found that using $d_{\rm in} = 20$ pc, $v_{\rm in} \approx 55.5$ km/s. Six simulations are generated with the relaxed and the non-relaxed NSCs with $\xi = 1.0$. We label the simulations where relaxed NSCs are used as `r_` simulations and the simulations where non-relaxed NSCs are used as `nr_`. To reduce the number of simulations performed due to computational constraints we perform mergers of relaxed NSCs with only other relaxed NSCs and non-relaxed NSCs with other non-relaxed NSCs. We expect the results of mixed simulations to lie in-between the results obtained in this study.

We expect most mergers to happen on eccentric orbits rather than circular. Thus, it is imperative to understand the effects of initial eccentricity and its evolution under the presence of relaxed and non-relaxed cusps. To understand the effects of an initially eccentric orbit, we perform four additional simulations with $q = 0.1$, $\xi = 0.5$ and $q = 0.1$, $\xi = 0.1$. The former simulations are labelled `ecc_1` whereas the latter simulations are labelled `ecc_2` The circular orbit models are used to understand how the evolution of the binary changes as a function of the mass-ratio while the eccentric models are used to understand how the eccentricity is affected by the different density profiles for a given mass ratio.

To demarcate the three stages of evolution, we need the influence radius and the hard binary radius. In order to determine the influence radius and the hardening radius, we consult Merritt (2013b) equation (8.71). We use the following definitions which are more suitable for N-body simulations; once a bound binary is formed, an influence radius of the primary can be defined as

$$d_{\rm infl} \equiv r_{\rm enc}(2M_1) \tag{3.10}$$

where $r_{\rm enc}$ is the radius enclosing the amount of mass in the parenthesis. The corresponding hard binary radius can be defined then as

$$a_{\rm h} = \frac{q}{(1+q)^2}\frac{d_{\rm infl}}{4}. \tag{3.11}$$

For more definitions, we refer the reader to Merritt and Szell (2006a).

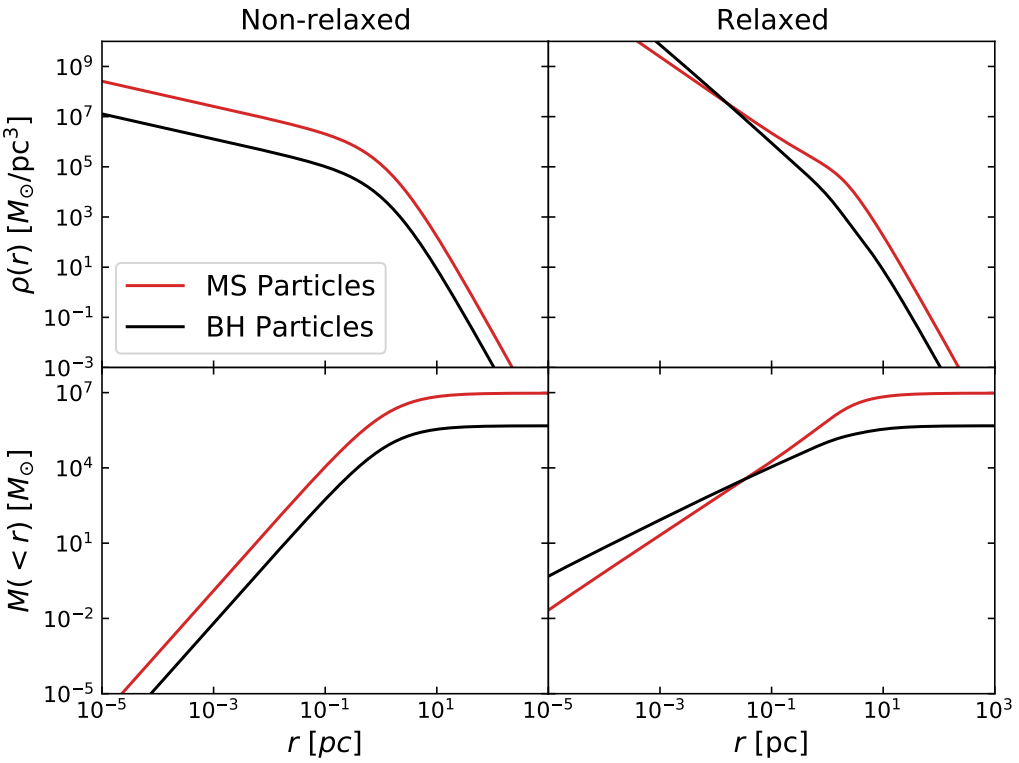

Figure 3.1: The analytic density $\rho(r)$ and the enclosed cumulative mass $M(< r)$ a function of r, the distance from the center of the cluster under the presence of a $10^6 M_\odot$ MBH at the center. The analytic profiles have been computed using `Phaseflow` The differences in the relaxed and the non-relaxed cases are evident with collisional relaxation implying mass segregation. The relaxation produces a denser cusp near the $10^6 M_\odot$ MBH and stellar mass black holes dominate the total mass for all radii < 0.1 pc. The MBH is dominant in regions with $r < 1$ pc.

All the simulations were run for a total time of 10 Myr. In most simulations, this was enough for the MBH binary to harden to $a_{\rm h}/5$ which is usually sufficient to study the effects of core scouring as reported in previous studies (e.g., Merritt and Szell, 2006a).

3.4 Results

To visually examine the evolution of the MBH binary over time, we present snapshots of the evolution over the first 1.5 Myr for the `r_q_0.1` model in Figure 3.2. The orbit of the MBHs is in the $x - y$ plane, and the plot's origin is the center of mass. We

Simulation ID	$\gamma_{\rm MS}$	$\gamma_{\rm BH}$	q	ξ
r_q_1.0	1.5	2.0	1.0	1.0
r_q_0.1	1.5	2.0	0.1	1.0
r_q_0.01	1.5	2.0	0.01	1.0
nr_q_1.0	0.5	0.5	1.0	1.0
nr_q_0.1	0.5	0.5	0.1	1.0
nr_q_0.01	0.5	0.5	0.01	1.0
r_q_0.1_ecc_1	1.5	2.0	0.1	0.5
r_q_0.1_ecc_2	1.5	2.0	0.1	0.1
nr_q_0.1_ecc_1	0.5	0.5	0.1	0.5
nr_q_0.1_ecc_2	0.5	0.5	0.1	0.1

Table 3.2: Summary of the model parameters used for the NSC-NSC merger simulations. All of the above simulations use the same number of particles. The first six models are on circular orbits initially and used to study the effect of relaxation as a function of q and the last four models are eccentric and used to study the effect of initial eccentricity and evolution of eccentricity at a fixed q. ecc_1 models are moderately eccentric whereas ecc_2 models are highly eccentric.

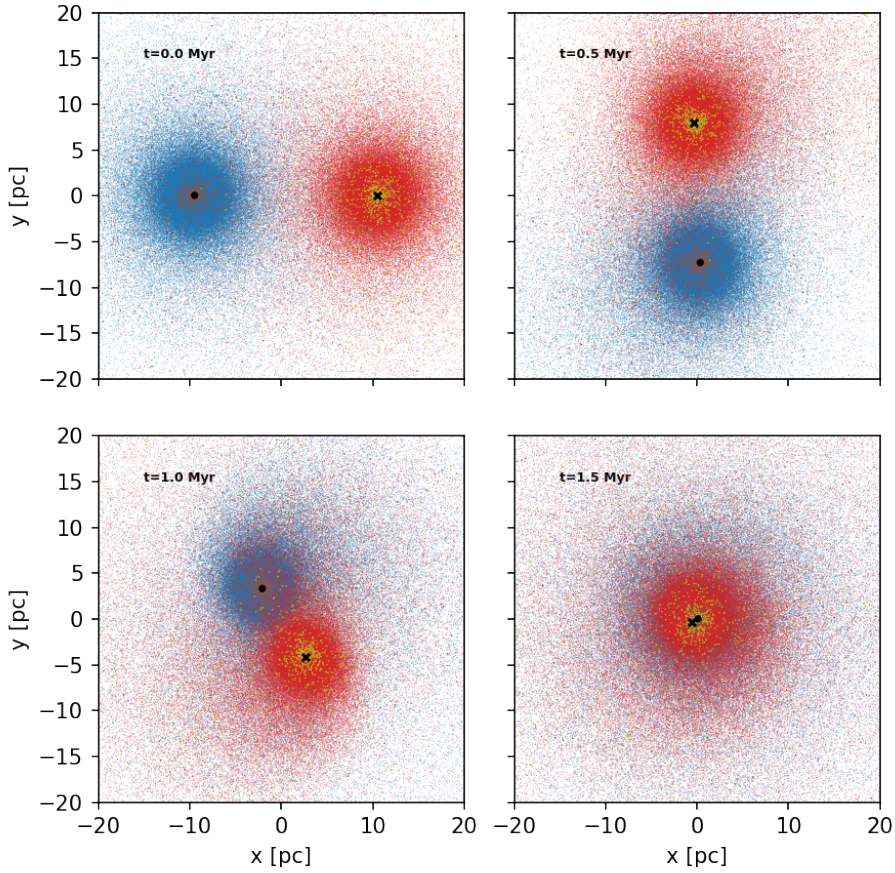

Figure 3.2: A scatter plot of the two NSCs with MBHs projected onto the $x - y$ plane at different points in time during the merger process. The simulation being pictured here is `r_q_0.1`. As the simulation proceeds, the NSCs belonging to the primary (black circle) and secondary (black cross) are brought closer to each other by the combined effects of dynamical friction and tidal forces from stripped stars leading to a mixture of the MS (blue, red dots) and BH particles (brown, yellow dots) from both NSCs. The NSCs merge within ~ 1.5 Myr resulting in the formation of a hard binary at the center.

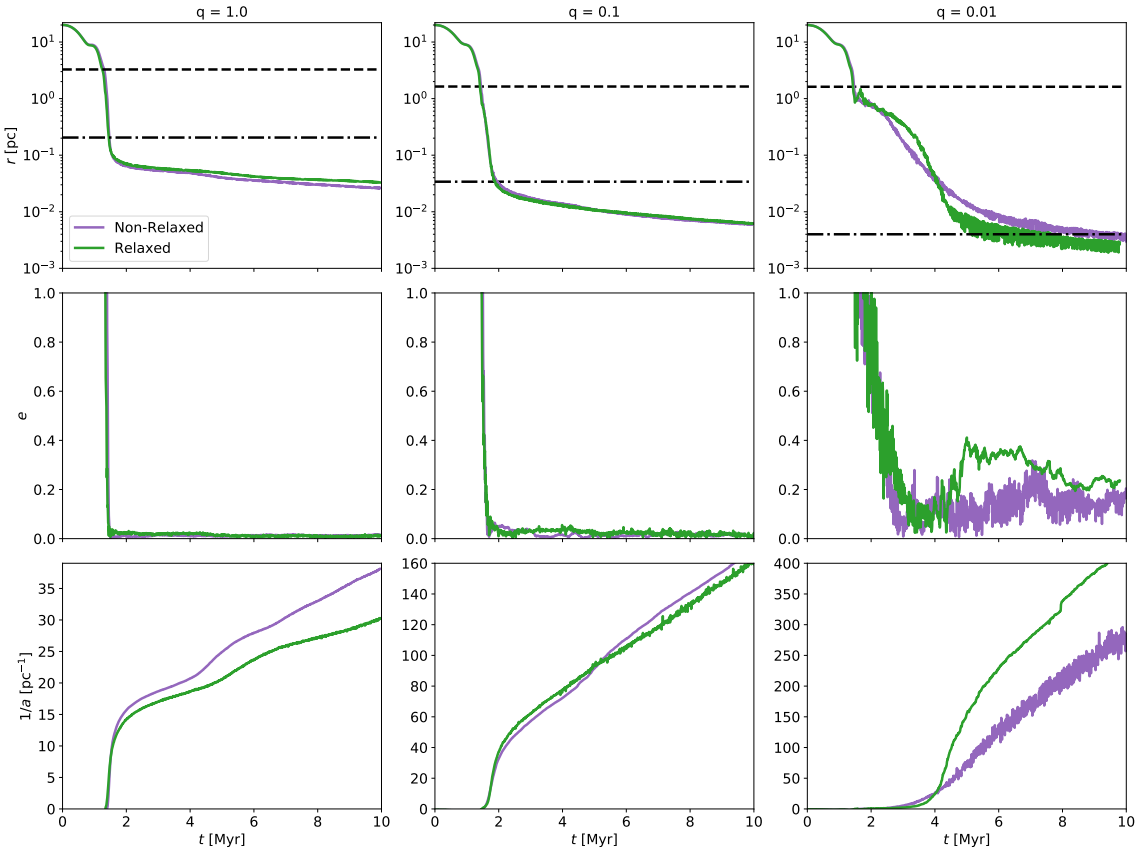

Figure 3.3: The evolution of the binary parameters as a function of time for the circular orbit models. The dashed line represents the influence radius of the binary and the dash-dotted line represents the hard-binary radius. Top: evolution of the separation (r) between the two MBHs as a function of time. Middle: evolution of the eccentricity (e) as a function of time. Bottom: evolution of the inverse semi-major axis ($1/a$) as a function of time. The different evolutionary tracks between the non-relaxed and the relaxed cases highlight the imprint of the surrounding NSC on the dynamics of the MBH binary. We find that while non-relaxed models reach hard binary radius and harden faster for $q = 1.0$, the opposite happens for $q = 0.01$.

find that the NSCs merge within $\sim$ 1.5 Myr leading to the formation of a hard MBH binary. To quantitatively understand the dynamics in more detail, we plot the evolution of the orbital separation, eccentricity, and inverse semi-major axis of the binary over time in Figures 3.3 and 3.6. They helps us understand the differences between relaxed and non-relaxed models as a function of the mass ratio q and the initial eccentricity. We analyze each step of the evolution in the sections below. We first present our analysis of the circular orbit models before moving on to the eccentric ones.

3.4.1 Pre-binary phase ($r > d_{\rm infl}$)

The first stage of evolution, the pre-binary phase, lasts until a bound binary has formed. Examining the orbital elements in Figure 3.3, we find that this phase lasts for the first $\sim$ 1.5 Myr for both relaxed and non-relaxed models across different q. The NSCs merge in roughly $\sim$ 1.5 Myr bringing the secondary to within the influence radius of the primary ($\sim$ 1 pc) leading to the formation of a bound binary.

In this phase the evolution is dominated by the dynamical friction of the stars and the drag force from the tidally stripped stars which help in reducing the angular momentum of the MBHs (Ogiya et al., 2020a). The latter effect is more important in this stage as dynamical friction typically acts on longer timescales. The process involves the transfer of angular momentum from the NSC cores to the stripped stars which expand their orbits. As demonstrated by Huang (1963) and later by Ogiya et al. (2020a), if the two NSCs are considered to be part of a binary system, then in the event of a mass loss, the change in specific angular momentum l can be written as

$$\delta l = (l_s - l)\,\frac{\delta m_s}{m} \tag{3.12}$$

where m is the mass of the NSC binary system and m_s and l_s are the mass and specific angular momentum of the stars that have been tidally stripped from the NSC. δm_s is large in the early phase of the dynamical evolution, while it can be negligible in the later phase. Under the assumption that the eccentricity of the stripped stars has not changed and that the mass loss through tidal disruption is negligible compared to that of the mass of the NSC binary, we can write the expressions for the specific angular momentum of the NSC binary and that of the stripped stars as follows:

$$l = \sqrt{Gma(1-e^2)} \tag{3.13}$$

$$l_s = \sqrt{Gm(a+\delta a)(1-e^2)}. \tag{3.14}$$

In order to satisfy the condition that $\delta m_s < 0$ and $\delta l < 0$, we find that $\delta a > 0$. Thus, as the expansion of orbit is associated with an increase in angular momentum, and the total angular momentum remains conserved, the distances between the NSC cores, and MBHs embedded in them, shrink as a result. This mechanism is especially

important for lower q cases as dynamical friction works inefficiently to decay the orbit of less massive BHs (Ogiya et al., 2020a). The results are consistent with those presented in Ogiya et al. (2020a). This is not surprising since we compared the amount of mass losing and gaining angular momentum between simulations and found that they were equivalent. This leads to negligible differences in the evolution during this period. However, we warn the readers that the effect of tidal stripping would be reduced in realistic galaxies. Interestingly, we find that the time of binary formation, which is dictated by the tidal effects, is consistent amongst simulations performed with $N = 4 \times 10^5$ and $N = 1.32$m (see appendix 3.7.1) suggesting that a lower resolution is sufficient enough to resolve tidal stripping effects and time of binary formation. A systematic study of the effect of resolution on tidal stripping is beyond the scope of this work.

3.4.2 Bound binary phase ($d_{\rm infl} < r < a_{\rm h}$)

In the second phase of evolution, after a bound binary has formed, orbital decay occurs due to a mix of dynamical friction and three-body scattering events. When the two MBHs are sufficiently far apart, dynamical friction acts on each body independently to shrink the binary (Merritt, 2013b, section 8.2.2). When the binary gets closer, hardening via scattering becomes more important and the efficiency of scattering depends on the binary mass ratio q (e.g., Merritt, 2013b, section 8.2.2).

Examining Figure 3.3 we find that for $q = 1.0, 0.1$, the combined phase proceeds extremely quickly leading to the formation of the hard binary immediately after the end of the first phase. The evolution in the $q = 0.01$ models is more gradual. In addition, we find some notable differences in this phase between the relaxed and the non-relaxed models that depend on q. We notice from the evolution of the inverse semi-major axis that in `nr_q_1.0`, the binary is able to settle at a smaller separation than in `r_q_1.0`. This is quite surprising since intuitively we would expect the denser relaxed cusp to yield a faster orbital decay. As the mass ratio is lowered, the situation changes and in the $q = 0.01$ case, we find that the the relaxed model actually accelerates the transition to the hard binary stage. Since the orbital energy of the binary is given as

$$E_{\rm binary} = -\frac{GM_1M_2}{2a} \tag{3.15}$$

where a is the semi-major axis of the binary, we deduce that for for the $q = 1.0$ case, the binary is able to lose more energy in the non-relaxed model compared to that in the relaxed model. As we lower the mass ratio, the energy loss in the relaxed models increases. We seek to understand the processes in work that change the results across mass-ratios.

In order to understand the differences, we first approximately quantify the amount of energy lost by dynamical friction and scattering. We follow Merritt (2013b) equa-

tions (8.73) and (8.74) which state

$$\left.\frac{dE}{dt}\right|_{\rm df} \approx -4.4\frac{G^2 M_2^2 \rho(r) \ln(\Lambda)}{\sigma} \tag{3.16}$$

and

$$\left.\frac{dE}{dt}\right|_{\rm s} = -\frac{H(a)}{2q}\frac{G^2 M_2^2 \rho(r)}{\sigma} \tag{3.17}$$

where $\left.\frac{dE}{dt}\right|_{\rm df}$ and $\left.\frac{dE}{dt}\right|_{\rm s}$ are the energy losses from dynamical friction and scattering respectively, $\ln(\Lambda)$ is the Coulomb logarithm, σ is the velocity dispersion and $H(a)$ is the dimensionless scattering rate. As the binary hardens, the energy losses via scattering become more important as the scattering efficiency increases as q^{-1}. Physically, this makes sense since the larger the mass of the secondary MBH, the more energy would have to be extracted by the intruder to harden the binary by a fixed amount.

From equations 3.16 and 3.17 we notice that $\frac{dE}{dt} \propto M_2^2$. Thus, the rate of loss of energy is faster for binaries with larger secondary mass, which explains why the combined phase proceeds rapidly for $q = 1.0, 0.1$ in contrast to the $q = 0.01$ case. Dynamical friction is less efficient in the $q = 0.01$ models leading to a more gradual orbital decay. However, this does not explain why the orbital decay is more efficient in the `nr_q_1.0` model compared to the `r_q_1.0` model.

To understand the differences in evolution of the inverse semi-major axis between the non-relaxed and relaxed models we note that $\frac{dE}{dt} \propto \rho(r)$. Therefore, we focus on the differences in the initial density profile of the non-relaxed and the relaxed models. For clarity, we compare the density profile for the MS particles in the case where the central MBH mass is $10^6 M_\odot$ in Figure 3.4. Since the total mass of the MS particles is much more than that of BH particles, we expect any major discrepancies to arise out of differences in the MS density and mass profiles.

Looking at all radii < 10 pc, we find that in the relaxed models, the density of MS particles is lower than that of the non-relaxed models for all radii > 0.1 pc. Since the NSC mass across our models is fixed and the relaxed models have higher central density, the density in the outskirts decreases. This counter-intuitive result was also reported in Gualandris and Merritt (2012a). To explain it, the authors accounted for the effect of the BH particles on the MS particles near the SMBH. Since the BH particles are more massive than the MS particles, they dynamically heat the MS particles leading to a lower density in the above-mentioned region. However, Gualandris and Merritt (2012a) showed that at smaller radii, due to the scattering effects of the BH particles, the MS particles end up forming a denser, Bahcall-Wolf cusp. This cusp has a higher density only for very small radii, typically $\sim 0.1 d_{\rm infl}$.

As the MBH binary hardens to $a_{\rm h}$ and energy loss due to scattering becomes more dominant, differences in evolution appear between the non-relaxed and the relaxed models. We find that $a_{\rm h} = 0.205$ pc for $q = 1.0$, the largest amongst all our models. This lies in the region where the density $\rho(r)$ for the relaxed models is lower than that

of the non-relaxed models implying that the energy loss in the non-relaxed models must be higher for `nr_q_1.0`.

The situation changes as we lower the mass-ratio and the hard-binary radius decreases. We find that the total density ρ increases after $r \sim 0.1$ pc. Owing to the higher density, the binary is able to compensate or even more than compensate for the differences in the energy loss in the beginning of the combined stage in the $q = 0.1, 0.01$ models. This is quite evident while comparing the evolution of the binary separation for non-relaxed and relaxed $q = 0.01$ models. The initial inspiral of the secondary is slower in `r_q_0.01` compared to `nr_q_0.01`. Once the secondary is about ~ 0.1 pc away, the inspiral accelerates and it reaches the hard binary radius faster than the non-relaxed model. The situation is quite similar to the `RUN3` model in Khan et al. (2015) where the authors found that initial inspiral of the secondary for the more centrally concentrated model to be slower in the beginning. In the semi-analytic model developed in Gualandris et al. (2022), the authors found that the inspiral due to dynamical friction was slower in the models with steeper inner cusps due to lower stellar density in the outskirts.

Incidentally, for the chosen set of initial conditions, `r_q_0.1` and `nr_q_0.1` show almost identical evolution of binary parameters. For $q \leq 10^{-1}$, higher central concentration drives the binary towards faster inspiral. For really low mass ratio binaries this has the potential to accelerate transition to hard-binary stages even faster but further studies with higher resolution models are required.

As a back-reaction to the shrinkage of the binary due to dynamical friction and scattering, energy is induced into the particles nearby leading to an expansion of their orbits and therefore, a disruption of the cusp. Figure 3.5 provides a visual description of said expansion. Comparing the Lagrange radii, i.e. the radii enclosing a particular fraction of the total mass, plots of both MS and BH particles, we find that in the combined phase, higher mass ratios inject more energy into the surrounding particles leading to a rapid expansion and disruption of the cusp, which also leads to reversal of mass-segregation. The disruption is less-violent in $q = 0.01$ case because of the lower mass of the secondary.

Since the MBHs form a bound binary, we can also examine the evolution of eccentricity during this stage. For the larger mass-ratio models we find that the binary's orbit is approximately circular as it reaches the hard binary radius. We see similar trends in both relaxed and non-relaxed models for $q = 1.0, 0.1$. This is similar to the observation made by Gualandris and Merritt (2012a) in their circular orbit models. For $q = 0.01$, the story is a little different. While the binaries initially start circular, the growth of eccentricity is more stochastic in this case because of the lower mass of the secondary. However, we find that evolution of eccentricity is generally in the direction of higher eccentricity. This was also noted in Merritt (2013b) for Intermediate Mass Black Holes (IMBHs) in this mass range. Curiously, we find that for `r_q_0.01`, the binary is able to reach a higher eccentricity in this phase compared to `nr_q_0.01`. Intuitively, we would expect the opposite because because when the

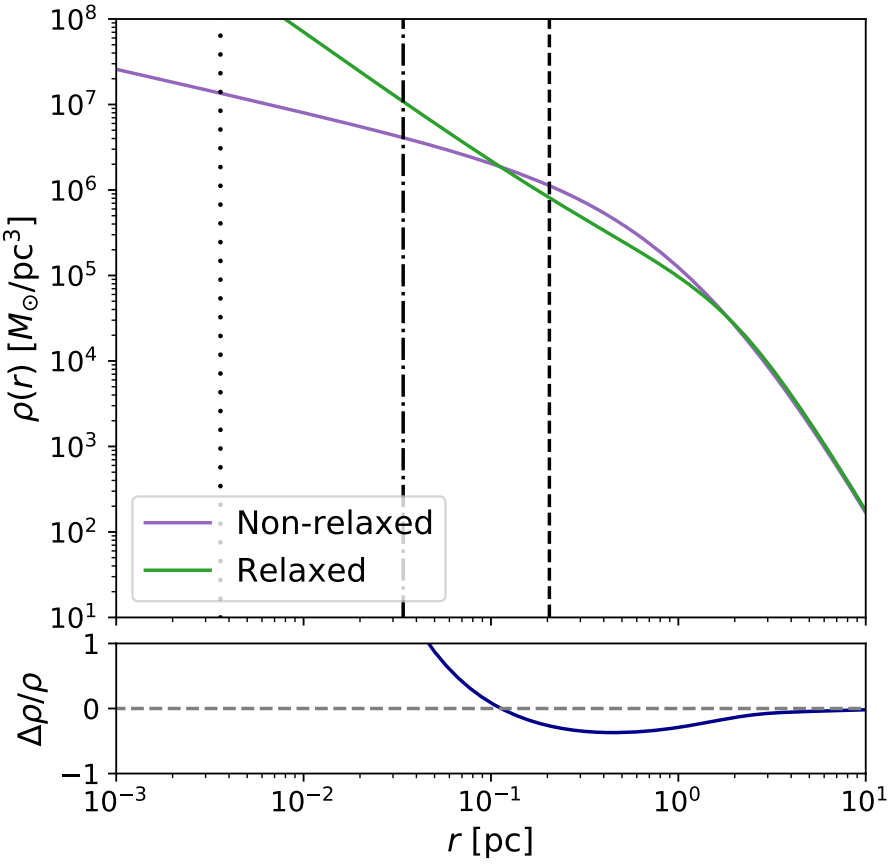

Figure 3.4: The initial density $\rho(r)$ and relative density difference $\Delta\rho/\rho$ of MS particles as a function of r, the distance from the center of the cluster under the presence of a $10^6 M_\odot$ MBH at the center. The dashed line, dash-dotted line, and the dotted line represent the hard-binary radii for the $q = 1.0, 0.1, 0.01$ models respectively. We find that for all radii within the influence radius of the primary, the density of the MS particles is lower in the relaxed models compared to the non-relaxed models until we reach $\sim 0.1 d_{\rm infl}$

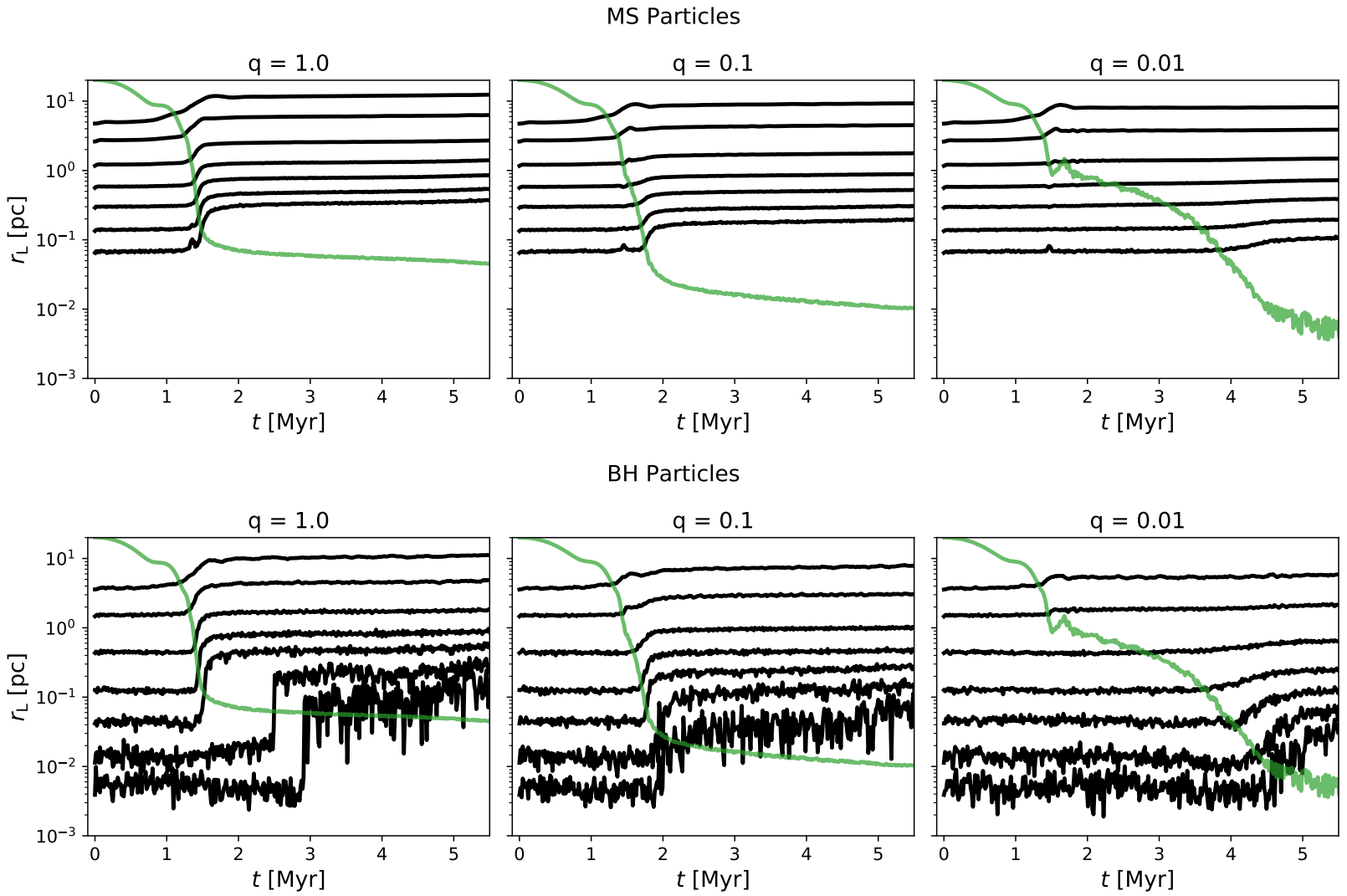

Figure 3.5: The evolution of the Lagrange radius ($r_{\rm L}$) of different mass-fractions as a function of time (t) for the relaxed models. The green curves denote the separation of the MBHs as a function of time. For clarity, only particles belonging to NSC 1 have been taken into account here. From bottom to top the mass fractions are 0.1%, 0.3%, 1%, 3%, 10%, 30%, and 50%. Top: The evolution of the Lagrange radius for the MS particles. Bottom: The evolution of the Lagrange radius for the BH particles. We notice that as the mass ratio decreases, the cusp is perturbed less, and mass-segregation is only partially reversed. The results are qualitatively consistent with similar ones presented in Gualandris and Merritt (2012a).

periapsis of the binary falls within the denser, relaxed cusp, it should circularize the binary. The increase in eccentricity happens around the time the secondary reaches the distances where the mass is dominated by BH rather than MS particles. The secondary hardens by having more encounters with the BH particles. In asymmetric MBH binaries the eccentricity growth is mostly driven by the companion-perturber mass ratio Sesana et al. (2008). A larger intruder mass in the relaxed model results in growth of eccentricity. This may not be physical because in realistic NSCs, the mass of the perturbing BH particles would be smaller.

We also noted that the evolution of eccentricity for $q = 0.01$ model is quite dependent on resolution. We found that a simulation using $N \sim 4 \times 10^5$ particles (see appendix 3.7.1) resulted in the MBH binary reaching extremely large values of eccentricity (0.9) in both non-relaxed and relaxed models. It highlights the importance of using larger resolution to estimate binary evolution parameters, especially for lower mass ratio binaries. Our results for this mass range are quite similar to those presented in Arca-Sedda and Gualandris (2018) where the authors found that unless the IMBH starts off in an highly eccentric orbit, it is not able to reach large values of eccentricity. We caution against consulting single simulations to track the evolution of eccentricity with the current resolution. Unlike semi-major axis, the evolution of eccentricity is a second order effect in angular momentum and subject to more stochasticity and more simulations are needed to model the evolution more realistically.

3.4.3 Hard-binary phase ($r < a_{\rm h}$)

The last phase before the GW emission state is the hard-binary phase where the binary hardens by three-body scattering. We investigate the differences in hardening rates between the relaxed and non-relaxed models. In the full loss-cone regime the binary should harden at a fixed rate. This would imply that,

$$\frac{d}{dt}\left(\frac{1}{a}\right) = s \tag{3.18}$$

where s is some constant. To find the value of s, we fit straight lines to the inverse semi-major axis plots. For $q = 1.0, 0.1$ we fit the lines to the values between 5 Myr and 10 Myr. For $q = 0.01$, we do it between 7 Myr and 10 Myr since the binary reaches the hard binary radius a little before 7 Myr. We notice that there is a sharp jump in the evolution of the inverse semi-major axis in the `r_q_0.01` model around ~ 8 Myr. Jumps in the evolution of the inverse semi-major axis was also noted in Khan et al. (2018). The authors attributed them to the ejection of an intruder following the formation of a triple system between the MBH binary and a third heavy particle. In our instance, this third particle happened to be a BH particle. The jumps do not dominate the overall evolution. To ensure consistency, the average value of s in the non-jump regions is taken to be the hardening rate in this particular scenario.

Simulation ID	s [pc^{-1}Myr^{-1}]
`r_q_1.0`	1.90
`r_q_0.1`	13.9
`r_q_0.01`	46.6
`nr_q_1.0`	2.6
`nr_q_0.1`	14.6
`nr_q_0.01`	39.4
`r_q_0.1_ecc_1`	8.2
`nr_q_0.1_ecc_1`	9.2
`r_q_0.1_ecc_2`	16.2
`nr_q_0.1_ecc_2`	15.7

Table 3.3: Summary of the slopes of the inverse semi-major axes of the various simulations. We find that hardening rates between relaxed and non-relaxed models are within ∼30% of each other contrary to the findings of Khan et al. (2018). In addition, we find that `nr_` models harden faster for $q \geq 0.1$ whereas the opposite is observed for $q = 0.01$

The values of s found for the different simulations are reported in Table 3.3. For the $q = 1.0$ models, the hardening rate in non-relaxed scenario is $\sim 35\%$ higher than that in the relaxed scenario whereas in the $q = 0.1$ models the hardening rates of both relaxed and non-relaxed models are within $\sim 10\%$ of one another. In the $q = 0.01$ models, the hardening rate of the non-relaxed model is 22% lower than that in the relaxed model. We notice that the value of s does not change much between relaxed and non-relaxed models indicating that even though the cusp might be disrupted less in lower mass cases, the primary mode of evolution in this phase is collisionless and dependant on the geometry of the merger product. The lower hardening rate in `r_q_1.0` is quite interesting. To understand it, we looked at the density profile during the hardening phase and found that the density profile in `r_q_1.0` was ∼20% lower than that in `nr_q_1.0` at the influence radius ($\sim$ 1 pc) of the binary. Since this is where the stars in the loss cone arise out of, the lower scattering rate is caused to to fewer particles in the loss cone in the `r_q_1.0` model.

We also examine the growth of eccentricity during this phase of evolution. $q = 1.0, 0.1$ models remain roughly circular whereas the $q = 0.01$ models show stochasticity in the evolution of eccentricity. We find that `nr_q_0.01` shows some mild growth resulting in a final eccentricity of 0.2 by 10 Myr whereas the eccentricity actually decreases in `r_q_0.01`. This is quite curious and we leave the investigation to a future study.

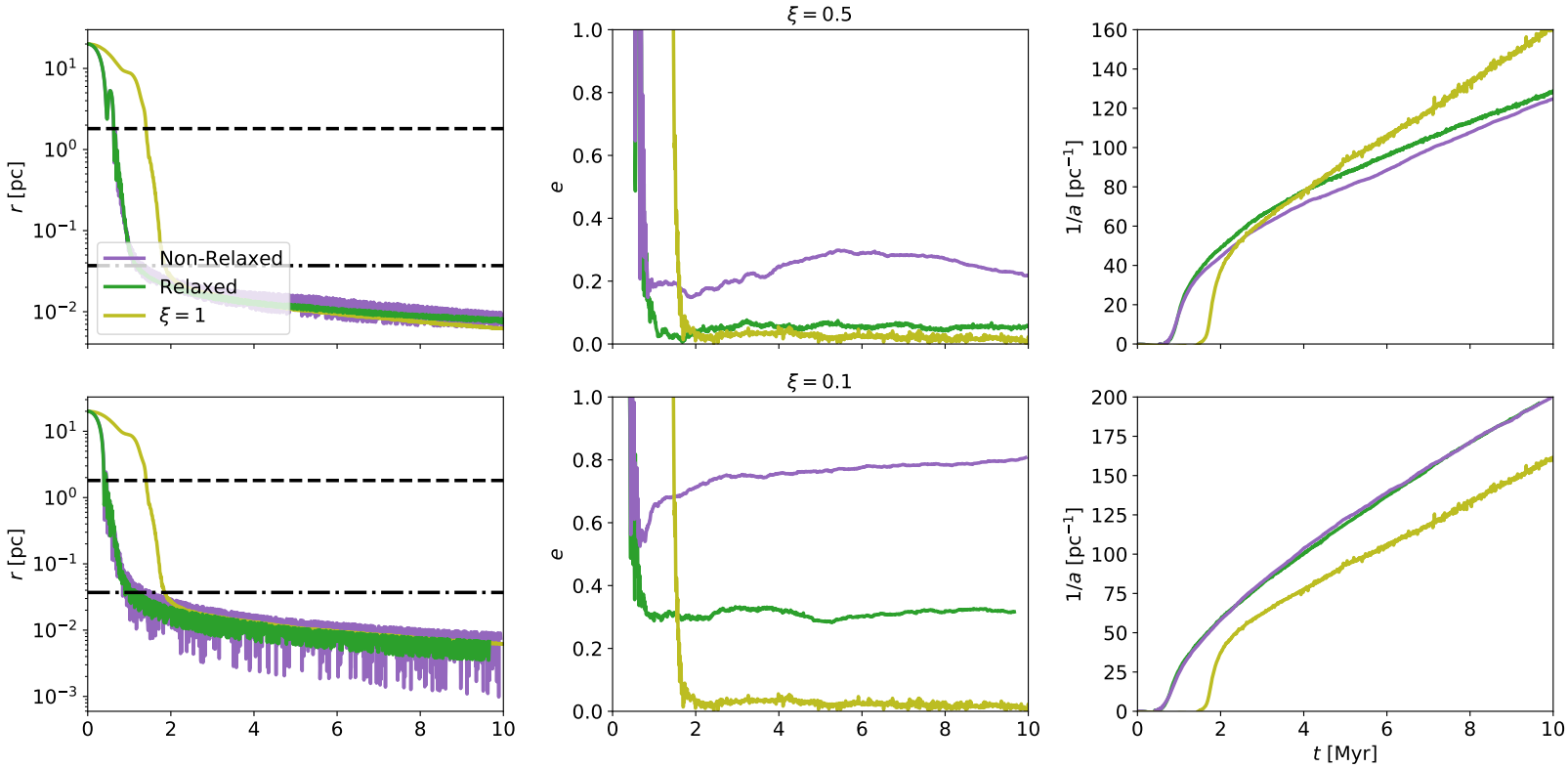

Figure 3.6: The evolution of the binary orbital parameters for the eccentric models as a function of time t in Myr. Top: `ecc_1` models that are mildly eccentric initially. Bottom: `ecc_2` models that are highly eccentric initally. We find similar trends as we found for the circular orbit models of $q = 0.1$ in Figure 3.3 for the binary hardening rates. We, however, notice that the presence of the denser relaxed cusp affects the eccentricity evolution of the binary. The relaxed cusp circularizes the binary more than the non-relaxed cusp. This is more evident in the highly eccentric scenario (bottom) where the binary in the relaxed cusp forms at much lower eccentricity and does not show any growth over time.

3.4.4 Eccentric Orbital Parameters

Realistic galaxy mergers are more likely to happen on eccentric orbits. The aforementioned parameter ξ can be changed to change the initial amount of angular momentum. Reducing the parameter can lead to more eccentric orbits. We studied the effect of eccentric initial conditions and how they interplay with relaxed and non-relaxed models. Figure 3.6 shows that the evolution of the separation and the binary hardening is qualitatively identical to the initial conditions of $\xi = 1$ and $q = 0.1$. The hardening rates between the relaxed and the non-relaxed models are within 10% of each other and is not affected by the initial eccentricity of the orbit. This indicates that the hardening rate is dependent on morphology of the merger product rather than the density of the central cusp. For the mildly eccentric orbits (`ecc_1`), we observe that the hardening rates are lower than that in the circular case whereas in the highly eccentric orbit models (`ecc_2`) demonstrate a larger hardening rate. This is due to the differences in the structure of the final merger product.

The presence of a relaxed cusp affects the evolution of eccentricity. The relaxed cusp leads to a lower eccentricity at binary formation. For `r_q_0.1_ecc_1` the eccentricity at binary formation is almost 0.0 even though the initial orbit was eccentric. Even for highly eccentric initial orbits like in the `r_q_0.1_ecc_2` scenario, we find that the eccentricity at formation is about 0.3. The situation for the non-relaxed models is quite different as `nr_q_0.1_ecc_1` and `nr_q_0.1_ecc_2` models demonstrate eccentricities of 0.19 and 0.55 at binary formation respectively. A This is because, in the presence of the relaxed cusp, the dynamical friction is able to circularize the binary leading to lower eccentricity as also explained in the previous section. A higher initial eccentricity results in a higher eccentricity at binary formation in our models. Gualandris et al. (2022) also report similar findings where the eccentricity at binary formation for the models with steeper cusps is systematically lower than that in the models with shallow cusps.

In the hard binary phase, the non-relaxed models demonstrate a slight growth in eccentricity whereas in the relaxed models, the eccentricity remains roughly constant. This was also observed in Gualandris and Merritt (2012a) where the authors found that the eccentricity evolution was roughly constant with time in the model with initial eccentricity. `nr_q_0.1_ecc_2` is able to reach a high eccentricity of 0.8 by 10 Myr whereas its relaxed counterpart only reaches 0.3. The eccentricity evolution in this stage is consistent with the findings of Sesana (2010) where the author report that shallower cusps with $q \approx 0.1$ demonstrate growth in eccentricity. The evolution of eccentricity in this phase has important consequences on determination of MBH merger timescales. Binaries that are able to reach high eccentricities can merge order of magnitudes faster than those on circular orbits. We plan on systematically studying the evolution of eccentricity as a function of the binary mass ratio in relaxed and non-relaxed models in a future study.

3.4.5 GW Emission from SMBH Binaries

We follow the evolution of the MBH binaries into the GW coalescence phase semi-analytically after the simulations have been stoppedat $t = 10$ Myr. To do so, we assume that the evolution of the inverse semi major axis is constant, i.e. that the hardening due to stellar scattering takes place in the full loss cone limit. This has been the strategy adopted in previous studies (e.g., Gualandris and Merritt, 2012a; Ogiya et al., 2020a). Under this assumption,

$$\frac{d}{dt}\left(\frac{1}{a}\right)_* = s \rightarrow \left(\frac{da}{dt}\right)_* = -a^2(t)s \tag{3.19}$$

where s is a constant and can be figured out from the inverse semi-major axis data by fitting a straight line through it and measuring its slope. The overall evolution of a can be written as follows

$$\frac{da}{dt} = \left.\frac{da}{dt}\right|_{GW} + \left.\frac{da}{dt}\right|_* \tag{3.20}$$

Peters (1964a) provides the rate of change of the orbital elements due to the emission of GW. The rate of change of semi-major axis and eccentricity are given as a set of coupled differential equations

$$\frac{da}{dt} = -\frac{64}{5}\beta\frac{F(e)}{a^3} \tag{3.21}$$

$$\frac{de}{dt} = -\frac{304}{15}\beta\frac{eG(e)}{a^4} \tag{3.22}$$

where

$$\beta = \frac{G^3}{c^5}\left(M_1 M_2 \left(M_1 + M_2\right)\right) \tag{3.23}$$

$$G(e) = (1-e^2)^{-5/2}\left(1 + \frac{121}{304}e^2\right) \tag{3.24}$$

$$F(e) = (1-e^2)^{-7/2}\left(1 + \frac{73}{24}e^2 + \frac{37}{96}e^4\right). \tag{3.25}$$

We numerically solve the coupled differential equations for both a and e using the orbital elements obtained at end of the N-body integration. We do not take into account the growth of eccentricity due to stellar scattering while solving the differential equations.

From Figure 3.7, we find that in all cases with $q \geq 0.1$, the MBH binary in the non-relaxed models merge faster. We find that the `nr_q_1.0` model undergoes coalescence almost 23% faster compared to the `r_q_1.0` model. This is due to the fact that the binary separation itself is lower in the non-relaxed case along with the fact that the scattering rate is also larger because of the reason explained in the previous section. In the $q = 0.01$ case, the binary in the relaxed model is at a smaller separation and

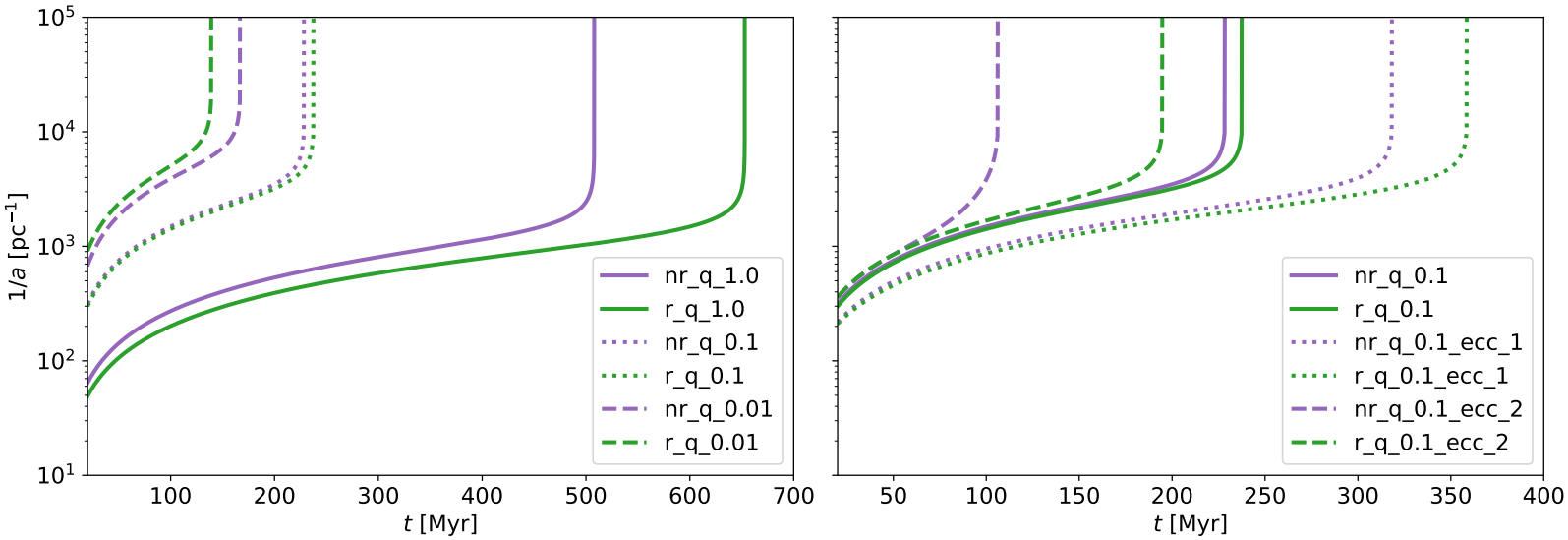

Figure 3.7: The evolution of the inverse-semi major axis as a function of time in the GW dominated phase for different mass-ratios. Left: Evolution and coalescence in models with circular orbits. Right: same but with eccentric models. The evolution is carried by taking the results from the simulations (solid lines) and evolving them semi-analytically (dashed lines) using the Peters (Peters, 1964a) equation. In the circular models (left), we find that the merger timescales for the non-relaxed models are smaller than their relaxed counterparts for $q = 1.0, 0.1$ and opposite for $q = 0.01$. In the eccentric models (right), we find that the binaries in the non-relaxed cusps always merge faster. All models merge within a Hubble time making NSCs a promising source of GWs.

hardens faster resulting in almost 15% faster coalescence compared to its non-relaxed counterpart. Since all of our models are roughly circular by the end of the integration time, we expect the timescales presented in 3.7 to be the upper-limit on the GW merger timescales.

Interestingly, the coalescence timescale for `r_q_1.0` and `nr_q_1.0` simulations is comparable to the relaxation timescale of those systems. It is reasonable to expect the erosion of the non-spherical nature of the system over such timescales in addition to the occurrence of mass-segregation which can affect the hardening rate at later stages for equal mass ratio binaries. However, this is beyond the scope of this study but presents a novel area that merits further investigation in future studies.

We know that the coalescence timescale of the binary depends sensitively on the eccentricity. As such, we expect the binary in the non-relaxed eccentric models to merge faster. Figure 3.7 shows that this is indeed the case. The MBH binary in `nr_q_0.1_ecc_1` merges by 320 Myr whereas its relaxed counterpart takes around 360 Myr. The effect is stark when considering binaries on highly eccentric orbits. `nr_q_0.1_ecc_2` merges within 95 Myr but `r_q_0.1_ecc_2` takes 2× longer. Binaries that form in very eccentric orbits thus more efficiently merge in galaxies where the central densities are lower. Our results are in line with those found by Gualandris et al. (2022) where the authors found that models with shallower slopes are more efficient at merging binaries. Our results also underscore the importance of including a mass-spectrum in N-body simulations as it can affect collisional relaxation in the NSC which in turn can affect the coalescence timescales.

3.4.6 Core Scouring

As the binary hardens, it displaces particles from the cusp. The effects of the binary can be strong enough to disrupt the cusp entirely and create a flat core. To understand the effects of the hardening of the binary on the particles, we plot the density profile of both the MS and BH particles in the merged system at different points of hardening. Due to computational limitations, we study the effects up until the time the binary hardens to a semi-major axis $a_{\rm h}/5$ for `r_q_1.0` and `r_q_0.1` and upto $\sim a_{\rm h}/2$ for `r_q_0.01`. The latter model is extremely computationally intensive to evolve longer because of the formation of stable multiple systems in the cusp. Improving the integration scheme to handle secular systems more efficiently (Rantala et al., 2022) can alleviate this issue.

In the $q = 1.0$ scenario, we find that the inner cusp is completely disrupted and a large flat core is produced as the binary hardens from $a_{\rm h}$ to $a_{\rm h}/5$. This is seen in the density profiles of both the MS and BH particles. The effective density of the core is $\sim 10^5 M_\odot {\rm pc}^{-3}$ signalling that the effects of the binary were so strong that the NSC itself was partially disrupted.

The situation is different for lower mass ratio binaries. For $q = 0.1$, the cusp of MS particles is partially disrupted as the binary hardens from $a_{\rm h}$ to $a_{\rm h}/3$ but further

disruption is not seen with more hardening. The original cusp is not retained. In fact, for the MS particles, the density profile of the merged system has a faint $\gamma_{\rm MS} = 0.5$ inner slope. For the BH particles, we find a faint $\gamma_{\rm BH} = 0.7$ slope. For $q = 0.01$, the effects are even more minuscule. However, since the binary could not be evolved to $a_{\rm h}/5$ due to computational limitations, we exclude it from this analysis.

The partial retention of the cusp has implications on the regrowth of the Bahcall-Wolf cusp post MBH binary coalescence. Whereas the time required to achieve the collisionally relaxed state for $q = 1.0$ may exceed the Hubble time because of the presence of a flat core in both the MS and the BH particles (Merritt, 2010), the same cannot be said for $q = 0.1, 0.01$ which will have faster regrowth. Crucially, we would need to understand how regrowth inter-plays with the galaxy geometry post merger. As such, N-body simulations are required to quantify the exact amount of time required for the regrowth of cusps.

Since Extreme Mass Ratio Inspiral (EMRI) rates are usually extrapolated under the assumption of a Bahcall-Wolf cusp at the center, we would expect EMRIs to arise out of galaxies that have undergone mergers with lower mass ratios. However, multiple mergers even with lower mass ratios can lead to the formation of a core and the exact number of mergers leading to the formation of a core as a function of q requires further studies. The time dependent rate of EMRIs post merger would be interesting to understand as well and we plan on exploring this in future studies.

How does the disruption of the cusp affect the velocity distribution? To analyze that, we plot the velocity anisotropy parameter for the `r_q_1.0` model in Figure 3.9 during different stages of hardening. The velocity anisotropy parameter is defined as

$$\beta = 1 - \frac{\sigma_{\rm t}^2}{2\sigma_{\rm r}^2} \tag{3.26}$$

where $\sigma_{\rm r}$ is the radial velocity dispersion and $\sigma_{\rm t}$ is the tangential velocity dispersion. We find that as the binary hardens, the velocity profile, which was initially isotropic, becomes tangentially biased . This is caused because the MBH binary preferentially ejects particles on radial orbits. The anisotropy parameter can act as an observational evidence for the presence of an MBH binary because of this reason. This was also noted in previous studies like Merritt and Szell (2006a).

3.5 Discussion

3.5.1 Impact of stochasticity

As there is inherent stochasticity involved while generating N-body samples, one may be curious as to whether the results mentioned in the previous sections are robust and reproducible. To understand the effect of stochasticity on the results, we generate four additional statistically independent merger models for the `r_q_1.0` and `nr_q_1.0` simulations. To save computational resources, we only perform the simulations for

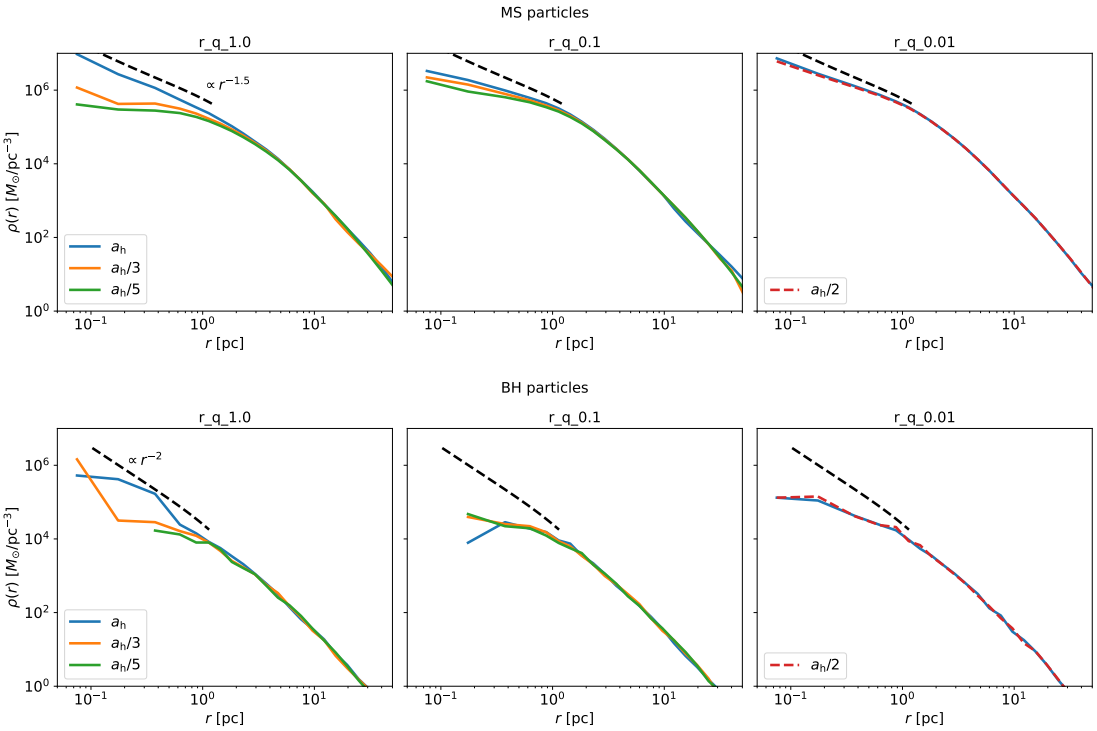

Figure 3.8: The density of particles ρ presented as a function of the distance r from the center of mass of the binary at different points in hardening for `r_` simulations. The initial cusp is also presented for comparison. Top: Density of MS particles. We can see that for $q = 1.0$ as the binary hardens, a core is formed. This is not observed for $q = 0.1$. Bottom: Density of BH particles. Similar observations are noted in this case.

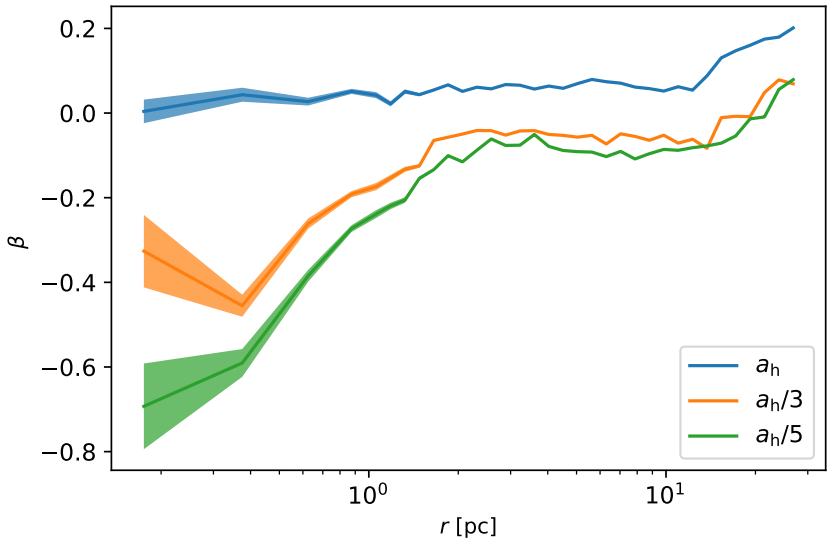

Figure 3.9: The evolution of the velocity anisotropy for `r_q_1.0` model. The anisotropy parameter (β) is plotted as a function of the distance from the binary center(r). The shaded portions denote the standard error in calculating the anisotropy parameter. The models start initially with an isotropic distribution of velocity. As the binary hardens, it preferentially ejects particles with radial velocities producing a tangentially biased velocity structure near the MBH binary.

the $q = 1.0$ models up to a termination time of 5 Myr with a resolution of $N \sim 4 \times 10^5$. This should be sufficient to study any discrepancies in the three phases of evolution despite the lower resolution. The impact of resolution for higher mass ratio binaries is minimal and is also further discussed in the next section and appendix 3.7.1.

In Figure 3.10, we plot the mean and standard deviation of the evolution of the inverse semi-major axis from the five simulations. We find that the results from the original simulations are reproducible and the stochastic scatter for the evolution of the separation and semi-major axis is low. This indicates that the differences in the results of the non-relaxed and relaxed models arise out of physical and not numerical reasons.

We caution the reader that such agreement might not be present for other orbital elements like eccentricity. Previous studies like Nasim et al. (2020) have highlighted this issue and found that it stems from insufficient numerical resolution. The stochasticity of the stellar encounters with the MBH binary affect eccentricty more leading to larger scatter among random realizations. This can potentially affect GW merger timescales since they are sensitive to the eccentricity of the binary. However, the resolution used in this work is sufficient to study the evolution of the semi-major axis and separation.

3.5.2 Comparison with previous studies

Gualandris and Merritt (2012a) and Khan et al. (2018) used a non-uniform mass function to model the galaxy mergers. The resolution used in our simulations is lower compared to that in Gualandris and Merritt (2012a) but is comparable to that used in Khan et al. (2018). Although our set up is quite different compared to Gualandris and Merritt (2012a) we find qualitatively similar results. Comparing Figure 7 from Gualandris and Merritt (2012a) to Figure 3.5 in our study, we observe that for the expansion of Lagrangian radius is quite sudden for mass ratios between $0.1 - 1.0$. Due to the energy injected into the cusp by the MBH binary there is an expansion in the orbits of stars leading to the destruction of the cusp as explained in the previous sections. Our work also suggests consistency with the evolution of the density profile. Although not directly comparable, we find that the effect of binary leading to core scouring presented for $q = 1/3$ in Gualandris and Merritt (2012a) lies between results from our $q = 0.1, 1.0$ simulations.

Our results are in contrast with those presented in Khan et al. (2018) where the authors found that the hardening rates in mass segregated cases were significantly higher than those in the non-mass segregated cases, which was not found in our case. The discrepancy could result from the usage of a different initial mass function in Khan et al. (2018) or the usage of a different mass ratio of the binary. The authors admit that the effects of relaxation in their work could be enhanced because of the lower resolution used while the galaxy itself is relaxing prior to the merger. This effect would be negligible in our case since the collisionally relaxed models have been

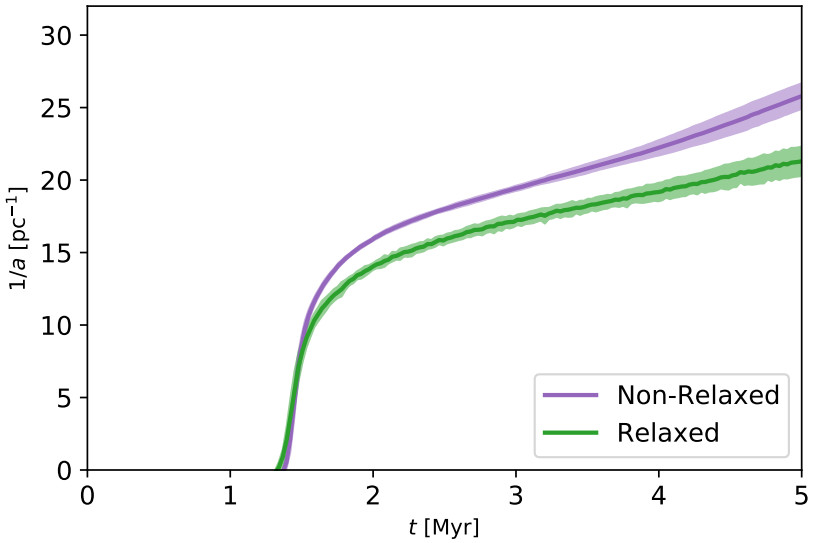

Figure 3.10: The evolution of the inverse semi-major axis $1/a$ as a function of time t for five independent realizations of `nr_q_1.0` and `r_q_1.0` simulations but with a resolution of $N \sim 4 \times 10^5$ for computational constraints. The solid lines represent the mean value of the inverse semi-major axis and the shaded region represents the standard deviation from five simulations. We find that our results are robust. This implies that the discrepancies in the results arise out of physical rather than numerical reasons.

generated from Fokker-Planck simulation rather than N-body simulations. It could also occur from different galaxy merger shapes. The interplay between triaxiality and mass segregation has not been properly studied and presents a further avenue of research.

We find similar results compared to Ogiya et al. (2020a). This verifies that the effect of NSCs in accelerating the orbital decline of MBH binaries is robust as indicated by Ogiya et al. (2020a). Our models are characterized by an initial shallow cusp (non-relaxed) or a dense cusp (relaxed), so the results are not directly comparable. The upper limit on the coalescence timescale in our simulations is $\sim$ 700 Myr for the $q = 1.0$ models whereas that found in Ogiya et al. (2020a) is $\sim$ 5 Gyr. This could be because the presence of a cusp rather than a core leads to a more efficient hardening. However, Khan and Holley-Bockelmann (2021) found coalescence timescales similar to ours for large mass-ratio mergers and attributed the discrepancy is to better resolution of the three-body scattering process with higher mass resolution. Nevertheless, we find that the merger timescale depends on the mass ratio of the binary with $q = 0.01$ mass ratio binaries merging in $\sim 150 - 170$ Myr. Interestingly, this is in contrast to the upper limit to the timescales found in Ogiya et al. (2020a).

We noted that the timescales for lower mass ratio binaries ($q \leq 10^{-2}$) is quite sensitive to the resolution. Lower resolution simulations with $N \sim 4 \times 10^5$ (see appendix 3.7.1) resulted in merger timescales of $\sim 90 - 100$ Myr for the $q = 0.01$ merger simulations. This is almost half the merger timescale found using the higher resolution simulations with $N \sim 1.32 \times 10^6$. As the mass ratio of the binary was increased, the discrepancies between the lower and the higher resolution simulations decreased. This is in contrast to the findings of Preto et al. (2011) where the authors noted that in MBH binaries formed in galactic mergers, the triaxiality of the non-spherical merger product ensured that the hardening rate is independent of N. However, our results seem to be in line with that presented in Vasiliev et al. (2015) where the authors found that in triaxial galaxies the hardening rate asymptotically reaches a fixed value as the resolution is increased. According to Vasiliev et al. (2014) collisional effects account for a non-trivial portion of the hardening rate and cannot be neglected while considering the hardening rates of MBH binaries. Since NSCs are collisional systems, it highlights the importance of resolving collisional effects properly for a proper determination of LISA timescales for MBH binaries, especially for $q \approx 0.01$. This would require the usage of even higher particle numbers, even for simulations where the initial orbit is circular. To determine the minimum resolution required, we need to perform more simulations with varying N, which is beyond the scope of this study. We plan on investigating the asymptotic limits of hardening rates as a function of the resolution in future studies. We would like to stress, however, that the binaries merge efficiently well within the Hubble time in line with the conclusions made in Ogiya et al. (2020a) about NSCs being potentially important sources for LISA detections.

3.6 Conclusions

MBH binaries are touted to be one of the most important sources of GW signals detectable by future generation of GW detectors like LISA. The presence of NSCs surrounding the MBHs can accelerate their evolution to the hard-binary and therefore, the GW emission phase. However, the dynamics of the binary can be quite sensitive to the composition of the NSC. The presence of a mass spectrum can lead to mass segregation since the NSC is a collisional system and can affect the evolution of the binary in non-intuitive ways.

In this work we have explored the effects of mergers of collisionally relaxed NSC on the dynamics of MBHs embedded in them. Using a suite of N-body simulations, we have demonstrated the non-intuitive ways in which a collisionally relaxed nuclei with a two-component initial mass function can affect the overall dynamics depending on the mass ratio of the binary. For simplicity, we considered the mass of the lighter objects to be $1M_{\odot}$ representing stars, white dwarfs, and neutron stars and the mass of heavier objects to be $10M_{\odot}$ representing stellar mass black holes.

Through the usage of a Fokker-Planck code, we evolved the NSCs to the collisionally relaxed state under the presence of an MBH at the center. We then set up the mergers with different MBH mass ratios and for comparison, also evolved mergers with non-relaxed NSCs.

During the three stages of evolution, we found that the dynamics during the pre-binary phase is similar amongst simulations, even with different mass ratios and is consistent with results from Ogiya et al. (2020a). However, due to changes in the density profile between relaxed and non-relaxed systems, differences arise during the combined phase.

The presence of a heavier mass species leads to a decline in the density profile of the lighter species for all radii greater than $\sim 0.1 d_{\rm infl}$ within the sphere of influence of the primary. As a result, in larger mass ratio binaries, the binary is able to settle at a lower separation in the non-relaxed models after the combined phase compared to the relaxed models. However, this trend slowly changes with the mass-ratio of the binary as the scattering efficiency of the binary increases with the decrease in mass-ratio and the binary is able to lose more energy by scattering the particles that are more tightly bound to the MBHs. For $q = 0.1$, we find that the evolution in the relaxed and non-relaxed models are similar and for $q = 0.01$, we find that the presence of the denser cusp actually accelerated the evolution of the binary to the hard binary stage.

In the hard binary stage, the evolution is similar amongst non-relaxed and relaxed models. The binaries harden at a fixed rate consistent with previous studies on NSC and galaxy mergers. This indicates that even when the cusp is disrupted less, the primary mode of evolution is collisionless where the loss-cone of the binary is populated by stars on centrophillic orbits. This is driven by the shape of the merger product rather than by any relaxation effects. However, we do find that the hardening rate depends on the particle number N in contrast with some previous studies. This

underscores the importance of accounting for properly accounting for collisional effects as they plan a non-trivial role in the evolution of MBH binaries in collisional systems like NSCs.

Crucially, we find that the relaxed cusp plays a big role in the determination of the eccentricity at the binary binding and hardening stages. Non-relaxed cusps show higher eccentricities at binding stage and show growth in eccentricity which is absent in relaxed cusps. The eccentricity "suppression" is quite large in the relaxed models and binaries on highly eccentric initial orbits can merge almost $2\times$ faster in non-relaxed cusps.

The expected GW coalescence time for all of our models (relaxed and non-relaxed) is significantly less than the Hubble time, making merging NSCs a promising GW source. In addition, we find that the expected coalescence time for the non-relaxed models is lower than their relaxed counterparts for all models with $q > 10^{-2}$. The eccentricity evolution of the MBH binary is strongly affected by the initial density profile of the NSCs with relaxed models exhibiting lower eccentricity. This underscores the importance of properly modeling the initial conditions of the NSC including the usage of a realistic mass-spectrum. Generation of LISA waveforms will require proper modeling of the environment surrounding the binary.

While our initial conditions are idealized in the sense that only two mass species are used, we demonstrate the necessity of modeling NSCs with multiple mass species as collisional relaxation can often evolve the density and mass profiles in non-intuitive ways which can affect the dynamics of the MBH binary. Our simulations open up avenues of further exploration with regards to IMBH-MBH mergers and the impact of triaxiality on mass segregation. We also demonstrate the effectiveness of `Taichi` at handling problems of this resolution and scale. `Taichi` is among the first N-body codes built with a fourth-order symplectic integrator with time symmetric step solver and regularization. Nevertheless, some of the simulations from this work clearly show that directly integrating over many orbits of the hard binaries can be computationally demanding. A proper treatment of stable hierarchical systems (e.g., Wang et al., 2020; Rantala et al., 2022) is worth the investment. With further improvements, `Taichi` presents an effective method to simulate galaxy mergers, MBH binaries and simulations of NSCs.

3.7 Appendix

3.7.1 Effect of resolution

To understand the effect of resolution on the results, we simulate lower resolution models of mergers on circular orbits with $N = 4 \times 10^5$. For clarity, we have only presented the full evolution for the $q = 0.01$ models. Looking at Figure 3.11, we do not find any differences in the pre-binary phase indicating that the lower resolution simulations are able to simulate the tidal stripping process as accurately as the higher

resolution simulations. The time of binary formation is consistent among the two sets of simulations as well. However, once the binary has formed, differences appear between the lower resolution and the higher resolution simulations. In simulations with $N = 4 \times 10^5$, we find that the binary is able to reach large values of eccentricity which is not observed in the simulations with $N = 1.32 \times 10^6$ particles. This was only observed in the $q = 0.01$ models but not in the larger mass-ratio models. In addition, as mentioned before, the hardening rate in the lower resolution simulation is about $2\times$ higher than that in the higher resolution simulation. To understand how the resolution plays a role in the hardening rate, we plot the average hardening rate as a function of the mass-ratio in Figure 3.12. We find that the hardening rates are consistent among the lower and higher resolution simulations for $q = 1.0$. As the mass ratio is lowered, the higher resolution models demonstrate lower rates of hardening. This indicates that although the primary method for loss-cone scattering in the hard binary phase may be collisionless, driven by the non-spherical nature of the merger product, collisional effects cannot be discounted, especially in collisonal systems like NSCs. The dependence of hardening rate on N is more prominent for models with lower q. This is in line with the observations of Vasiliev et al. (2014) where the authors found that even in non-spherical galaxies, collisional loss cone refilling can play a significant part.

3.7.2 Comparsion between NBODY6 and Taichi

We use the initial conditions for the $q = 1.0$ model directly obtained from Ogiya et al. (2020a) to test `Taichi` against `NBODY6++GPU`. We present the evolution of the orbital elements and show that using the parameters chosen in this study, `Taichi` is able to simulate the system as accurately as `NBODY6++GPU`. Even without the usage of specialized hardware such as GPUs, `Taichi` is able to simulate the systems $2\times$ faster than `NBODY6++GPU` using 2 GPUs. The energy conservation at the end of the simulation for `Taichi` was $\sim 0.1\%$ whereas that for `NBODY6++GPU` was $\sim 1\%$. For these simulations, the number of particles was the same as that used in Ogiya et al. (2020a), $N = 131072$. We find that there are no systemic differences between the evolution of the binaries using the two codes. We find the differences in the evolution of binary separation are minuscule. Even though the eccentricity evolution is inherently stochastic, qualitatively the evolution is similar in both cases. From the plot of the inverse semi-major axis, we deduce that in both scenarios the binary is hardening at similar rates. This indicates that `Taichi` is able to model both collisional and collisionless processes as accurately as `NBODY6++GPU` and is suitable to handle the class of problems mentioned in this work.

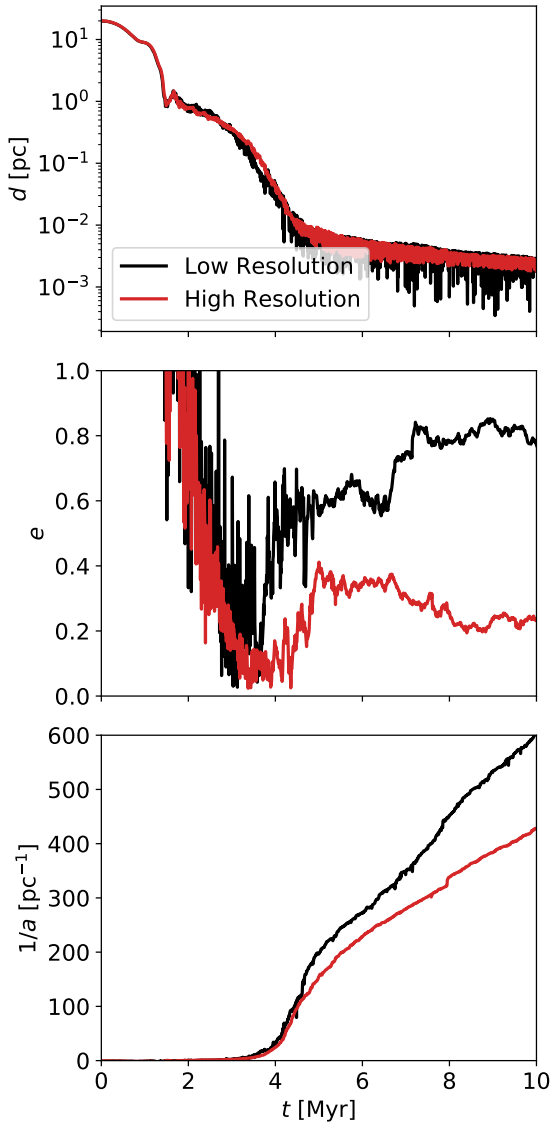

Figure 3.11: Evolution of the binary parameters for the relaxed $q = 0.01$ model presented as a function of time for lower resolution and higher resolution models. We find that although there are no differences in pre-binary phase and the time of binary formation between the lower resolution and higher resolution models, differences appear once the binary is in the bound-binary and the hard binary phases. This is quite notable for the evolution of eccentricity and the rate of hardening where the lower resolution model demonstrates a higher value compared to the high resolution model. The results are in contrast with Preto et al. (2011) as we find the hardening rate depends on N indicating that the effects of collisional loss-cone refilling cannot be discounted.

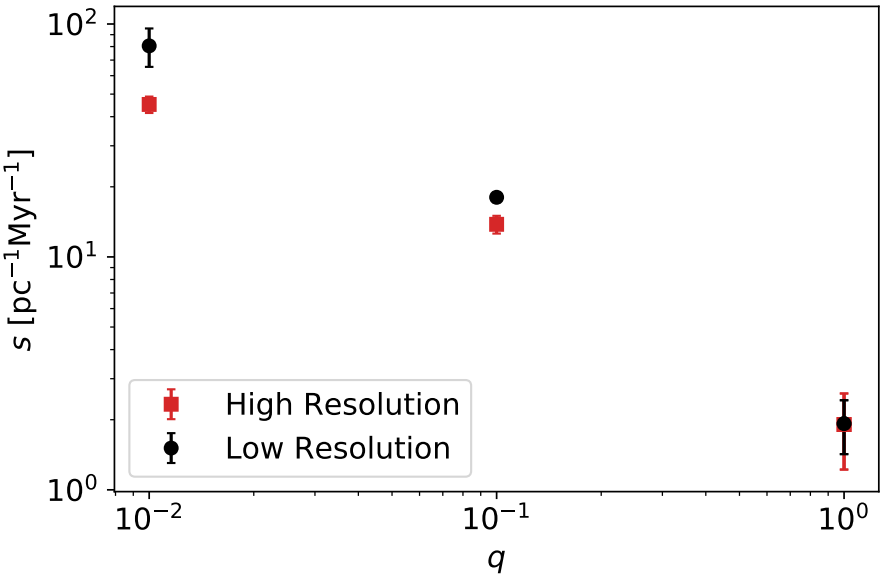

Figure 3.12: Hardening rates of circular relaxed models with different resolutions presented as a function of the mass-ratio q. The hardening rates have been computed by taking the average of the hardening rates every 1 Myr after a hard binary has been formed. The error bars correspond to the standard deviation. We find that the hardening rate strongly depends on q and resolution. As the mass-ratio is lowered, the hardening rate decreases as we increase N. Similar observations were noted for the non-relaxed scenario.

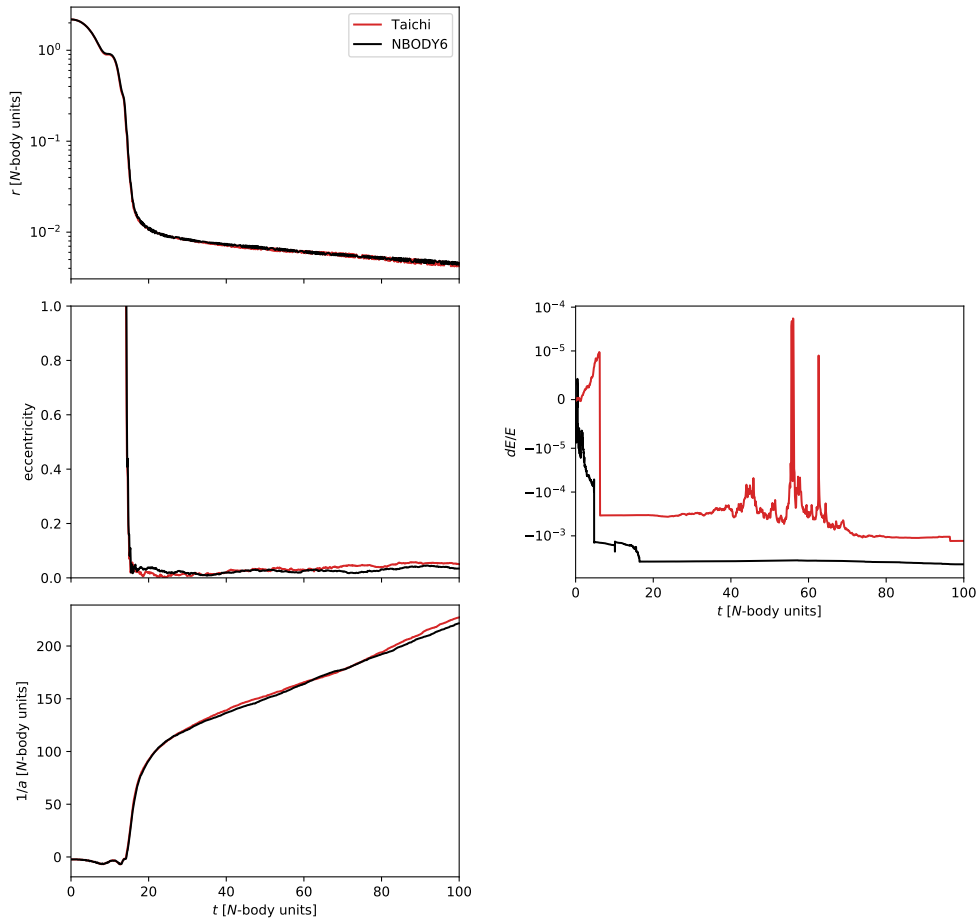

Figure 3.13: Evolution of the orbital parameters as a function of time for the $q = 1.0$ model presented in Ogiya et al. (2020a). Top: Evolution of binary separation. Middle: Evolution of eccentricity. Bottom: Evolution of inverse semi-major axis. Right: Evolution of the relative energy error. We find the results between `NBODY6++GPU` and `Taichi` are consistent with each other and `Taichi` is better at energy conservation by a factor of ~ 10 compared to `NBODY6++GPU` .

Chapter 4

Seeds sink swiftly: nuclear star clusters dramatically accelerate seed black hole mergers

Diptajyoti Mukherjee[1], Yihao Zhou[1], Nianyi Chen[1], Ugo Niccolò Di Carlo[2], Tiziana Di Matteo[1]

[1]McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213
[2] SISSA - International School for Advanced Studies, via Bonomea 365, I-34136 Trieste, Italy

Abstract

Merger rate predictions of Massive Black Hole (MBH) seeds from large-scale cosmological simulations differ widely, with recent studies highlighting the challenge of low-mass MBH seeds failing to reach the galactic center, a phenomenon known as the seed sinking problem. In this work, we tackle this issue by integrating cosmological simulations and galaxy merger simulations from the MAGICS-I and MAGICS-II resimulation suites with high-resolution N-body simulations. Building on the findings of MAGICS-II, which showed that only MBH seeds embedded in stellar systems are able to sink to the center, we extend the investigation by incorporating nuclear star clusters (NSCs) into our models. Utilizing N-body resimulations with up to 10^7 particles, we demonstrate that interactions between NSCs and their surrounding galactic environment, particularly tidal forces triggered by cluster interactions, significantly accelerate the sinking of MBHs to the galactic center. This process leads to the formation of a hard binary in $\lesssim 500$ Myr after the onset of a galaxy merger. Our results show that in 8 out of 12 models, the high stellar density of the surrounding NSCs enhances MBH hardening, facilitating gravitational wave (GW) mergers by redshift $z = 4$. We conclude that at $z > 4$, dense NSCs serve as the dominant channel for

MBH seed mergers, producing a merger rate of 0.3–0.6 yr^{-1} at $z = 4$, which is approximately 300–600 times higher than in non-NSC environments. In contrast, in environments without NSCs, surrounding dark matter plays a more significant role in loss-cone scattering.

4.1 Introduction

The formation and evolution of massive black holes (MBHs) stand as one of the most intriguing phenomena in astrophysics, playing a pivotal role in shaping the cosmic landscape from the early universe to the present epoch. Supermassive Black Holes (SMBHs) are quite ubiquitous, as observations suggest, and present in almost all galactic centers in local galaxies (e.g., Miyoshi et al., 1995; Tremaine et al., 2002; Kormendy and Ho, 2013b). Intermediate Mass Black Holes (IMBHs) ranging from $10^3 M_\odot$ to $10^6 M_\odot$ have also been found to exist at the centers of several local dwarf galaxies (e.g., Reines et al., 2013; Greene et al., 2020). Observations, particularly from the James Webb Space Telescope (JWST), have unveiled a wide population of MBHs with masses $10^6 M_\odot - 10^8 M_\odot$ in the high redshift universe (e.g., Larson et al., 2023; Übler et al., 2023; Kocevski et al., 2023; Harikane et al., 2023; Goulding et al., 2023; Maiolino et al., 2023, 2024; Matthee et al., 2024). Scaling relations derived from local galaxies show that these MBHs are typically overmassive compared to their hosts (Pacucci et al., 2023; Goulding et al., 2023) and pose new challenges to our understanding of the formation and growth of MBHs in the high redshift universe.

In recent years, significant strides have been made in understanding the origins of MBHs, suggesting that they may originate from a seed population at $z \sim 20 - 30$ (Barkana and Loeb, 2001) which then grows in mass via mergers and accretion (e.g., Dayal et al., 2019; Pacucci and Loeb, 2020; Piana et al., 2021; Bhowmick et al., 2024). These seeds can form from the direct collapse of pristine gas clouds (e.g., Begelman et al., 2006; Mayer et al., 2010) or from the collapse of massive PopIII stars and runaway growth in dense stellar systems (e.g., Portegies Zwart et al., 1999) such as nuclear star clusters (NSCs) (e.g., Devecchi and Volonteri, 2009; Lupi et al., 2014; Das et al., 2021; Askar et al., 2023; Kritos et al., 2023) or young star clusters (e.g., Di Carlo et al., 2019, 2021). However, reconciling theoretical predictions of different seeding models with observations remains a challenge due to the low masses of these seeds and their faint electromagnetic signatures (Pacucci et al., 2017).

With the advent of gravitational wave (GW) astronomy using LIGO-Virgo interferometers (e.g., Abbott et al., 2017, 2020c,b,a) and Pulsar Timing Array (PTA) (Mingarelli et al., 2017; Kelley et al., 2018; Agazie et al., 2023a,b), a new pathway has emerged to understand and constrain various seeding models in the early universe, if seed BHs can form MBH binaries and merge. In particular, MBH binaries with masses $10^4 M_\odot - 10^7 M_\odot$ are one of the main targets of the space based GW detectors like LISA (Amaro-Seoane et al., 2017) or TianQin (Luo et al., 2016) which should be able to detect mergers of MBHs out to $z > 20$. However, the dynamics of these

$10^3 M_\odot - 10^6 M_\odot$ MBH seeds are notoriously difficult to model as they involve resolving processes that can bring them from kiloparsec scales to coalescence, producing a wide range in theoretical predictions for merger rates.

The process leading to the GW coalescence of MBH binaries is often conceptualized as a three-step journey (Begelman et al., 1980; Merritt, 2013a). Initially, the dynamical friction (DF) (Chandrasekhar, 1943) exerted by stars, dark matter (DM), and interstellar gas comes into play, reducing the angular momentum of the MBHs, leading them to sink to the center of the merged galaxy. When the two MBHs get close enough, they form a bound binary, marking the onset of the second stage. During this phase, the binary's separation reduces due to a combination of dynamical friction and three-body scattering events. In the penultimate stage, further orbital decay ensues, primarily driven by three-body scattering processes. However, if insufficient scattering interactions between MBH binaries and stars occur, the orbital decay of the binary can stall before the binary can coalesce via the emission of GWs (e.g., Milosavljević and Merritt, 2003; Vasiliev et al., 2015). This highlights the importance of the environment surrounding the MBH binary on its fate.

The self-consistent evolution of MBH seeds and their galactic environments is generally treated using cosmological simulations, which offer a comprehensive framework for understanding the intricate processes governing seed growth and mergers (see Amaro-Seoane et al., 2023, for a review). However, large-volume cosmological simulations lack sufficient resolution at scales of $\lesssim$ kpc. Physical processes like dynamical friction are usually treated in a subgrid fashion (Tremmel et al., 2015a; Chen et al., 2022a; Ma et al., 2023; Damiano et al., 2024) leading to uncertainties in the dynamics of MBH seeds, especially at separations of $\leq$kpc from the galactic centers. A number of recent studies have found that MBH seeds are inefficient at sinking to the center of the galaxy leading to longer merger timescales than previously expected (Pfister et al., 2019; Ma et al., 2021; Partmann et al., 2023), creating the so-called "seed-sinking problem". Crucially, large softening lengths and lower mass resolution in cosmological simulations prevent the resolution of dense stellar systems such as NSCs, which can play a major role in the formation and evolution of MBH seeds (e.g., Devecchi and Volonteri, 2009; Das et al., 2021; Lupi et al., 2014; Askar et al., 2021).

NSCs are some of the densest known stellar systems, and galaxies hosting both NSCs and MBHs at their centers are quite prevalent across galaxy masses and morphologies (Neumayer et al., 2020a). Due to the large stellar density at their centers, they provide an ideal environment for the formation and growth of MBH seeds. Using semi-analytical models, Kritos et al. (2023) find that NSCs in dwarf galaxies can provide suitable environments for the production of MBH seeds of masses up to $10^6 M_\odot$. In the densest and most compact NSCs, runaway growth and gas accretion can lead to formation of $10^6 M_\odot$ MBHs in ~ 100 Myr.

Shi et al. (2024) perform multi-physics simulation and find that low mass MBH seeds present in star clusters hierarchically merge, leading to the formation of a proto-

bulge or a dense NSC. The star clusters surrounding the seeds help them migrate efficiently to the center of the galaxy, resulting in a rapid formation of seed MBH binaries with separations ≤ 1 pc. The process is much faster than the DF timescale of such MBH seeds, if they were isolated. On a similar note, recent works have found that MBHs embedded in NSCs undergo accelerated mergers(Ogiya et al., 2020b; Khan and Holley-Bockelmann, 2021; Mukherjee et al., 2023; Chen et al., 2024). The additional mass surrounding the MBHs helps them sink more efficiently to the galactic center. Once these MBHs reach a separation of $\lesssim$ 50pc, tidal interactions of the NSCs help them sink from 20-50 pc to a few milliparsec in a span of $\sim$ Myr, leading to the formation of a hard binary faster than the DF timescale.

Recent studies have systematically explored favorable conditions under which MBH seeds can merge, leveraging multi-scale examinations. Large-volume cosmological simulations, such as `ASTRID` (Bird et al., 2022; Ni et al., 2022; Chen et al., 2022b), along with galaxy merger re-simulations like MAGICS-I (Chen et al., 2024), have demonstrated that galaxy mergers, particularly those involving dwarf galaxies, play a critical role in driving MBH seed mergers in the high-redshift universe. Building on MAGICS-I, MAGICS-II (Zhou et al., 2024) employs higher-resolution models of galaxies, and incorporates an AR-chain integrator (Mikkola and Tanikawa, 1999c; Rantala et al., 2017, 2020) and finds that only MBH seeds embedded within extended stellar systems (which remains largely unstripped during the merge) can effectively migrate BHs to the center of the merged galaxy, where they may form a binary. In contrast, significant tidal stripping during the initial infall of the MBH host galaxy lead to naked MBH seeds in a merger remnant. In this case, MBHs stall at distances of 0.1 - 1 kpc (Zhou et al., 2024).Therefore, stellar structures, such as NSCs surrounding the MBHs, are crucial for creating conditions that facilitate MBH seed mergers. However, the resolution limits of MAGICS-II preclude the proper examination of realistic nuclear structures around MBHs, such as NSCs, highlighting the need for high-resolution N-body simulations to accurately capture the detailed dynamics of MBHs within dense stellar environments.

There is a wealth of literature available on modeling stellar clusters in cosmological simulations. While this is hard to do self-consistently, studies such as E-MOSAICS (e.g., Pfeffer et al., 2018) and FIRE (e.g., Rodriguez et al., 2023) have employed multi-scale techniques combining high resolution simulations with cosmological simulations to study the evolution of globular clusters in galaxies. In the same vein as these previous works, we employ high resolution N-body models informed from cosmological simulations in this work. Upon the merger of two galaxies, the NSCs present in either galaxy would be subject to the tidal effects from the stellar bulge or the DM halo as they sink to the center, resulting in mass loss over time and thereby affecting their sinking efficiency. A systematic study of the effect of MBH seed mergers due to NSCs in realistic galactic environments is missing. Our work, therefore, is primarily motivated to understand the efficiency of MBH seed mergers embedded in NSCs in galactic environments informed by cosmological and galaxy merger simulations.

In this work, we introduce the MAGICS-NSC suite of high-resolution N-body simulations, using up to 10^7 particles, to study the evolution of MBH seeds from a separation of hundreds of parsecs to GW emission stage under the influence of NSCs. Our models are directly informed from the MAGICS-I (Chen et al., 2024, hereafter MAGICS-I) and MAGICS-II (Zhou et al., 2024, hereafter MAGICS-II) suite of galaxy merger resimulations . We use the Fast Multipole Method (FMM) (Greengard and Rokhlin, 1987; Cheng et al., 1999) based N-body code `Taichi` (Zhu, 2021b; Mukherjee et al., 2021a, 2023) which has been shown to be as accurate as direct-summation based N-body codes while using a fraction of the computational time. Our simulations are state-of-the-art and present an effort where the evolution of the seeds is traced accurately from kiloparsec to milliparsec scales, including not only the stellar bulge but the surrounding DM halo.

We begin by introducing the MAGICS-I and MAGICS-II resimulation suites in section 4.2 and then briefly describe the computational methods in section 4.3 and our models in section 4.4. This is followed by the results in section 4.5. We then describe some of the implications of this work in section 4.6 and conclude in section 4.7.

4.2 The MAGICS Project

The simulations in this work are based on the initial conditions of galaxy mergers in MAGICS-I and MAGICS-II, which are directly informed by the MBH mergers in the `ASTRID` cosmological simulation. Here we briefly describe `ASTRID` and, MAGICS-I and MAGICS-II simulation suites.

`ASTRID` is a cosmological hydrodynamical simulation with $250h^{-1}$Mpc per side and 2×5500^3 initial tracer particles comprising dark matter and baryons (Bird et al., 2022; Ni et al., 2022). The simulation includes a full-physics, sub-grid models for galaxy formation, SMBHs and their associated supernova and AGN feedback, as well as inhomogeneous hydrogen and helium reionization. BHs are seeded in haloes with $M_{\rm halo} > 5 \times 10^9 h^{-1} M_\odot$ and $M_* > 2 \times 10^6 h^{-1} M_\odot$, with seed masses stochastically drawn from a power-law distribution within the mass range between $3 \times 10^4 h^{-1} M_\odot$ and $3 \times 10^5 h^{-1} M_\odot$, motivated by the direct collapse scenario proposed in (e.g.; Lodato and Natarajan, 2007).

The stellar mass criterion for seeding in `ASTRID` ensures that there is some cold gas remaining in the halo which is crucial to form and grow the MBH. Recent experiments involving simulations of dense stellar environments find that they provide the ideal environment to grow MBHs of masses $\geq 10^3 M_\odot$ (Partmann et al., 2024). A dense environment is necessary to anchor and grow light MBH seeds. Additionally, in `ASTRID` the stellar component is a crucial source for providing DF force onto the BHs. Thus, the halo mass criterion is chosen so that the BHs are only seeded in haloes that have sufficient stellar mass to provide sufficient DF force to the BH seeds.

The gas accretion rate onto the black hole is estimated via a Bondi-Hoyle-Lyttleton-like prescription (Di Matteo et al., 2005). The black hole radiates with a bolometric luminosity $L_{\rm bol}$ proportional to the accretion rate $\dot{M}_\bullet$, with a mass-to-energy conversion efficiency $\eta = 0.1$ in an accretion disk according to Shakura and Sunyaev (1973). 5% of the radiated energy is coupled to the surrounding gas as the AGN feedback. Dynamics of the black holes are modeled with a sub-grid dynamical friction model (Tremmel et al., 2015b; Chen et al., 2022c), yielding well-defined black hole trajectories and velocities. We boost the dynamical mass of the black holes to $1.2 \times 10^7\,M_\odot$ for gravitational interactions to alleviate the artificial dynamical heating from other particle species. Two black holes can merge if their separation is within two times the gravitational softening length $2\epsilon_g = 3\,{\rm ckpc/h}$ and they are gravitationally bound to the local potential.

`ASTRID` has the largest MBH merger population at high-redshift (Chen et al., 2022a) with MBH masses in the range $5 \times 10^4\,M_\odot < M_{\rm BH} < 5 \times 10^{10}\,M_\odot$. In MAGICS-I, we select 15 out of 2107 $z \sim 6$ MBH mergers from `ASTRID` to perform high-resolution resimulations. The resimulations have a dark matter and gas particle mass of $8000\,M_\odot$ and a stellar particle mass of $2000\,M_\odot$. The gravitational softening is 80 pc for dark matter and gas, 20 pc for stars, and 10 pc for black holes (the inter-species softening length is the maximum value of the two species). The subgrid physics models in MAGICS-I are similar to those of `ASTRID` , except that we alleviate the boost in the MBH dynamical mass due to the much higher particle resolution. The dynamics of MBH pairs are also followed to smaller separations ($2\epsilon_{\rm g,BH} = 20\,{\rm pc}$) compared with `ASTRID` .

To trace MBH binary dynamics down to smaller scales, MAGICS-II uses the `KETJU` code (Rantala et al., 2017; Mannerkoski et al., 2023) to study 6 merging systems identified in MAGICS-I. `KETJU` replaces the leapfrog integration with the algorithmically regularized `MSTAR` integrator (Rantala et al., 2020) in regions around each MBH. In these regions, interactions of BH-BH, BH-star, and BH-DM are unsoftened. Post-Newtonian (PN) corrections up to the order of 3.5 (Mora and Will, 2004) are included to account for general relativistic effects on the MBH binary. In principle, this could resolve MBH evolution down to separations of tens of Schwarzschild radii. Particle splitting is implemented in MAGICS-II. All particles, including DM, gas and stars, within 1 kpc from the binary center of mass are split to 500 $\rm M_\odot$. The gravitational softening is 20 pc for gas particles, and 5 pc for DM, stellar, and BH particles.

4.3 Computational Methods

To simulate large-N systems within a reasonable amount of time, we utilize the FMM (e.g., Greengard and Rokhlin, 1987; Cheng et al., 1999) based code `Taichi` (Zhu, 2021b; Mukherjee et al., 2021a, 2023). Mukherjee et al. (2021a) showed that `Taichi` can simulate systems as accurately as direct-summation based collisional N-body codes while scaling as $\mathcal{O}(N)$. The accuracy of the force calculation in `Taichi`

can be tuned via the usage of an input accuracy parameter (ϵ) which controls the opening angle, and a multipole expansion parameter (p) which controls the number of expansion terms used in the force calculation. Using `Taichi` we can simulate large-N systems without the usage of specialized hardware within a physically reasonable amount of time. Mukherjee et al. (2023) extended `Taichi` to include a fourth-order force-gradient integrator using the `HHS-FSI` scheme (Rantala et al., 2021c) and regularization using the `AR-Chain-Sym6+` scheme (Mikkola and Tanikawa, 1999c; Wang et al., 2021a) using the `SpaceHub` library. `Taichi` includes adaptive, individual time-symmetrized timesteps which help conserve energy better than non time-symmetrized schemes (Pelupessy et al., 2012b).

When the separation between the MBHs, Δr, is > 30 pc, we set $\epsilon = 2 \times 10^{-5}$, $p = 12$, and a timestep parameter $\eta = 0.3$. In this stage, we do not use regularization to save on computational expenses. Our experiments find that regularization does not make a significant difference in this era. For smaller separations with $\Delta r \leq 30$ pc, we set $\epsilon = 2 \times 10^{-6}$, $p = 15$, and a timestep parameter $\eta \leq 0.1$ and enable regularization. This ensures an accurate evolution of the MBH binary in the three-body hardening stage.

The interactions of the MBHs with other particles are never softened. However, we soften the interactions between particles other than the BHs using a Plummer-type softening. Different particle types have different softening lengths. The softening length for interactions between particle types i and j, ε_{ij}, is determined as:

$$\varepsilon_{ij} = \sqrt{\frac{\varepsilon_i{}^2 + \varepsilon_j{}^2}{2}} \tag{4.1}$$

where ε_i is the softening length of particle type i and ε_j is the softening length of particle type j. We do extensive experimentation to understand the effect of softening and find that there are minimal differences in the orbit of the MBHs in the simulations that use softening versus those that do not. We re-emphasize that the interactions of MBHs with other particles are *never* softened in either case. The values of softening used for different particle types are provided in table 4.1 in section 4.4.

Despite the average simulation containing $4\text{m} - 8\text{m}$ particles, our simulations take about 7-14 days of wall-clock time to run, which includes the evolution in the hard binary stage. All simulations presented in this study were performed using 48-64 threads on an `AMD Epyc 7742` node. Energy in all of our simulations is conserved to the order of 0.01% - 0.1%. Extra care is taken to ensure that the energy error is always at least an order of magnitude below the fraction of hard binary's energy to the total energy of the system.

4.4 Models

Our initial conditions are derived directly from the MAGICS-I and MAGICS-II suites of galaxy merger resimulations. The positions and velocities of all particles are ex-

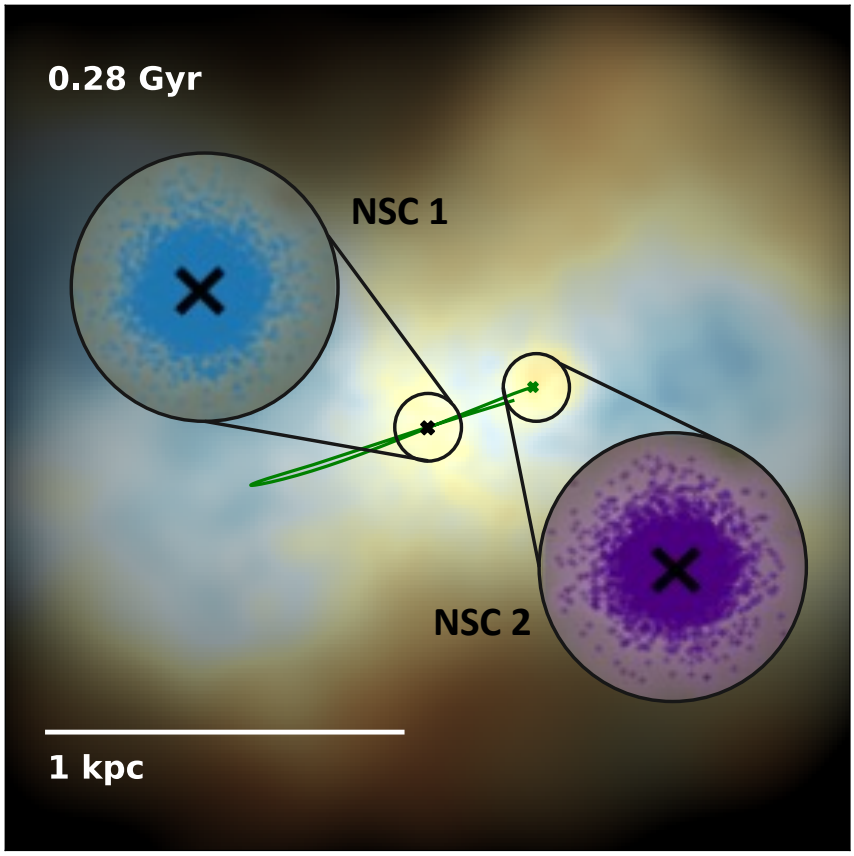

Figure 4.1: Visualization of the stellar density field of system 12 from the MAGICS-II suite. Brightness represents density, while color indicates stellar age: blue for younger stars and yellow for older stars. Overdense regions are observed around the MBHs (crosses), marking the nuclei of the original galaxies. These nuclei correspond to nuclear clusters surrounding the MBHs, but they are less dense and less massive than realistic NSCs. To assess the impact of denser and more massive NSCs, we introduce NSCs (blue and purple particles in the zoomed-in circles) around the two MBHs and continue their evolution in this study.

tracted from the MAGICS-I N-body data. Once the data is set up and the NSCs are added, our simulations proceed in two stages: `STAGE-I` where the system is evolved until the MBHs reach a separation of $\Delta r \approx 30$ pc, and `STAGE-II` where the MBHs subsequently form a hard binary and shrink further via three-body interactions. To determine the hard-binary radius $r_{\rm h}$, we first find the influence radius $r_{\rm infl}$. Following Merritt (2013a) equation (8.71), the influence radius of the binary $r_{\rm infl}$ is defined as

$$r_{\rm infl} \equiv r_{\rm enc}(2(M_p + M_s)) \tag{4.2}$$

where M_p and M_s are the masses of the primary and secondary MBHs respectively, and $r_{\rm enc}$ is the radius enclosing twice the mass of the binary. We note that M_1 and M_2 are also used to denote the masses of the MBHs in the simulations. When the MBHs get sufficiently close to begin the process of formation of a bound binary, we switch to M_s and M_p. The hard binary radius $r_{\rm h}$ can then be determined as

$$r_{\rm h} = \frac{q}{(1+q)^2}\frac{r_{\rm infl}}{4}. \tag{4.3}$$

where $q \equiv \frac{M_s}{M_p}$ is the mass-ratio of the MBH binary.

Dynamical processes such as DF and three-body hardening are sensitive to the MBH mass to particle mass ratio (e.g., Pfister et al., 2019; Genina et al., 2024). To ensure that the ratio is large enough, we perform particle splitting in the same manner as Khan et al. (2012, 2016). We then add the NSCs around the MBHs, assuming that they follow the Dehnen (1993) density profile with a shallow $\gamma = 0.5$ cusp. We describe the steps used to generate the initial conditions in further details below.

4.4.1 System and snapshot selection

We select a subset of 6 systems from the original MAGICS-I set where the MBH binaries merge within 1.2 Gyr. We also simulate an additional system, system 6, from the MAGICS-I set in order to expand the range of environments in which the mergers happen. Once the systems are chosen, we identify the time when the separation between the MBHs is greater than 300 pc and the gas density is about an order of magnitude lower than the stellar and DM density within a kiloparsec from the potential center of the galaxy. This is done since `Taichi` cannot handle gas effects. The gas particles in our simulations only interact via gravity. The lower gas density ensures that active star formation in the inner few hundred parsecs does not significantly affect our results.

4.4.2 Particle splitting

The stellar, gas, and DM particle masses in the resimulations suite are $2 \times 10^3 M_\odot, 8 \times 10^3 M_\odot$, and $8 \times 10^3 M_\odot$ respectively. This is somewhat insufficient to resolve DF and

three-body hardening effects directly. To increase the particle mass resolution, we split each particle into more particles. Particle splitting is performed in the same manner as Khan et al. (2012, 2016). Each DM or gas particle is successively split over a uniform sphere of radius equaling its softening length until the mass of each split particle is $500M_{\odot} - 1000M_{\odot}$ while each stellar particle is split until the mass of each split particle is $250M_{\odot} - 500M_{\odot}$. The split particles are assigned the same velocity as that of the parent particle. This conserves energy and momentum. In each system, the mass ratio of the least massive MBH to the particle mass is ensured to be greater than 100.

Splitting increases the number of particles in the simulations by factors of $5 - 20$. To ensure that the simulations can be performed within a physically reasonable amount of time, splitting is only performed within 1 kpc from the potential center of the galaxy. Afterwards, radial cut at a distance of 2 kpc is performed. Experiments are performed to ensure that the radial cuts did not significantly affect the dynamics of the binary or the overall mass profile in the region of interest. The above procedures result in each system containing $N \approx 4 \times 10^6 - 8 \times 10^6$. The systems are evolved until the MBHs reach $\Delta r = 30$ pc. This marks the end of `STAGE-I` of our simulations.

When the MBHs reach $\Delta r = 30$ pc, `STAGE-II` begins. We perform a radial truncation of the particle dataset at 1 kpc from the minimum potential center and perform particle splitting again to ensure that *all* particles are $250M_{\odot} - 500M_{\odot}$. In our higher resolution models, the stellar particles are split until they have a mass of $62.5M_{\odot}$. The three-body evolution is sensitive to the mass-ratio of the particle to the secondary, necessitating the particle splitting procedure for all particles. This results in a sufficient resolution to resolve the binary binding stage and the three-body interactions that lead to hardening of the binary over time. In most models, this leads to $N_{\rm hb} \approx 3.5 \times 10^6 - 7.5 \times 10^6$, where $N_{\rm hb}$ is the number of particles in the hard binary stage. Across our suite of simulations, we are able to achieve $10^2 < M_{\rm MBH}/M_{*} < 1.5 \times 10^3$ The mass ratio of the secondary MBH to that of the stellar particles is comparable to the values provided in Khan and Holley-Bockelmann (2021). For most models, the evolution of the MBHs is followed down to a semi-major axis $a \leq r_{\rm h}/10$.

The split particles are assigned softening values as denoted in table 4.1 whereas the unsplit particles in `STAGE-I` of the simulations retain the softening values from the original MAGICS-I set. The NSC stars are assigned zero softening.

4.4.3 Generating the nuclear star clusters

In MAGICS-II, the authors find that the presence of extended stellar systems surrounding the MBH seeds are critical to the sinking process. These stellar systems represent the unstripped nuclei of the original galaxies that grow over time due to star formation. A visualization of these stellar systems for system 12 from their simulations is provided in Figure 4.1. The authors find that when the extended stel-

Simulation stage	Particle type	Softening length
STAGE-I	Bulge stars	5 pc
	NSC stars	0 pc
	Gas	5 pc
	DM	20 pc
STAGE-II	Bulge stars	0.01 pc
	NSC stars	0 pc
	Gas	0.01 pc
	DM	10 pc

Table 4.1: List of inter-particle softening values used for different particle types that undergo splitting in our simulations. Plummer-type softening is incorporated into the force calculations. The interactions between the NSC stars are never softened. In STAGE-II we reduce the softening to better resolve loss-cone scattering. We remind the reader that the interactions between MBHs and other particles are never softened.

MAGICS-I/II system name	Model name	$N_{\rm gal}$ $[10^6]$	$N_{\rm NSC}$ $[10^3]$	$M_{*,\rm gal}$ $[10^8 M_\odot]$	$M_{\rm NSC,1}$ $[10^6 M_\odot]$	$M_{\rm NSC,2}$ $[10^6 M_\odot]$	$M_{\rm MBH,1}$ $[10^4 M_\odot]$	$M_{\rm MBH,2}$ $[10^4 M_\odot]$
System 1	sys1_FRe_O	7.7	44	2.7	9.03	1.88	6.00	16.7
System 2	sys2_FRe_O	6.2	97	9.2	41.1	7.20	13.2	26.7
	sys2_VRe_O							
System 3	sys3_FRe_O	5.2	10	1.2	2.45	2.26	8.69	5.15
	sys3_FRe_HI		80					
System 6	sys6_FRe_O	3.0	8	1.9	2.34	1.26	25.4	12.5
System 7	sys7_FRe_O	5.6	17	1.8	5.16	3.25	14.5	8.65
System 10	sys10_FRe_O	5.7	34	6.8	12.3	4.67	19.5	25.9
	sys10_VRe_O							
System 12	sys12_FRe_O	5.8	11	1.8	2.65	2.65	54.3	8.76
	sys12_FRe_HI		88		2.65	2.65		
	sys12_LOWM		11		0.35	0.35		

Table 4.2: Initial conditions for `STAGE-I` of our simulations. The columns denote the number of particles in the galaxy $N_{\rm gal}$ after splitting and truncation including stellar, DM, and gas particles, the total number of particles used to model both NSCs $N_{\rm NSC}$, the stellar mass of the galaxy $M_{*,\rm gal}$, masses of each NSC $M_{\rm NSC,1}$ and $M_{\rm NSC,2}$ and the MBHs present in those NSCs. $N_{\rm gal}$ for a system is kept constant between different models for each system. $N_{\rm NSC}$ depends on the mass resolution of each NSC particle. For the original resolution `O` models, we set the particle mass to be $500 M_\odot$ ($250 M_\odot$ for system 1) whereas in the higher resolution `HI` and low mass `LOWM` model we set that to $62.5 M_\odot$. Our average simulations consists of $N \sim 6 \times 10^6$ particles in this stage.

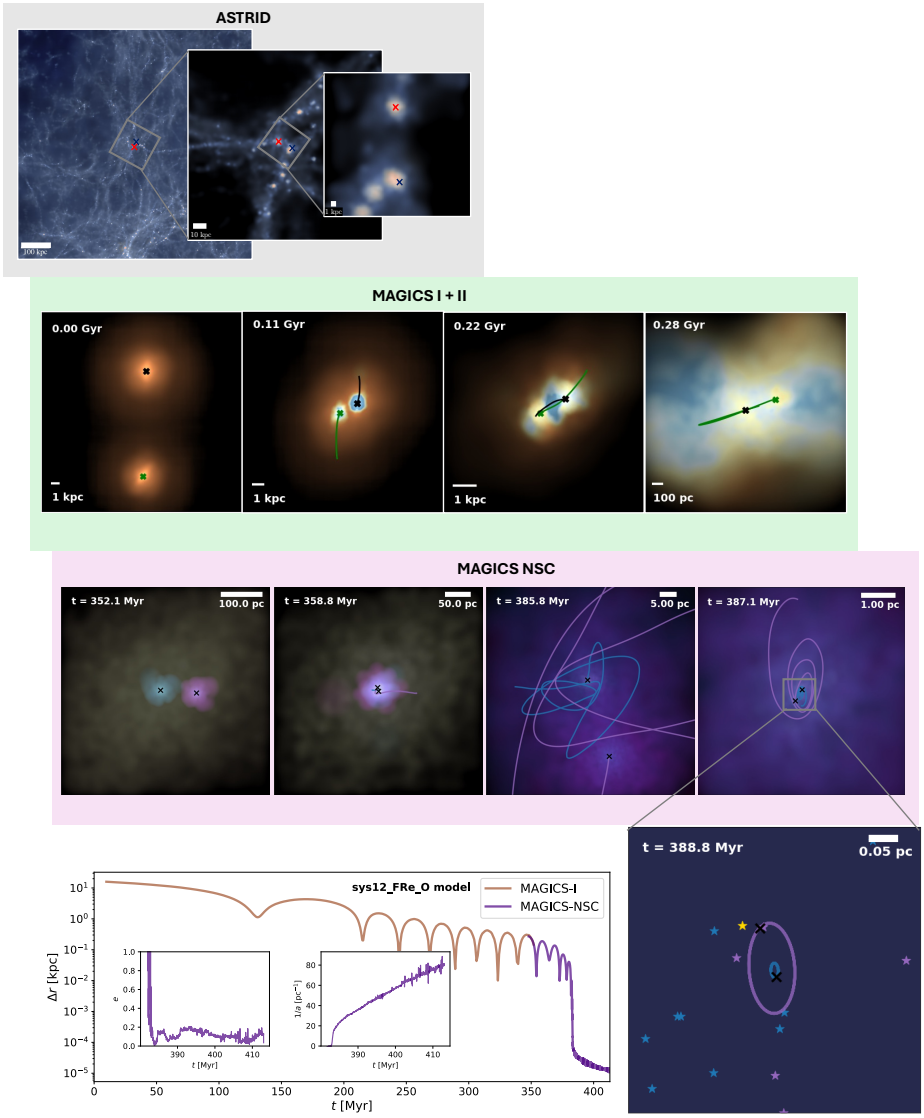

Figure 4.2: Visualization of the MBHs from kpc to mpc scales for system 12. Top: Zoom-in of the ASTRID volume showing the galaxies hosting MBHs (crosses), as resimulated in MAGICS-I and MAGICS-II. Middle: High-resolution merger resimulation from MAGICS-II displaying two galaxies from ASTRID with brightness indicating density and color indicating age (blue for younger stars, red for older). MBHs reach a separation of $\sim$ 300 pc within$\sim$ 300Myr. Bottom: Central 600 pc of the galaxy, embedding two seeds inside NSCs (blue and purple spheres) within the stellar bulge background (yellow). Interactions between NSCs enhance MBH sinking, forming a hard binary by 388.8 Myr. Bottom left: Evolution of the binary separation Δr as a function of the time t. The initial phase (brown line) is followed in MAGICS-I whereas the later phase with the MBHs inside clusters (violet line) is followed in this work. Insets show the evolution of the eccentricity e and inverse semi-major axis $1/a$ after a bound binary has formed. Bottom right: Binary visualization at a few mpc, showing the secondary's orbit (purple line) around the primary (blue line). Star colors indicate origins: blue (left NSC), purple (right NSC), yellow (bulge).

lar systems surrounding the MBH seeds are fully stripped, the MBHs stall at large separations and are unable to efficiently sink to the center to form a binary.

The finite particle limit and force softening hinder the accurate resolution of nuclear substructures such as nuclear star clusters (NSCs) and their dynamical effects in MAGICS-II. The limited mass resolution introduces mass loss, primarily due to artificial stripping caused by two-body interactions. This energy loss is proportional to $\frac{\ln(N)}{N}$, where N is the number of particles. Consequently, achieving a more accurate estimation of stellar system mass loss over time necessitates higher particle numbers. Additionally, the softening length employed for stellar particles results in a cored density profile within $4 \times \epsilon_s$ ($\sim$ 100 pc), where ϵ_s denotes the softening length. Studies have shown that such cored density profiles are subject to enhanced tidal stripping (e.g., Errani and Navarro, 2021; Du et al., 2024), leading to a significant reduction in density after a few orbits. In contrast, in the presence of an NSC, we expect a cuspy profile in the central few parsecs of the nucleus. These cuspy profiles are more resistant to tidal stripping and tend to retain a significant fraction of their mass, suggesting that, under ideal conditions, the NSC would survive. The aforementioned limitations lead to the presence of an extended stellar system only surrounding system 12. A detailed analysis of the stellar profile of system 12 shows that the mass ratio of the two galactic nuclei is initially equal. In contrast, other systems with unequal mass ratios experience complete disruption of the less massive nucleus during the merger process, leading to the formation of a naked MBH seed. However, this outcome is an artifact of force softening, as the presence of initial cusps, representative of NSC, would have resulted in their survival. This observation motivates further investigation into the effects of NSCs in systems that experienced stripped nuclei in this work.

As a first step, we analyze system 12 from MAGICS-II to determine the mass contained within 50 pc of each MBH when they are more than 100 pc apart. We find that roughly $3.5 \times 10^5 M_\odot$ is contained within each nucleus. We, then, assume the mass represents an NSC and follows the Dehnen (1993) density profile which is defined as

$$\rho(r) = \rho_0 \left(\frac{r}{r_0}\right)^{-\gamma_{\rm cl}} \left(1 + \frac{r}{r_0}\right)^{\gamma_{\rm cl}-4} \tag{4.4}$$

where ρ_0 is a normalization parameter, r_0 is a scale radius, and $\gamma_{\rm cl}$ controls the slope of the inner density profile. We set $\gamma_{\rm l} = 0.5$ indicating a shallow inner cusp, and $r_0 = 1.4$ pc. r_0 is related to the effective radius $R_{\rm eff}$ of the NSCs as

$$R_{\rm eff} \approx 0.75 r_0 \left(2^{1/(3-\gamma_{\rm cl})} - 1\right)^{-1}. \tag{4.5}$$

For our choice of r_0, we find $R_{\rm eff} = 3.3$ pc.

The NSCs are generated taking into account the cluster potential and the MBH potential. We perform experiments to understand the effect of the galactic potential on the stability of our profiles and we find that our NSCs remain sufficiently stable as the galactic potential is subdominant compared to the cluster and MBH potentials. The NSCs are assumed to be spherical and isotropic initially and are generated using

`Agama` (Vasiliev, 2019a). They are then placed around the MBHs. We note that while the mass of each NSC in this case is determined from MAGICS-II, the positions and velocities of other particles, including the MBHs, are determined from the same system in the MAGICS-I set. This is done in order to be consistent with the other models. This method produces a lower bound on the mass estimates for the NSCs and this model is labeled as `sys12_LOWM`. The mass of each NSC particle is set to $62.5 M_{\odot}$ for this particular model. Given that the effective radii of typical NSCs are comparable to, or less than the stellar softening length used in the MAGICS-II simulations, the mass estimates are a lower bound on NSC masses.

To better ascertain the masses of the NSCs used in other models, we follow a simple prescription using the particle dataset from MAGICS-I. We ensure that our masses are informed from the galactic environment by assuming that the total initial mass of both NSCs is equal to the mass contained within 100 pc from the center of mass of the MBH binary at the end of MAGICS-I resimulations. We compare the obtained NSC mass $M_{\rm NSC}$ to the galaxy stellar mass $M_{*,\rm gal}$ for each model to ensure that it is physically realistic. We find that the total NSC mass is on the upper end of the $M_{*,\rm gal} - M_{\rm NSC}$ mass relation from Neumayer et al. (2020a) and is also consistent with the $M_{*,\rm gal} - M_{\rm NSC}$ relationship from Pechetti et al. (2020).

Once the total mass in NSCs for each system is found, we then determine the mass of each individual NSC in a particular system. The masses of the two NSCs, $M_{\rm NSC,1}$ and $M_{\rm NSC,2}$, are based on the mass ratio of the inner 100 pc of the stellar bulge of each galaxy right before the galaxies merge. This helps capture some information related to initial star formation in the centers of each galaxy. Once the mass of each individual NSC has been found, we follow the same steps as before to generate N-body representations assuming a Dehnen (1993) density profile with a $\gamma = 0.5$ cusp. For simplicity we first generate models where we set $r_0 = 1.4$ pc for all clusters, thereby fixing the effective radius. We label these models as `FRe` models indicating all such NSCs have an initial *fixed* $R_{\rm eff}$. Although $R_{\rm eff}$ stays constant as a function of $M_{\rm NSC}$, the generated clusters are consistent with observed NSCs with similar masses (e.g., Georgiev et al., 2016). To understand the effect of changing $R_{\rm eff}$ as a function of NSC mass, we perform simulations on a subset of systems, systems 2 and 10, where we calculate the initial $R_{\rm eff}$ of the clusters using the following relation from Pechetti et al. (2020):

$$\log\left(\frac{\mathrm{R_{eff}}}{\mathrm{pc}}\right) = 0.33 \log\left(\frac{\mathrm{M_{NSC}}}{10^6 \mathrm{M_{\odot}}}\right) + 0.36. \tag{4.6}$$

These models are labeled with `VRe` indicating that the NSCs have a *varying* $R_{\rm eff}$. While the overall generation of the clusters is not fully self consistent since we are adding extra mass to the initial system, we find that in absence of tidal interactions between clusters, the NSCs remain quite stable. The added NSC mass varies between 2-4% of the overall stellar mass in the galaxy and does not affect the global potential.

For most of our models, the mass of each individual NSC particle is set to be

Model	$N[10^6]$	M_s/M_*
sys1_FRe_0	4.49	240
sys2_FRe_0	5.16	266
sys2_VRe_0	5.46	
sys3_FRe_0	4.19	104
sys3_FRe_HI	7.59	825
sys6_FRe_0	4.08	250
sys7_FRe_0	3.89	173
sys10_FRe_0	4.19	390
sys10_VRe_0	4.15	
sys12_FRe_0	3.86	175
sys12_FRe_HI	10.2	1401
sys12_LOWM	7.46	1401

Table 4.3: Initial conditions for STAGE-II of our simulations. Once the MBHs reach a separation of $\Delta r \leq 30$ kpc, the second stage of our simulation begins. We perform further truncation to 1 kpc and ensure *all* particles have a mass of $500M_\odot$ ($250M_\odot$ for sys_1_FRe_0) in the 0 models. For the HI models, all stellar particles are split until they have a mass of $62.5M_\odot$. The mass ratio of the secondary to the stellar particles M_s/M_* is always greater than 100 for a sufficiently smooth evolution in the hard-binary stage.

$500M_\odot$. For better resolution, we set the NSC particle mass to be $250M_\odot$ for system 1. These models are labeled with 0 suffix. For two of our systems, systems 3 and 12, we generate higher resolution NSCs to test the convergence of our results. The mass of each NSC particle in these models are set to be $62.5M_\odot$. The higher resolution models are labeled with HI suffix. In all of our models, the NSC particles are set to have zero softening. The initial conditions for all our models for both stages of the simulations are summarized in tables 4.2 and 4.3.

4.5 Results

4.5.1 Initial evolution and sinking of the MBHs

Our simulations all start roughly when the MBHs are $\Delta r \approx 300$ pc apart. We present visualizations of the sinking process for two of our models, sys2_FRe_0 and sys6_FRe_0, in Figure 4.3. In sys2_FRe_0, we find that the process is extremely efficient with the MBHs forming a bound binary and sinking to $\sim 10^{-3}$ pc within 16 Myr from the start of the simulation. The fast sinking is aided by the high stellar density of the surrounding bulge and the large mass of the NSC hosting M_2 with $M_{\rm NSC,2} \approx 7.2 \times 10^6 M_\odot$. In fact, sys2_FRe_0 is the fastest sinking model amongst all of the models simulated. This is in contrast to sys6_FRe_0 where the MBHs take 85

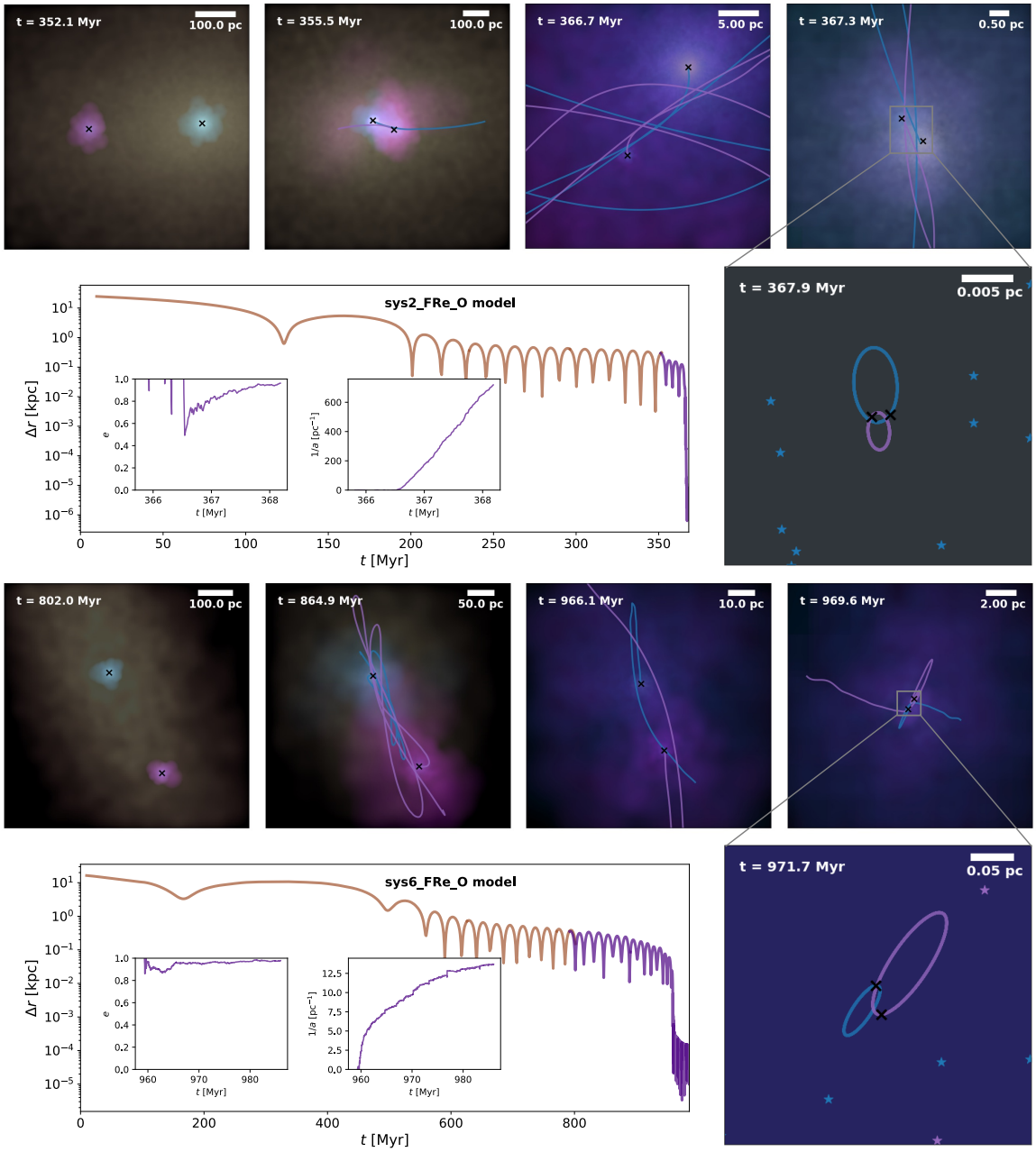

Figure 4.3: Visualization of two different models: `sys2_FRe_0` and `sys6_FRe_0` from our suite of simulations. The color scheme is same as that used in Figure 4.2. Top: `sys2_FRe_0`, a fast shrinking system. Due to the high stellar density of the surrounding galactic stellar medium and the large NSC masses, the binary shrinks quickly forming a hard binary with 15 Myr from the start of the simulation. As shown in the inset axes, the formed binary reaches a high eccentricity of 0.95 by 368 Myr while hardening to $a \sim 10^{-3}$ pc. Bottom: `sys6_FRe_0`, a slow shrinking system. Contrary to the other system, the MBHs take 85 Myr to shrink to a bound binary stage in this case. The formed binary is very eccentric (0.98-0.99) as the other model (`sys2_FRe_0`) but the hardening rate is about three orders of magnitude lower. Consequently, the binary is only able to shrink to ~ 100 mpc by the end of our simulation.

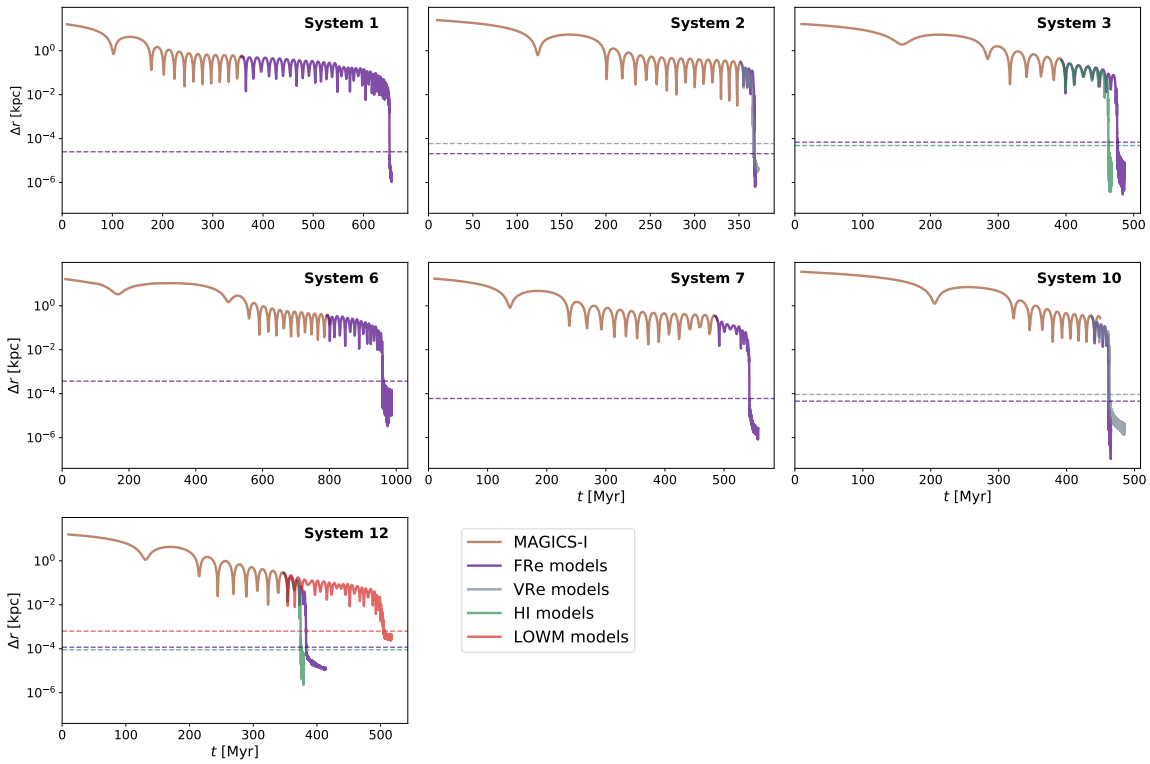

Figure 4.4: The evolution of the MBH separation Δr as a function of time t for all our models. Across all models, the binaries shrink to below the hard binary radius (dashed line) indicating that NSCs are efficient at accelerating MBH sinking times. The initial effective radii $R_{\rm eff}$ of the NSCs are *fixed* in the `FRe` models (purple lines) whereas they are *varied* in the `VRe` models (grey) following equation 4.6. The `VRe` models show qualitatively similar evolution to that of the `FRe` models with differences of at most $1-2\%$ in the sinking times. We notice that our higher resolution models `HI` models (green) merge faster than their original `O` resolution counterparts. Increasing the resolution leads to a less efficient tidal stripping by the galactic environment and causes the clusters to retain more mass. In both systems 3 and 12, the `HI` models sink about 20% faster. The `LOWM` model (red line) shrinks to the largest hard binary radius as the NSCs are about an order of magnitude less massive than the `FRe` model.

Myr from the start of the simulation to form a bound binary. The NSC hosting M_2 is only 17% as massive as that in `sys2_FRe_0` model. The central stellar density is also $10\times$ lower than that in `sys2_FRe_0` reducing the efficiency of DF leading to a longer inspiral.

We examine the evolution of Δr as a function of time in Figure 4.4 for all models, finding that the MBHs are able to sink and form a hard binary, shrinking significantly below $r_{\rm h}$. In all but one model, `sys1_FRe_0`, the MBHs sink within 100 Myr from the start of our simulations. `sys1_FRe_0` is unique in that it takes $\approx$ 300 Myr for the MBHs to sink to below the hard binary radius. This is primarily caused due to the fact that the NSC hosting the secondary is only $1.5 \times 10^6 M_\odot$ and the initial separation of the MBHs is slightly larger than that in other simulations. Incidentally, Zhou et al. (2024) find that system 1 also ends up stalling at the largest scale ($\gtrsim$ 400 pc) amongst all systems studied in MAGICS-II.

While the sinking time is dependent on the galactic environment and the masses of the MBHs, it is still informative to examine the mean sinking time $\tau_{\rm sink,avg}$ across all our simulations. To calculate $\tau_{\rm sink}$ for each model, we note the time when the MBHs reach $\Delta r = 0.1$ pc, close to the average of influence radii across all systems. Using only the data from the `FRe` models to be consistent, we find that, $\tau_{\rm sink,avg} \approx 540$ Myr. We compare this to the average sinking time for the MBHs derived from MAGICS-I suite and find that the our sinking times are 20 % shorter. However, we emphasize that the the sinking time in the MAGICS-I suite is calculated when the MBHs reach $\Delta r \approx 20$ pc. Thus the actual sinking time without the NSCs may be substantially higher. This is corroborated by comparing our models to those from the MAGICS-II suite where the authors follow the evolution down to sub-pc scale and find sinking times of the order of $\sim$Gyr in simulations where the binary does not stall.

One might ask whether this accelerated sinking is caused due to the additional mass of the NSCs surrounding the MBHs which enhances the DF force experienced by the MBH+NSC system or if it is due to the tidal interactions between the NSCs and the galactic environment. To examine this, we perform an experiment with `sys10_FRe_0` where, instead of adding the NSC surrounding the MBHs, we boost the mass of the MBHs by the same amount and plot the evolution of the MBH separation in Figure 4.5. We find that when the MBHs are separated by $\Delta r \gtrsim 100$ pc, the orbits shrink by roughly the same amount. There are only minor differences between the orbits in the first 20 Myr. This indicates the primary phenomenon driving the orbital shrinkage in this case is DF from the additional mass. However, once the MBHs reach a separation of ≈ 50 pc, there is a period of rapid orbital shrinkage when the MBHs are surrounded by NSCs. This is driven by the tidal interactions between the two NSCs which exert torques on the MBHs leading to a rapid decrease in their angular momentum and energy. The MBHs form a hard binary at the end of this phase, which is very rapid and lasts ≤ 1 Myr. This is consistent with the findings of Ogiya et al. (2020b) where the authors note that tidal interactions dominate over DF at separations of $\Delta r \sim 50$ pc. In the boosted MBH mass case the binary continues to

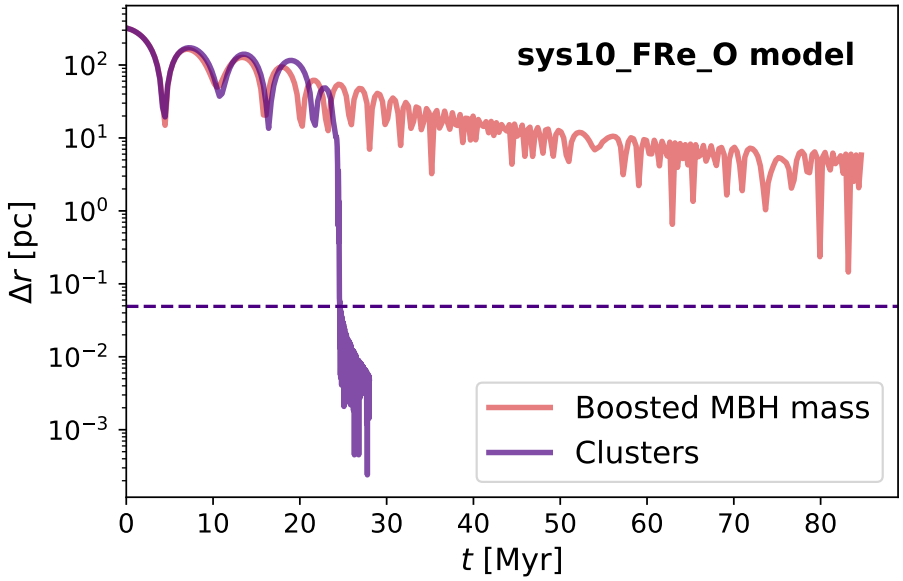

Figure 4.5: The binary separation Δr as a function of time t for `sys10_FRe_0` model with MBHs in NSCs (purple) and with the MBH masses boosted by the amount of mass present in the respective NSCs (red). While the early evolution is similar in both cases, indicating that it is dominated by DF from the increased mass, the same cannot be said for $t > 20$ Myr. The orbits shrink by roughly the same amount until the MBHs reach $\Delta r \sim 50$ pc, after which the tidal interactions between the NSCs drive the MBHs to rapidly sink to the center of the merged NSC in < 0.5 Myr and form a hard binary. However, the boosted MBH mass model does not demonstrate this rapid shrinking leading to a much longer sinking timescale. This indicates that NSCs are a key ingredient of rapid MBH binary formation.

shrink due to DF and is only able to reach $\Delta r = 10$ pc in 85 Myr. This indicates the importance of extended stellar systems such as NSCs in accelerating the formation of bound MBHs. We also perform an additional simulation where we only embed the secondary MBH in system 12 in an NSC and find that the MBHs do not form a binary within $\gtrsim 200$ Myr.

`sys12_LOWM` takes substantially longer to sink than its more massive counterpart, `sys12_FRe_0` . This is caused due to the low mass of the NSCs, almost a factor of 10 smaller than that in `sys12_FRe_0`. Nevertheless, a hard binary forms around ≈ 500 Myr. The binary settles to separation of $\lesssim 1$ pc and continues further hardening via three-body interactions. We compare the sinking time between this model and the equivalent system from the MAGICS-II and find that they are similar. Zhou et al. (2024) find that the MBHs sink to ~ 1 pc around 600 Myr from the beginning of their simulation. The similarity in the sinking separations and timescales indicates the consistency between our simulations.

4.5.2 Density profile of the formed NSC

Once the two NSCs have merged and before a hard binary has formed, we measure the overall stellar density profiles and compare them to observations of known NSCs. We choose the set from Nguyen et al. (2017, 2018), which includes density profiles of M32, NGC5102, NGC5206, NGC404, and NGC205, as well as sets from Pechetti et al. (2020) and Georgiev et al. (2016). All these sets contain local NSCs, while our galaxies are present in the high redshift universe. Nevertheless, these comparisons allow us to understand how our models relate to present-day NSCs in similar mass galaxies. We first provide a comparison of our formed NSCs against the Nguyen et al. (2017, 2018) set as the galaxy stellar masses used in that set are comparable to those in our models.

Comparing density profiles across our models to the set from Nguyen et al. (2017, 2018) in Figure 4.6, we find that stellar density in the inner 10 pc is quite consistent with that from Nguyen et al. (2017, 2018). For `sys2_FRe_0` , we notice $\rho_{1\,\mathrm{pc}} \approx 2\times10^6 M_\odot \mathrm{pc}^{-3}$, highest among all our models and about a factor of 10 larger than the densest cluster from Nguyen et al. (2017, 2018), M32. This is caused due to the high initial density of the NSC harboring the primary MBH. However, we notice that the `VRe` counterpart for the same system produces a density profile that is in agreement with the observations over the entire radii. `sys6_FRe_0` produces the least dense cusp among all our models (other than the `LOWM` model), with $\rho_{1\,\mathrm{pc}} \approx 4\times10^3 M_\odot \mathrm{pc}^{-3}$. This is unsurprising given the low initial masses and densities of the NSCs in that particular case.

The densities in the `VRe` models are systematically lower than their `FRe` counterparts in the inner parsec, although they are higher in the outer parts at $r \gtrsim 10$ pc. This is caused due to the fact that the `VRe` models have a lower central density. Since the masses of the NSCs are not changed between the `FRe` and `VRe` models, they

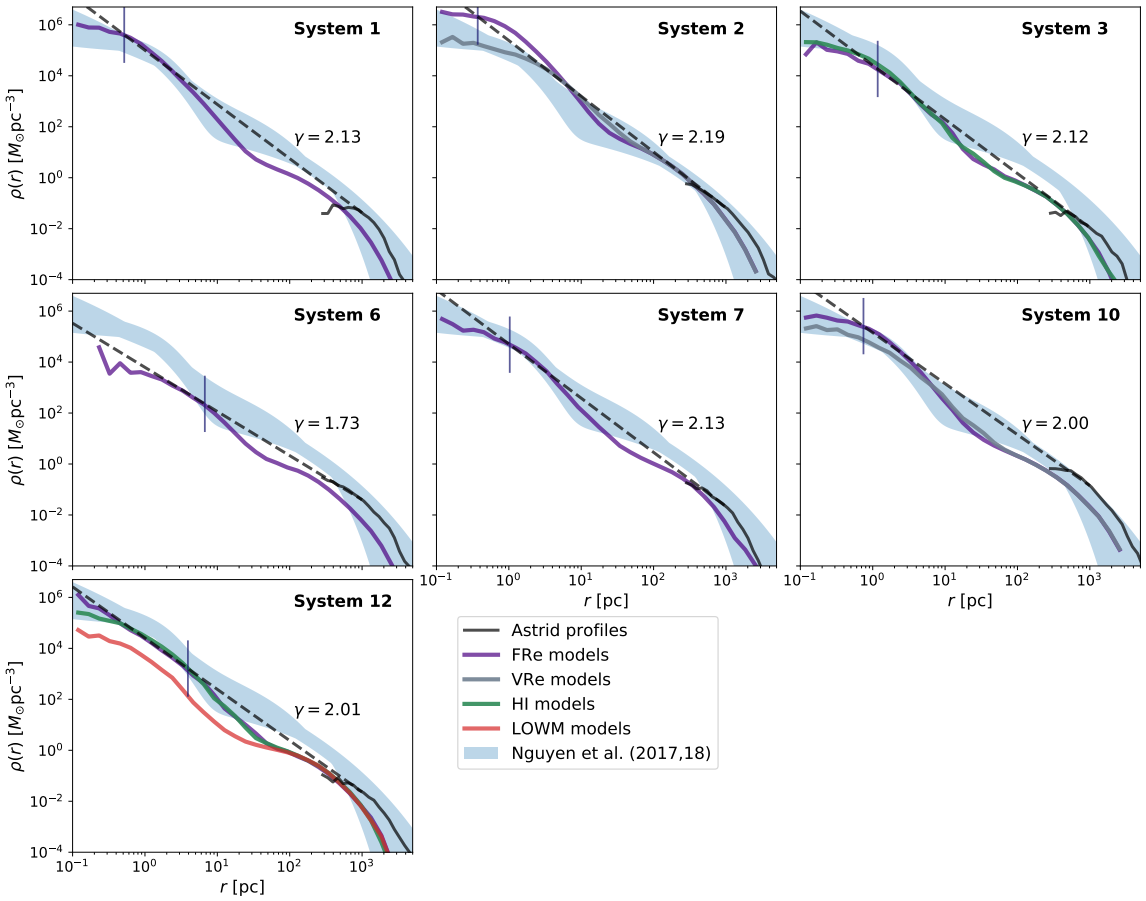

Figure 4.6: The stellar density $\rho(r)$ at the moment the NSCs merge as a function of distance from the center of potential of the system r for our models and the equivalent systems from ASTRID (black line). The different line colors indicate the different models, similar to Figure 4.4. The profiles are measured before any significant scouring effects due to the formation of the hard MBH binary. The ASTRID stellar profiles are more dense in the outskirts than our profiles as we neglect further galaxy mergers. The stellar density in $r < 10$ pc is quite consistent with observations (blue shaded region) of nucleated local dwarfs with similar galactic stellar masses (Nguyen et al., 2017, 2018) as the galaxies in our suite. We also calculate the slope extrapolation parameter γ, and the extrapolated stellar density profile (black dashed line), extrapolating the ASTRID density profiles from 1 kpc to the influence radius of the binary (vertical line). Averaging across simulations, we find $\gamma \approx 2.04$, with denser NSCs having a slightly higher $\gamma \approx 2.1 - 2.2$.

have higher densities in the outskirts. Our higher resolution NSC models, the HI models, produce profiles that are in strong agreement with their original resolution O counterparts. We compare the density profiles at various radii between the O and HI models and find that the profiles differ by a factor of 2-3 at $r \sim 1$ pc at most. The HI models retain more mass than their O counterparts, leading to a slightly higher density in certain parts.

We do notice that our stellar bulge density to NSC density is somewhat larger than the observations from Nguyen et al. (2017, 2018). This is primarily caused by our initial NSC generating prescription which produces slightly more massive NSCs compared to the galaxy stellar mass. However, core scouring caused by the MBH binary can lead to significant erosion (e.g., Merritt and Szell, 2006b; Merritt et al., 2007a). As we will show in the next section, it can lead up to a 70% decrease in the density profile at $r = 1$ pc which can produce density profiles that are more consistent with observations.

In typical cosmological simulations, stellar profiles can only be reliably measured up to approximately 1 kpc. Post-processing studies often extrapolate the stellar density profiles using a fixed power law to calculate the density at the influence radius, $\rho_{\rm infl}$, of the binary (Chen et al., 2022b). This is crucial because the hardening timescales are sensitive to $\rho_{\rm infl}$. The extrapolated profiles are defined by the slope parameter γ, which determines how the density $\rho_{\rm extra} \propto r^{-\gamma}$ behaves. For clarity, we note that the extrapolated γ present here is different from $\gamma_{\rm cl}$, the inner slope of the initial density profile of the clusters. Typically, γ values between 1.5 and 2 are used, producing a wide range in the binary's hardening timescales. Here, we investigate the γ parameter obtained from our simulations, which resolve the density down to sub-pc scales. We use the original stellar profiles from ASTRID for the same systems and measure γ based on $\rho_{\rm infl}$ from our SR models and $\rho_{1\ \rm kpc}$ calculated from ASTRID. Examining γ and the extrapolated density profiles in Figure 4.6, we find that for most systems $\gamma \gtrsim 2.0$. For system 6, we find a lower gamma on account of two factors: the ASTRID profile is denser at 1 kpc since the galaxy undergoes more mergers in the full volume simulation, and the NSC is less dense compared to NSCs in other systems. The galaxies in system 6 undergo more mergers than other systems as the MBHs are recorded as merging at a later snapshot in ASTRID compared to the other systems. This results in a lower $\gamma = 1.73$ extrapolated profile. Taking the average across all systems, we find $\gamma_{\rm avg} = 2.04$. For post-processing in cosmological simulations, we recommend using this value to account for the effects of dense nuclear structures.

We also compare the stellar density in our models at $r = 5$ pc with the $\rho - M_{*,\rm gal}$ relationship provided in Pechetti et al. (2020) derived from a volume-limited sample of 27 galaxies. According to Pechetti et al. (2020),

$$\log\left(\frac{\rho_{5\ \rm pc}}{\rm M_\odot pc^{-3}}\right) = (0.61 \pm 0.18)\left(\frac{\rm M_{*,gal}}{10^9 \rm M_\odot}\right) + (2.78 \pm 1.68). \tag{4.7}$$

Using the $M_{*,\rm gal}$ values from our models, we find that $\frac{\rho_{\rm sim}}{\rho_{\rm obs}} \sim 5$ when $M_{*,\rm gal} \leq 5 \times$

$10^8 M_\odot$ but increases to ~ 20 for $M_{*,\rm gal} \sim 10^9 M_\odot$. However, our values are within the limits of the uncertainty in the Pechetti et al. (2020) relation. In future studies, we plan on investigating the effect of choosing $M_{\rm NSC}$ more conservatively to try and better replicate the observed relationship from our simulations. `sys12_LOWM` produces a density profile that is lower than that calculated using equation 4.7 by a factor of 3. This indicates that a realistic NSC mass lie in between our `0` and `LOWM` models.

We compare the effective radii $R_{\rm eff}$ of our formed NSCs to those from Georgiev et al. (2016) in Figure 4.7. The simulated NSCs across various models match observations. For the `FRe` models, we find that $R_{\rm eff}$ is negatively correlated with the mass of the merged NSC $M_{\rm NSC}$. For $M_{\rm NSC} \leq 10^7 M_\odot$, we produce NSCs that are slightly underdense than observed ones whereas for $M_{\rm NSC} \geq 10^7 M_\odot$, we produce NSCs that are denser than the ones observed. Nonetheless, our simulations produce NSCs with $R_{\rm eff}$ that lie within the limits of the observations. The `VRe` counterparts for systems 2 and 10 produce NSCs are on average more consistent with the observed data. While they are less dense than their `FRe` counterparts, they still sink in roughly the same amount of time. This indicates that despite the lower density, NSCs are still efficient at accelerating MBH mergers. The higher resolution `HI` models produce NSCs that have very similar $R_{\rm eff}$ as their `0` counterparts indicating that our results are convergent and robust.

4.5.3 Core scouring

The formation of a hard binary and the subsequent hardening process can lead to a substantial decrease in the central density of the formed cluster (e.g., Merritt and Szell, 2006b; Merritt et al., 2007a). The binary undergoes complicated scattering events with the stellar matter surrounding it leading to a complete or partial ejection of the particle from the system. This leads to a mass deficit in the center of the formed cluster. In order to examine how the density profile evolves as a function of the binary's semi-major axis a, we plot the density profile for `sys10_VRe_0` and `sys12_FRe_0` at several times in Figure 4.8. These models were chosen since they had been evolved for sufficiently long time in the hard binary phase. For the `sys10_VRe_0` model, we find that the formation of the binary and subsequent hardening to $a = r_h/30$ leads to a significant decrease in the central stellar density profile. At $r = 1$ pc, the scouring effect of the binary leads to a $5\times$ decrease in the density. A qualitatively similar observation is made for the `sys12_FRe_0` model. However, we find that the scouring of the density profile is more prominent here as a $\approx 7\times$ decrease in the density profile is observed by the time the binary hardens to $a = r_{\rm h}/10$. `sys12_FRe_0`, in fact, shrinks more slowly than the `sys10_VRe_0` model caused due to the larger binary mass and lower density of the NSC. This implies that a larger amount of mass needs to be ejected for the binary to shrink by the same amount as that in `sys10_FRe_0` model leading to a more significant erosion of the stellar density within the same amount of time.

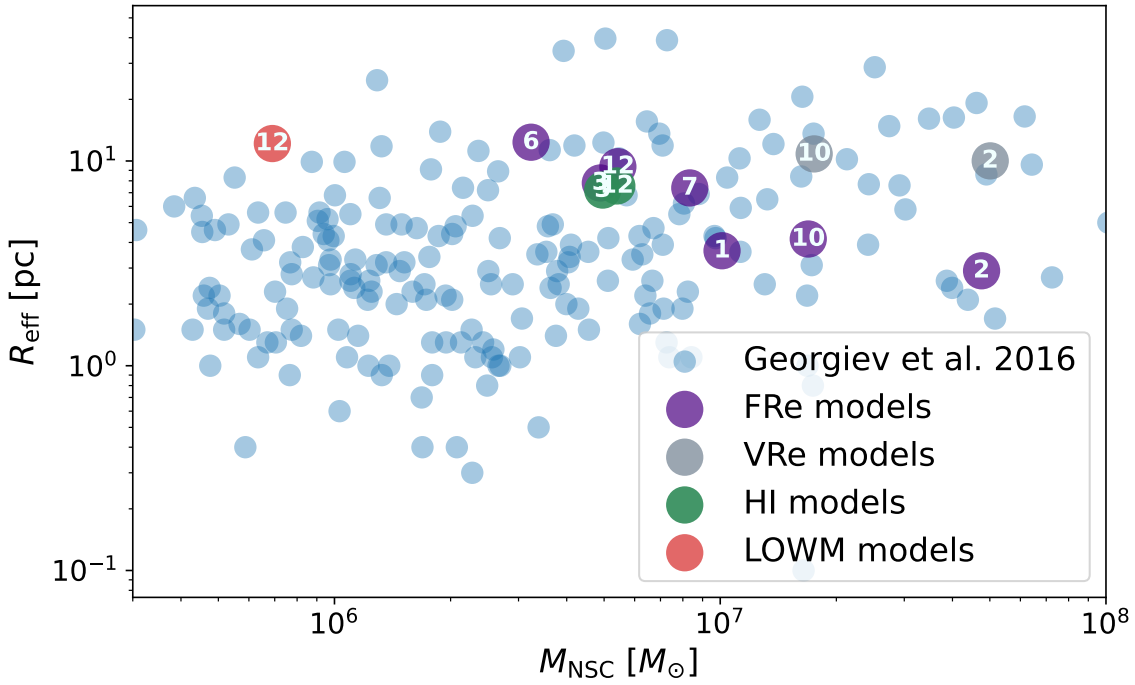

Figure 4.7: The effective radius of the final NSC $R_{\rm eff}$ as a function of the mass of the NSC $M_{\rm NSC}$ upon the merger of the two individual NSCs. The blue circles represent values obtained from Georgiev et al. (2016) while the larger circles are values obtained from our simulations, with the color indicating the model type (similar to those from previous figures), and the numbers indicating the system number. We find that $R_{\rm eff}$ of the merged NSCs are quite consistent with those from observations. For the `FRe` models (purple circles), there is an inverse correlation between $R_{\rm eff}$ and $M_{\rm NSC}$. Thus the NSCs with $M_{\rm NSC} \geq 10^7 M_\odot$ are somewhat denser than the ones that are observed. In the `VRe` models (grey circles), we find that $R_{\rm eff}$ of the formed NSC is about a factor of 5 larger and more consistent with observations. We find that the effect of resolution is subdominant with minor differences in $R_{\rm eff}$ between the original resolution `O` and higher resolution `HI` models (green circles). For the `LOWM` model (red circle), we find that the merged NSC has $R_{\rm eff}$ that is about an order of magnitude larger than NSCs with similar masses. This is a result of the large initial $R_{\rm eff}$, larger than what we expect from observations.

To understand how core scouring varies across all of our models, we compute the relative change in density $\Delta\rho/\rho$ computed at $r = 1$ pc as a function of the binary hardening rate s. $\Delta\rho$ is computed by taking the difference between the initial density profile and that when the binary has hardened to $r_{\rm h}/5$ across all the models. The hardening rate s is defined as

$$s \equiv \frac{d}{dt}\left(\frac{1}{a}\right). \tag{4.8}$$

We calculate s from our simulations by fitting straight lines to the evolution of the inverse semi-major axis $1/a$ as a function of time every 0.3 Myr after the formation of the hard binary and then taking the mean value. Plotting $\Delta\rho/\rho$ as a function of s in Figure 4.9, we find that, a lower hardening rate correlates to a larger relative change in the density profile. For our slowest shrinking model a $\gtrsim 70\%$ decrease in density is observed whereas for the fastest shrinking model, it only leads to a change of $\lesssim 10\%$. This is qualitatively similar to the findings from Merritt et al. (2007a) where the authors find that the relative rate of mass loss is inversely proportional to the hardening rate. While our slower shrinking systems form a large flat core by the time they shrink to GW dominated stage, whereas for the denser systems, a cusp, albeit shallower, might still remain. Core scouring can also play a role in lowering the density such that our simulated systems match observations better. For example, for the `sys12_FRe_0` model, we observe that $\rho_{5\,\rm pc}$ reduces by a factor of 2 by the time the binary has hardened to $r_{\rm h}/10$. This reduction in density leads to more consistency between the observed Pechetti et al. (2020) relation and our simulated systems.

One can also calculate the total mass deficit by taking the difference between the final and initial density profiles and then integrating over some radius. We follow this procedure for a variety of radii and calculate the mass deficit $M_{\rm def}$ by finding where the curves peak. To be consistent among models, we calculate the mass deficit when the binaries have hardened to $a_{\rm h}/5$. We find that the deficit is quite independent of the density profile used. For the bulk of our models we find that $M_{\rm def} \approx 1.4 \times M_{\rm bin}$ while for one of our models it peaks at $1.81 M_{\rm bin}$. This is qualitatively similar to the results from Khan and Holley-Bockelmann (2021) although they find higher $M_{\rm def}$ in their models. Incidentally, similar to Khan and Holley-Bockelmann (2021) we find that the $M_{\rm def}$ does not peak at $r_{\rm infl}$ but rather at distances greater than $r_{\rm infl}$. This indicates that stars beyond $r_{\rm infl}$ also contribute significantly to the hardening process, necessitating the inclusion of the bulge and DM halo.

4.5.4 Evolution of the hard binary

After the formation of a hard binary, we follow the evolution down to $a \lesssim r_{\rm h}/10$ for all models other than the `LOWM` model. In this stage, the binary hardens via three-body interactions with the surrounding material. Under the full-loss cone approximation, we expect the MBH binary to harden at a fixed rate. Accordingly, we determine the hardening rate $s \equiv \frac{d}{dt}\left(\frac{1}{a}\right)$ as mentioned in the previous section and study its evolution

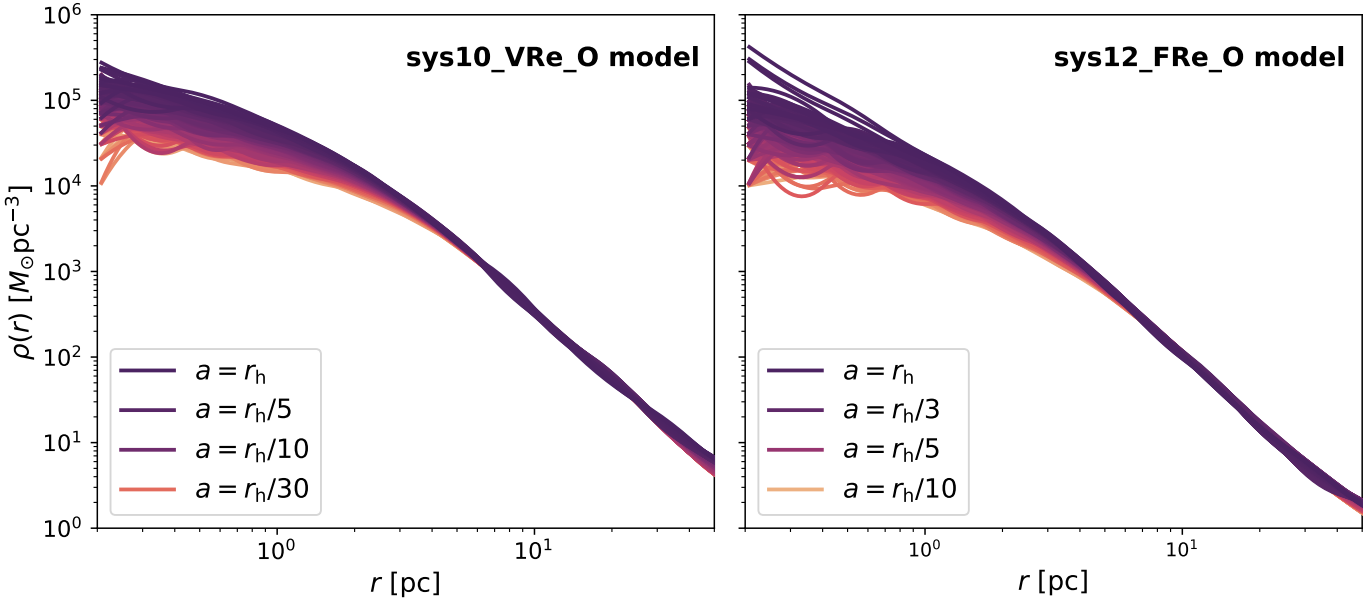

Figure 4.8: Evolution of the stellar density $\rho(r)$ as a function of the distance from the MBH binary r over time as the MBH binary hardens. The binary hardens by ejecting particles from the cusp leading to the slow scouring of the initial cusp. Left: evolution of the density profile in the `sys10_VRe_0` model where the binary is evolved for a total duration of 26 Myr in the hard binary phase. Right: same but for the `sys12_FRe_0` model and the evolution is followed for 33 Myr. The mass of the binary in this model is larger than that in `sys10_VRe_0` while the NSC mass is lower. This leads to a slower hardening and a larger erosion of the cusp.

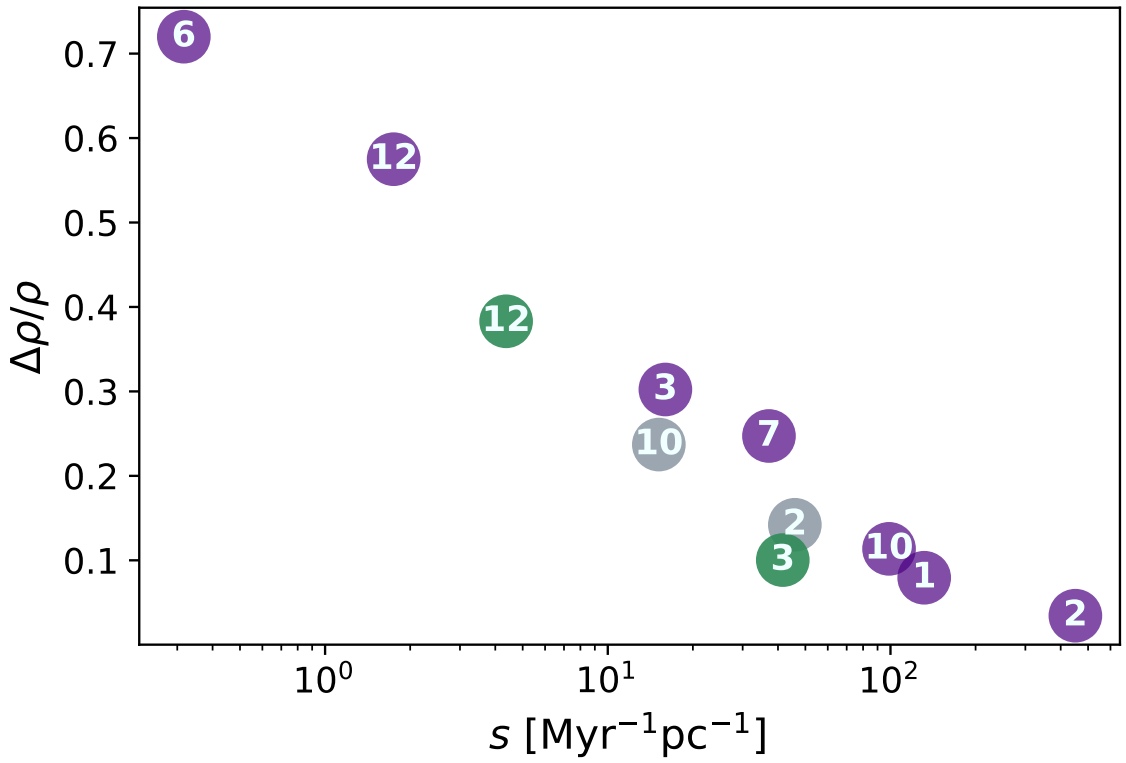

Figure 4.9: Relative change in the density $\Delta\rho/\rho$ at $r = 1$ pc as a function of the binary hardening rate s. The circle number and color indicate the system number and model type respectively, as in Figure 4.7. The values are measured when the binary has hardened to $a = r_{\rm h}/5$. There is an inverse correlation between $\Delta\rho/\rho$ and s. A faster hardening implies a slower rate of erosion of the cusp as the initial cusp density is higher and/or the mass of the binary is lower, similar to findings from Merritt et al. (2007a).

over time. We find that the hardening rate remains roughly constant, as expected, indicating that the binary loss-cone is filled by particles on centrophillic orbits. This is primarily caused due to the non-spherical geometry of the merged product which produces a torquing effect, leading to replenishment of the loss-cone.

Quinlan (1996); Sesana et al. (2006) provide a description of the energy loss during the hardening process. The hardening rate can be represented as

$$\frac{d}{dt}\left(\frac{1}{a}\right) = \frac{GH\rho}{\sigma} \tag{4.9}$$

where σ is the velocity dispersion, and H is a dimensionless hardening parameter that is almost constant once the binary is sufficiently hard. Sesana et al. (2006) performed three-body scattering experiments to numerically determine the hardening rate and found that $H \approx 16 - 20$. To determine the value of H from our simulations, we measure ρ and σ at the influence radius $r_{\rm infl}$ of the binary as done in Sesana and Khan (2015). We find that $H \approx 10$ for all our simulations. This is about 40% lower than the theoretically predicted value of 16 but consistent with the findings from other N-body simulations (e.g., Fastidio et al., 2024). This implies that although loss-cone interactions drive the binary towards GW stage, the binary does not harden in the full loss-cone regime (Vasiliev, 2019b). We also compare the effect of resolution on the the value of H by comparing the calculated values from the `O` and the `HI` models. For `sys3_FRe_O`, we find that $H = 9.03$ whereas for `sys3_FRe_HI`, we find $H = 9.48$, only about 5% larger. Similarly, for `sys12_FRe_O`, we find $H = 11$ while for `sys12_FRe_HI` we find $H = 12.9$, about 17% larger. This indicates that the dimensionless hardening parameter is only weakly sensitive to the resolution and is convergent in our `FRe` models. We do note that slight variations between ρ and σ among the `O` and `HI` models are present which affect the overall hardening rate s. Owing to the slightly higher density in the `HI` models, we find that s is larger by a factor of 2 in those models compared to the `O` models.

Given the diverse environments in our simulations, we aim to understand how the hardening rate s compares across different models. In Figure 4.10 we compare s as a function of $\rho_{\rm infl}$ and find a strong correlation between the hardening rate and the density at the influence radius across all our models. The model `sys2_FRe_O` hardens the fastest with $s = 450.8 {\rm Myr}^{-1}{\rm pc}^{-1}$ which is unsurprising given its $\rho_{\rm infl} = 2.1 \times 10^6 M_\odot {\rm pc}^{-3}$. Similarly, we find $s = 0.04 {\rm Myr}^{-1}{\rm pc}^{-1}$ for `sys12_LOWM` given the very $\rho_{\rm infl} = 11 M_\odot {\rm pc}^{-3}$. Zhou et al. (2024) find a hardening rate of $\approx 0.02 {\rm Myr}^{-1}{\rm pc}^{-1}$ for system 12 from their simulations. Despite substantial methodological differences, the similarity in hardening rates between the two sets of simulations underscores the robustness of our results. Recently Khan et al. (2024) performed simulations of IMBH binary mergers in non-nucleated dwarfs. Their `D1.5C` model is has similar $\rho_{\rm infl}$ as our `sys12_LOWM` model and we find hardening rates comparable to their results.

For models with intermediate $\rho_{\rm infl}$ such as `sys7_FRe_O`, $s = 37.1 {\rm Myr}^{-1}{\rm pc}^{-1}$. We find that the relationship between s and $\rho_{\rm infl}$ can be described by a simple power-law.

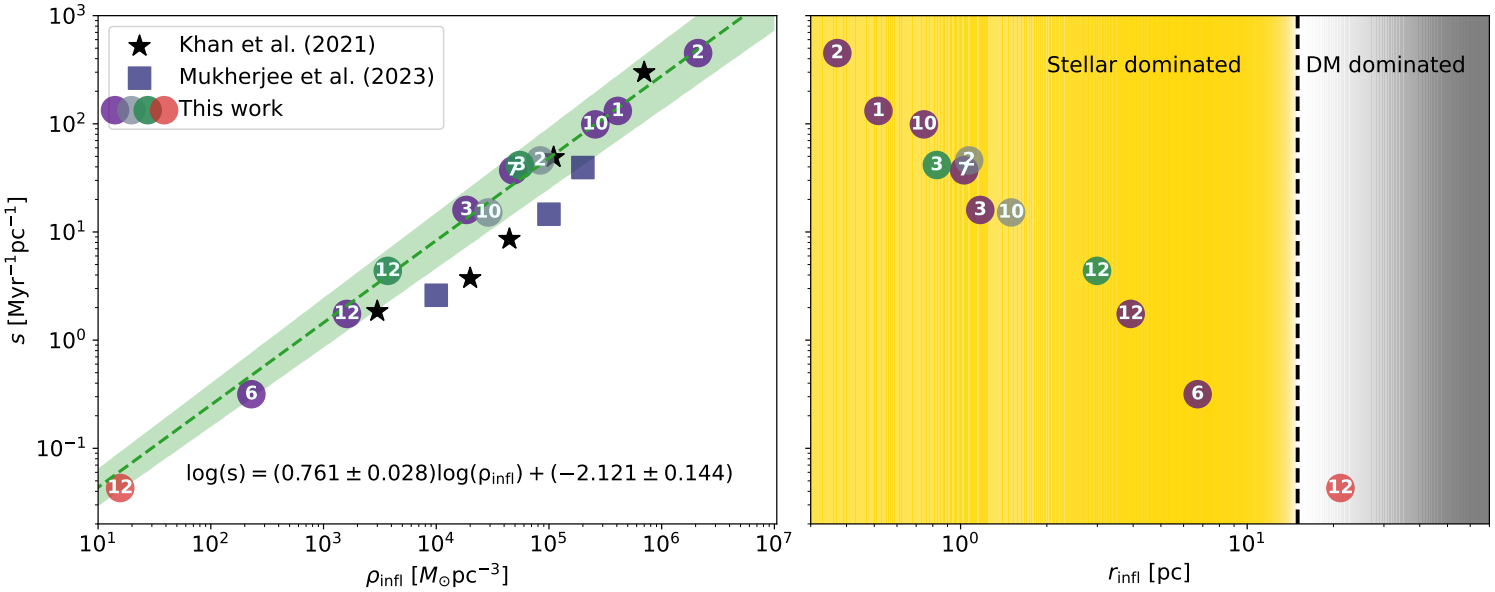

Figure 4.10: Left: The hardening rate s as a function of the density at influence radius $\rho_{\rm infl}$. We find that s is strongly correlated to $\rho_{\rm infl}$ and can be described by a power law (green line). The shaded region represents the uncertainty of our fit. Theoretically, we expect s to be a linear function of $\rho_{\rm infl}/\sigma_{\rm infl}$. In our simulations $\sigma_{\rm infl}$ does not change a lot across our models, allowing us to obtain a one-parameter fit between s and $\rho_{\rm infl}$. The obtained values are consistent with those obtained from previous simulations with NSCs including Khan and Holley-Bockelmann (2021) and Mukherjee et al. (2023) differences of factors of 2-3 at most. Right: s but as a function of the influence radius $r_{\rm infl}$. Here we observe a negative correlation between s and $r_{\rm infl}$ which is consistent with our expectations as a smaller $r_{\rm infl}$ implies a higher $\rho_{\rm infl}$. We also find that in all but one case, the binary hardens primarily via three-body interactions with stars from the NSC and the bulge. However, for `sys12_LOWM`, $r_{\rm infl} \approx 25.2$ pc and we find that DM contributes twice as much as the stars in the hardening rate. The black dashed vertical line represents where the contribution due to DM and stars is equal.

We fit the values obtained from our simulations and find the relationship to be

$$\log_{10}\left(\frac{\mathrm{s}}{\mathrm{Myr^{-1}pc^{-1}}}\right) = (0.761 \pm 0.028)\log_{10}\left(\frac{\rho_{\rm infl}}{\mathrm{M_\odot pc^{-3}}}\right) + (-2.121 \pm 0.144). \tag{4.10}$$

Interestingly, Khan et al. (2024) in simulations of IMBH binaries in non-nucleated dwarfs, find a linear relationship between $\rho_{\rm infl}$ and s in the low $\rho_{\rm infl}$ regime. We additionally compare our values to those obtained from previous simulations with NSCs and find that for similar $\rho_{\rm infl}$, our hardening rates are consistent with those obtained from previous works. Khan and Holley-Bockelmann (2021) study the hardening of IMBH binaries in a variety of nucleated dwarf galaxies, taking into account the effect of the bulge and NSC. We find that for $\rho_{\rm infl} \geq 10^5 M_\odot \mathrm{pc}^{-3}$, the hardening rates predicted by our simulations are fully in agreement with those from Khan and Holley-Bockelmann (2021). For example, the authors find that $s = 49 \mathrm{Myr}^{-1}\mathrm{pc}^{-1}$ for NGC 404 where $\rho_{\rm infl} = 1.1 \times 10^5 M_\odot \mathrm{pc}^{-3}$. For similar $\rho_{\rm infl}$ such as in `sys2_VRe_0` where $\rho_{\rm infl} = 8.3 \times 10^4 M_\odot \mathrm{pc}^{-3}$, we find $s = 45.9 \mathrm{Myr}^{-1}\mathrm{pc}^{-1}$. At lower densities, there are differences of at most factors of $2-4$. For NGC 5206 Khan and Holley-Bockelmann (2021) find $\rho_{\rm infl} = 2 \times 10^4 M_\odot \mathrm{pc}^{-3}$ and $s = 3.73 \mathrm{Myr}^{-1}\mathrm{pc}^{-1}$, whereas in `sys3_FRe_0`, we obtain $s = 15 \mathrm{Myr}^{-1}\mathrm{pc}^{-1}$ for a similar $\rho_{\rm infl}$. These differences may be caused due to the differences in initial conditions and resolution. We also compare our results to those from Mukherjee et al. (2023) where the authors studied the effect of MBH mergers in mass-segregated NSCs and find that our results are quite consistent especially for higher $\rho_{\rm infl}$. The differences at lower $\rho_{\rm infl}$ are, again, at most factors of 2-3. These could be caused due to differences in initial conditions, or due to the lack of a stellar bulge in Mukherjee et al. (2023) as Khan and Holley-Bockelmann (2021) showed that a non-trivial amount of the hardening is caused to the bulge stars, especially for equal mass ratio binaries. Theoretically we expect a linear relationship between s and $\frac{\rho_{\rm infl}}{\sigma_{\rm infl}}$. We fit s versus $\frac{\rho_{\rm infl}}{\sigma_{\rm infl}}$ and find that $s \propto \left(\frac{\rho_{\rm infl}}{\sigma_{\rm infl}}\right)^{0.96 \pm 0.04}$ in agreement with theory.

Since we have both stars and DM in our simulations, an interesting question one might ask is which component contributes more to hardening. Examining Figure 4.10 where we compare the s as a function of the influence radius $r_{\rm infl}$, we find that most of our systems harden by scattering with stars. The hardening rate is negatively correlated to $r_{\rm infl}$ implying that binaries with large s harden by scattering with particles closer to the binary. Examining where $M_{\rm def}$ peaks along with $r_{\rm infl}$ values, we find that stars mainly from the cluster contribute more to the hardening of the binary for systems with large s. However, in particularly low density environments such as `sys6_FRe_0` and `sys12_LOWM`, the situation is different. For the latter, we find that $r_{\rm infl} \approx 25.2$ pc. Here $\rho_{\rm DM,infl} = 11 M_\odot \mathrm{pc}^{-3}$ while $\rho_{*,\rm infl} = 2 M_\odot \mathrm{pc}^{-3}$. Comparing the hardening rates of different components using equation 4.9, we find that the DM contributes $2\times$ as much to the hardening rate as stars. Therefore, in DM dominated dwarf galaxies lacking dense NSCs, the effect of the DM halo cannot be neglected

as the binary primarily hardens by scattering with DM, consistent with the findings from Partmann et al. (2023). Therefore, we expect DM to dominate the hardening in non-nucleated dwarfs and the effect of the DM halo cannot be neglected. The addition of the DM halo could potentially alleviate the issue of non-mergers of IMBH binaries in non-nucleated dwarf galaxies as noticed by Khan et al. (2024).

We also study the evolution of the eccentricity during this stage. The MBHs are initially on highly eccentric orbits given that Chen et al. (2024) find that a combination of high eccentricity and central stellar density is essential for the MBHs to sink sufficiently. Before the MBHs become bound, the orbital eccentricity using Keplerian elements is undefined for the pair. Thus, in order to measure the unbound eccentricity of the pair in the global potential, we use the pericenter and apocenter distances. The **unbound** orbital eccentricity is measured as

$$e = \frac{r_{\rm apo} - r_{\rm peri}}{r_{\rm apo} + r_{\rm peri}} \tag{4.11}$$

where $r_{\rm apo}$ and $r_{\rm peri}$ are the relative apocenter and pericenter distances. Initially, the MBH pairs have unbound orbital eccentricities of $0.75 - 0.95$ across all systems. The mean eccentricity initially is 0.86 with a standard deviation of 0.06. During the DF phase, we notice a systematic decrease in the eccentricity of the pairs, similar to Gualandris et al. (2022). By the time the MBHs are 30 pc apart, the mean eccentricity reduces to 0.67. This is caused due to the the circularizing effect of DF (Gualandris et al., 2022). However, we notice that this phase introduces a lot of scatter in our eccentricity values, increasing the standard deviation to 0.2. This is unsurprising given stochasticity introduced due to our finite resolution. Nasim et al. (2020) find that this scatter is introduced by the merger process and affects binaries with $e > 0.9$ more substantially. This makes inferring bound binary eccentricities as a function of the initial eccentricity challenging.

To understand how the eccentricity of the bound binary behaves as a function of the unbound eccentricity, we measure the unbound eccentricity $e_{\rm unbound}$ when the MBHs are separated by 30 pc and compare it to the bound eccentricity e_b in Figure 4.11. The latter is measured according to equation 6 from (Gualandris et al., 2022) where the eccentricity is measured at a separation $r_{\rm bound}$ where

$$M_{\rm enc}(r_{\rm bound}) = 0.1(M_p + M_s) \tag{4.12}$$

For the former, we choose to measure the unbound eccentricity at $\Delta r = 30$ pc as we find that the initial eccentricity produces a wide range of e_b erasing any correlations. Similar findings were also noted by Fastidio et al. (2024) where the authors found a wide range of e_b when the initial eccentricity e_0 was greater than 0.9. Examining Figure 4.11, we notice that $e_{\rm unbound}$ is correlated to e_b but there is significant scatter in e_b when the $e_{\rm unbound}$ is high. In `sys10_FRe_0` we find both $e_{\rm unbound}$ and e_b are greater than 0.9, the situation for `sys2_FRe_0` and `sys2_VRe_0` are different where the pair has $e_{\rm unbound} \approx 0.9$ but e_b is significantly lower. We also note that there

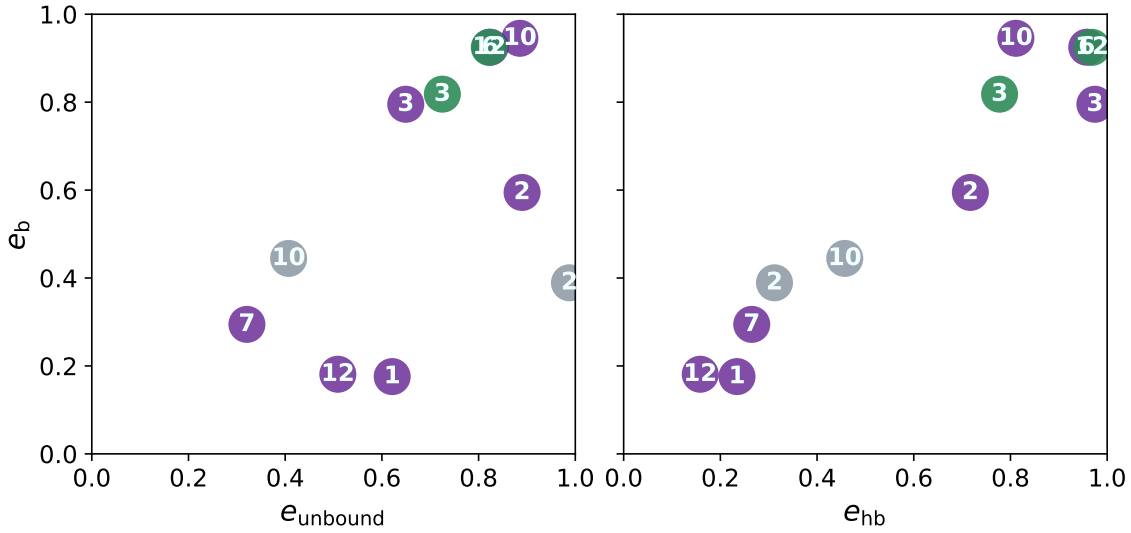

Figure 4.11: Evolution of eccentricity across all our models. Left: The bound eccentricity e_b as a function of the unbound orbital eccentricity as calculated using equation 4.11 of the pair measured at a separation of $\Delta r = 30$ pc. e_b is measured before the black holes form a hard binary and when they are separated by a distance r_{bound} as defined in equation 4.12. Although the initial orbital eccentricity of all the black holes are 0.85-0.95, $e_{unbound}$ shows a wide scatter with the mean around 0.75. e_b is correlated to $e_{unbound}$ but stochasticity associated with higher $e_{unbound}$ produces a wide variety of e_b. Right: e_b as a function of the eccentricity at the hard binary separation e_{hb}. We notice that a higher e_b implies a higher e_{hb}. During the hardening stage, we find `sys3_FRe_O`, `sys12_FRe_HI` and `sys6_FRe_O` are able to achieve almost radial orbits. The differences among `HI` and `O` models persist in this stage with `sys12_FRe_O` containing a binary with an eccentricity of 0.15 while that in `sys12_FRe_HI` is 0.95.

are differences between the formed binary among our `0` and `HI` models. While both `sys3_FRe_0` and `sys3_FRe_HI` have similar $e_{\rm unbound}$ and e_b, the same cannot be said for the `sys12_FRe_0` and `sys12_FRe_HI`. In `sys12_FRe_0`, the bound binary has a low eccentricity of 0.2 whereas in the `sys12_FRe_HI`, the bound binary has an eccentricity of 0.9. Our investigations reveal that NSCs in `sys12_FRe_HI` merge somewhat earlier when the unbound orbital eccentricity is still large. In the `0` model, the pair undergoes circularization before the NSCs merge leading to a lower bound eccentricity. This is an unfortunate effect of stochasticity, especially for initially eccentric binaries, making inferences of bound binary eccentricity from the unbound eccentricity challenging , which is consistent with Rawlings et al. (2023). On the other hand, examining the right panel of Figure 4.11, we find there is a tighter correlation between the hard binary eccentricity $e_{\rm hb}$ and e_b. For binaries with $e_b < 0.6$, we find that $e_{\rm hb}$ remains nearly the same as e_b, whereas for higher e_b, we notice a slight increase in the eccentricity. Longer evolution reveals that binaries with high eccentricities grow their eccentricity over time in the hard binary phase, leading faster mergers via the emission of GWs.

The evolution of eccentricity is also affected by the slope of the cusp in the original NSCs. Gualandris et al. (2022) find that for denser cusps, the eccentricity of the formed binary is systematically lower. Similarly, due to collisional relaxation under a two-component mass species, Mukherjee et al. (2023) find that relaxed systems which have denser cusps, form binaries with lower eccentricities. In our models, the density profile of the NSCs initially follows a shallow $\gamma = 0.5$ cusp. Thus eccentricities of our formed binaries can be considered to be an upper limit. In future studies, we plan on varying the initial cusp of the NSCs, and/or incorporating a mass function to understand how that affects our results.

4.6 Discussion

4.6.1 Gravitational wave timescales

We use a semi-analytic method similar to that used in previous studies (e.g., Gualandris and Merritt, 2012b; Gualandris et al., 2022) to calculate the GW merger timescale $\tau_{\rm merge}$ of our models. We assume that the hardening rate s remains constant and that there is no change in eccentricity. Under this assumption, we can write the change in semi-major axis a due to three-body hardening over time as

$$\left.\frac{da}{dt}\right|_{*} = -s^2 a. \tag{4.13}$$

Then, the overall evolution of a can be written as follows:

$$\frac{da}{dt} = \left.\frac{da}{dt}\right|_{GW} + \left.\frac{da}{dt}\right|_{*} \tag{4.14}$$

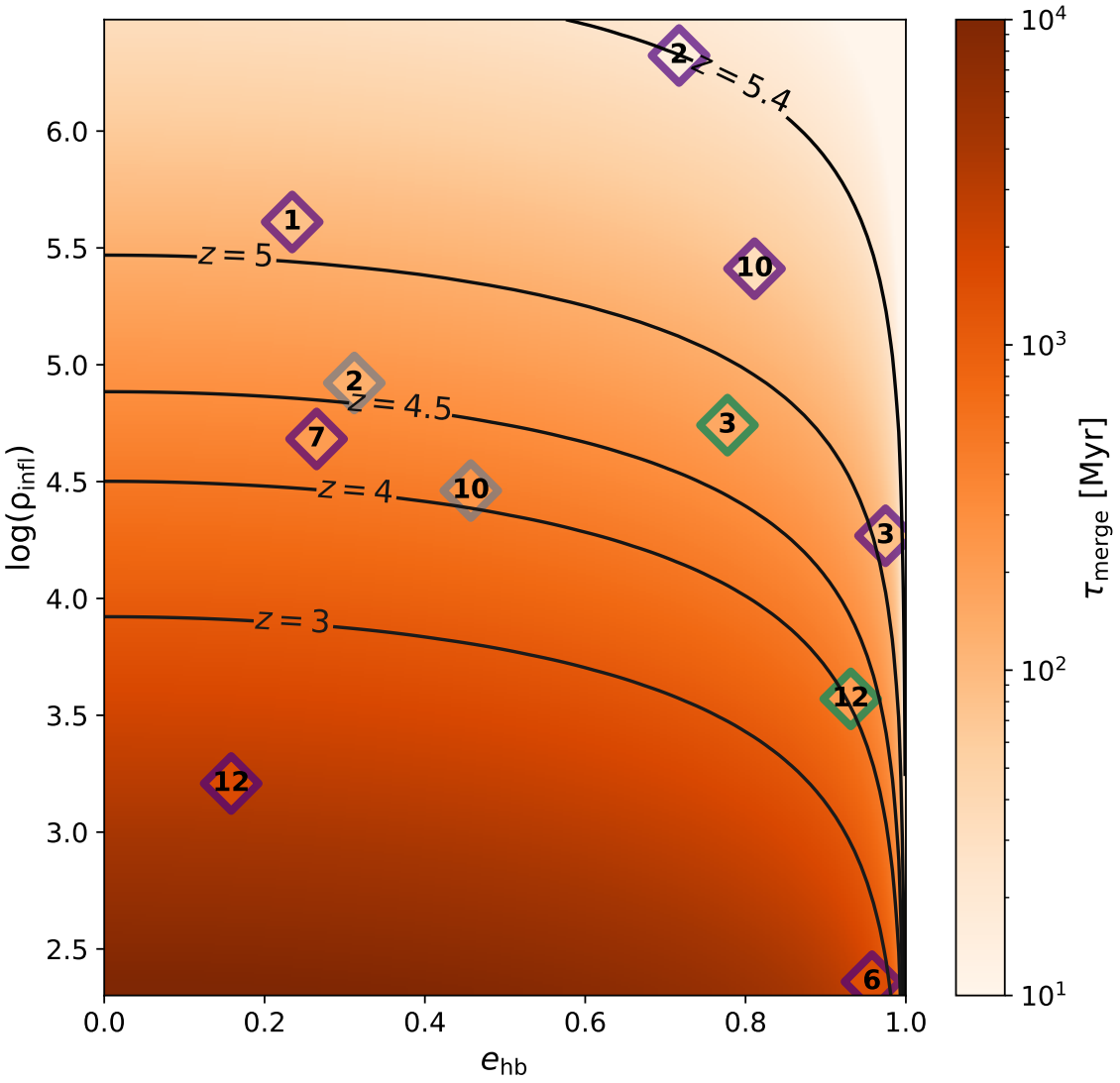

Figure 4.12: The merger timescale $\tau_{\rm merge}$ as a function of the density at influence radius $\rho_{\rm infl}$ and the eccentricity at hard binary radius $e_{\rm hb}$. The diamonds represent $\tau_{\rm merge}$ of our different models with the color and number indicating type of model and system number respectively, similar to that used in previous figures. Using $\rho_{\rm infl}$, we calculate the hardening rate s using equation 4.10 which allows us to approximate the merger time following equations 13 - 19. The contours are drawn approximating that the galaxy merger process begins at $z = 9$ and it takes 500 Myr for the MBHs to sink to hard binary radius. Under this approximation, we find that binaries that harden in environments with $\rho_{\rm infl} > 10^{4.5} M_{\odot}{\rm pc}^{-3}$ merge by $z = 4$. Eccentric binaries merge faster and can lead to high redshift mergers even in somewhat lower density environments. Nevertheless, we find that seed black holes require a high stellar density environment to merge at high redshifts, suggesting that most mergers that happen at high redshifts occur in NSC dominated environments. We find in 8 of the 12 models we simulate, the MBH binary merges before $z = 4$.

We use Peters (1964c) equations to determine $\left.\frac{da}{dt}\right|_{\rm GW}$ and $\left.\frac{de}{dt}\right|_{\rm GW}$. According to Peters (1964c),

$$\left.\frac{da}{dt}\right|_{GW} = -\frac{64}{5}\beta\frac{F(e)}{a^3} \tag{4.15}$$

$$\left.\frac{de}{dt}\right|_{GW} = -\frac{304}{15}\beta\frac{eG(e)}{a^4} \tag{4.16}$$

where

$$\beta = \frac{G^3}{c^5}\left(M_p M_s \left(M_p + M_s\right)\right) \tag{4.17}$$

$$G(e) = (1-e^2)^{-5/2}\left(1 + \frac{121}{304}e^2\right) \tag{4.18}$$

$$F(e) = (1-e^2)^{-7/2}\left(1 + \frac{73}{24}e^2 + \frac{37}{96}e^4\right). \tag{4.19}$$

We solve the set of coupled differential equations assuming that the initial semi-major axis and eccentricity for the differential equations are $a_{\rm hb}$ and $e_{\rm hb}$ respectively. Equation 4.10 allows us to to calculate $\tau_{\rm merge}$ as a function of $\rho_{\rm infl}$ by fixing $e_{\rm hb}$. For simplicity, we calculate $\tau_{\rm merge}$ as a function of $\rho_{\rm infl}$ and $e_{\rm hb}$ assuming $M_p = 10^5 M_\odot$ and $q = 0.5$. To simplify our calculations, we estimate $\tau_{\rm merge}$ by starting from $a = 0.02$ pc. This is roughly close to the mean of $a_{\rm hb}$ across the models considered in this study. We present $\tau_{\rm merge}$ as a function of $\rho_{\rm infl}$ and $e_{\rm hb}$ under these set of assumptions in Figure 4.12 with merger timescales from our models overlaid. We note that $\tau_{\rm merge}$ is measured from the time the binary reaches $a_{\rm hb}$. We also draw contours assuming that the galaxies merge at $z = 9$, and that the MBHs take 500 Myr to sink to form a hard binary. Examining Figure 4.12 we find that mergers at $z > 4$ require $\rho_{\rm infl} > 10^{4.3} M_\odot {\rm pc}^{-3}$ even for moderately eccentric binaries with $e_{\rm hb} = 0.8$, indicating that high redshift MBH seed binaries with masses $\lesssim 2 \times 10^5 M_\odot$ merge only in NSC dominated environments. Among our models, binaries embedded in high density NSCs such as `sys2_FRe_0` or those with $e_{\rm hb} > 0.9$ such as `sys12_FRe_HI` merge before $z = 4$. One caveat of this calculation is that we do not consider the growth of eccentricity during the hard binary phase which may lead to faster mergers. Additionally, recent studies have shown that rotation in galactic nuclei can affect the eccentricity of the binary at formation (Rasskazov and Merritt, 2017). Nuclei counter-rotating with respect to the binary can lead to an almost radial binary at formation accelerating merger timescales by factors up to 10 whereas nuclei co-rotating lead to systematically lower eccentricities (e.g., Mirza et al., 2017). Nevertheless, our results strongly suggest that high-z mergers happen in dense stellar environments that are only possible in NSCs. Additionally, our results coupled with those from MAGICS-II suggest that naked seed MBHs do not even make it to the galactic center efficiently.

Zhou et al. (2024) find that in only one of the six systems simulated the MBHs seeds form a binary. This is attributed to the presence of the nuclei around both MBHs. Even a low mass NSC such as that present in our `LOWM` model helps the MBHs sink efficiently. Such stellar structures around the seeds were neglected in previous studies (Ma et al., 2021; Partmann et al., 2023) and the authors found that the sinking process of seeds, especially low mass ones, is highly erratic. Our study in association with Zhou et al. (2024) corroborates this but finds that in the presence of extended dense stellar systems such as NSCs, the seed sinking problem can be resolved. Low mass seeds are more likely to be formed in such high density environments indicating that future studies cannot neglect the effect of NSCs.

4.6.2 Implications for future detection

What do our results imply for future detection? Using the results from our simulations, we can estimate the effect of different environments on the merger rate. We use the mergers recorded in the `ASTRID` simulation (Ni et al., 2022; Bird et al., 2022) and select a subset where the MBH masses are: $M_p \leq 5.5 \times 10^5 M_\odot$ and $M_s \leq 2 \times 10^5$. We record the redshift of the mergers and add a delay time based on the hardening timescale of the binary by calculating s from $\rho_{\rm infl}$ using equation 4.10. For simplicity we assume $e_{\rm hb} = 0.8$ and $a_{\rm hb} = 0.02$ pc. This calculation does not take into account the time delay due to DF as the mergers in `ASTRID` are recorded when the MBHs are $\sim$ kpc apart from each other. As noted in previous sections and in Zhou et al. (2024), the delay time due to DF dominated sinking can be of the order of ~ 100 Myr for MBH seeds surrounded by NSCs whereas for naked MBH seeds it is larger. Nevertheless, our calculation provides an upper limit on merger rate in a particular environment. Following Chen et al. (2022b), once the delay time has been added, we calculate the merger rate as a function of the redshift $\frac{dN}{dzdt}$ as

$$\frac{dN}{dzdt} = \frac{d^2n(z)}{dzdV_c}\frac{dz}{dt}\frac{dV_c}{dz}\frac{1}{1+z} \tag{4.20}$$

where V_c is the comoving volume. To calculate $\frac{d^2n(z)}{dzdV_c}$ from the simulation, we approximate it as

$$\frac{d^2n(z)}{dzdV_c} = \frac{N(z)}{\Delta z V_{\rm sim}} \tag{4.21}$$

where Δz is the width of the redshift bin, $N(z)$ is the total number of mergers in that particular bin, and $V_{\rm sim}$ is the volume of the simulation. For `ASTRID`, $V_{\rm sim} = (250{\rm Mpc/h})^3$. We add the delay time based on $\rho_{\rm infl} = 10^3, 10^4,$ and $10^5 {\rm M_\odot pc^{-3}}$ and present the merger rate in Figure 4.13. We find that at $z = 4$, $\frac{dN}{dzdt} = 0.63{\rm yr}^{-1}$ when $\rho_{\rm infl} = 10^5 M_\odot$, $\frac{dN}{dzdt} = 0.33{\rm yr}^{-1}$ when $\rho_{\rm infl} = 10^4 M_\odot$, and $\frac{dN}{dzdt} = 10^{-3}{\rm yr}^{-1}$ when $\rho_{\rm infl} = 10^3 M_\odot$. This suggests that most, if not all, detectable high redshift seed mergers happen in dense stellar environments, similar to the conclusions found by Khan et al. (2024). In non-nucleated bulges stellar density is typically $\lesssim 10^3 M_\odot$ which results in

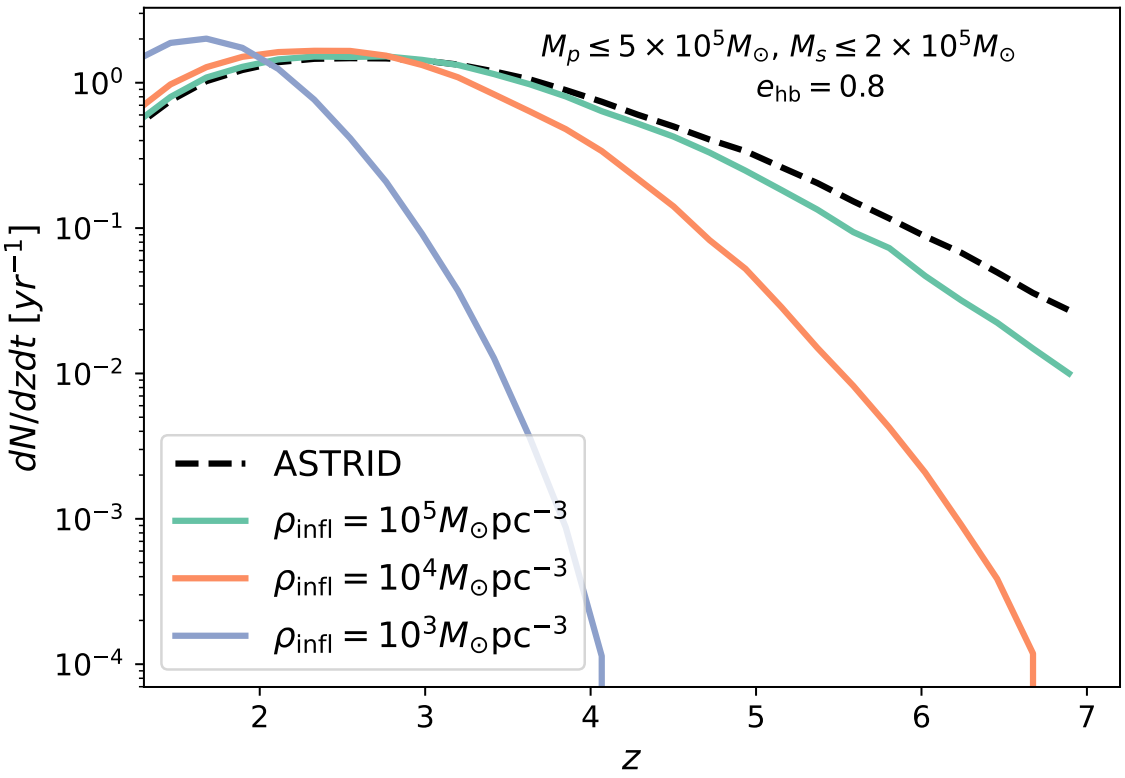

Figure 4.13: The merger rate per year $\frac{dN}{dzdt}$ as function of the redshift z. We extract the merger data from the `ASTRID` simulation for MBH pairs where $M_p \leq 5 \times 10^5$ and $M_s \leq 2 \times 10^5$, and add a hardening delay time based on different $\rho_{\rm infl}$ using equation 4.10 and the semi-analytic approach described in Section 4.6.1. For simplicity, we assume that the MBH binaries have hard binary eccentricity $e_{\rm hb} = 0.8$. We find dense stellar environments dominate high-z mergers at $z > 4$ strongly suggesting that NSCs are the dominant channel for MBH seed mergers at high redshift. In bulge-like environments with $\rho_{\rm infl} = 10^3 M_\odot$, the merger rate is much lower by factors of 600 compared to $\rho_{\rm infl} = 10^5 M_\odot$ at $z = 4$ and peaks at $z = 1.67$. Our calculation assumes that all mergers take place in nucleated environments, which may not hold in reality. However, for similar mass galaxies, Neumayer et al. (2020a) report that 60% may be nucleated. This would imply that NSC dominated evolution would be the primary channel for seed MBH mergers.

three orders of magnitude fewer mergers at $z = 4$. We also find that when $\rho_{\rm infl} = 10^5 M_\odot$, the merger rate peaks at $z = 2.76$ with a rate of $1.52{\rm yr}^{-1}$, while for $\rho_{\rm infl} = 10^4 M_\odot$ the rate peaks at $z = 2.33$ at a rate of $1.66{\rm yr}^{-1}$. In low density environments such as $\rho_{\rm infl}$, we find that mergers peak later around $z = 1.67$ at a rate of $2.01{\rm yr}^{-1}$. Our calculations are approximate as we do not consider the fraction of galaxies that have nucleation and additional effects such as hierarchical three-body hardening which can also accelerate mergers. For the former, observations suggest anywhere between 20% to 60% of the galaxies with $M_* = 10^8 - 10^9 M_\odot$ may be nucleated (Neumayer et al., 2020a). We will present a more detailed analysis of mergers taking into account the DF time delay, evolution of eccentricity, and nucleation fraction in a future work. Despite this, the implications of our work are robust and we find that the seed sinking problem can be resolved if the MBH seeds are embedded in NSCs. Additonally, our results suggest that high-redshift MBH seeds are an important source of GWs for future GW missions such as LISA which should be able to detect a few such mergers. This would be crucial to further constrain MBH seeding models.

4.7 Conclusions

MBH seeds are one of the most important sources of gravitational waves (GWs) that will be detectable by future low-frequency GW detectors like LISA. The merger rates of MBH seeds at high redshifts can shed light on, and constrain seeding models. Previous studies have shown that the process by which MBH seeds sink to the centers of galaxies is inefficient, a phenomenon known as the seed-sinking problem. However, these analyses overlook the influence of extended stellar systems, such as nuclear star clusters (NSCs) surrounding the seeds. Combined with the findings of MAGICS-II, which highlight the importance of these stellar systems in aiding the sinking process, our results strongly indicate that NSCs are a critical factor in accelerating both the sinking and merger of MBH seeds.

In this study, we utilize high-resolution N-body re-simulations of MBH seeds embedded in NSCs within high-redshift dwarf galaxies, employing up to 10^7 particles using the FMM-based code `Taichi`. Our objective is to examine how NSCs influence the sinking and merger dynamics of MBH seed binaries, motivated by the findings of MAGICS-II, which demonstrated that only MBH seeds retaining extended stellar systems around them are capable of sinking and forming binaries in high-z dwarf galaxies. Building on MAGICS-I and MAGICS-II, we analyze the detailed dynamics of MBH seeds within NSCs orbiting the remnant host galaxy. For these models, we adopt a simple prescription for determining NSC masses, following MAGICS-I and MAGICS-II, where the total mass of both NSCs is set equal to the mass enclosed within the central 100 pc of the nucleus of the merged galaxy. The resulting NSCs in our simulations have masses ranging from $3 \times 10^5 M_\odot$ to $3 \times 10^7 M_\odot$ and are spherically symmetric, isotropic, and follow the Dehnen (1993) density profile with a shallow cusp near the MBH.

The main conclusions of our work can be summarized as follows:

- MBHs that are embedded in NSCs of all masses and sizes investigated in this study form a hard binary. NSCs assist MBH seeds in sinking to sub-pc scale. For more massive and denser NSCs, this process is extremely efficient, with MBHs sinking to $\lesssim 0.1$ pc within 15-20 Myr from the start of our simulation. The overall sinking times are, on average, 20% faster than those found in MAGICS-I. When the NSCs reach a separation of ≈ 50 pc , we observe an accelerated decline in separation due to tidal interactions between the NSCs, leading to the formation of a hard MBH binary in $\lesssim 0.5$ Myr. Notably, this effect is absent when the MBH masses are boosted by the same amount, indicating that tidal effects of NSCs are crucial in helping MBHs form bound binaries. In the absence of NSCs surrounding either MBH seed, we find that the sinking process is inefficient. Our results converge upon increasing the resolution and we find that higher resolution models sink $10-20\%$ faster.

- The density and effective radius of the central NSC produced after the merger are consistent with those of observed NSCs. Specifically, our density profiles fall within the range of known nucleated dwarfs drawn from Nguyen et al. (2017, 2018). Additionally, the density of our NSCs aligns with the $\rho - M_{*,\mathrm{gal}}$ relationship from Pechetti et al. (2020). While the final effective radii of the merger product are somewhat smaller than expected for more massive NSCs in models using a fixed initial effective radius, our investigations reveal that selecting the initial effective radius based on Pechetti et al. (2020) mitigates this issue. This paves the way for future studies to investigate more realistic models of NSCs derived from observational relations.

- We find that binaries harden efficiently when embedded in sufficiently dense NSCs. As expected from theoretical predictions, the hardening rate s is proportional to the density at the influence radius ρ_{infl}. In our densest model, $s \approx 450 \mathrm{Myr}^{-1}\mathrm{pc}^{-1}$, while in the least dense model, $s \approx 0.04 \mathrm{Myr}^{-1}\mathrm{pc}^{-1}$. Comparing our hardening rates to recent studies of MBH binaries in NSCs, we find consistent results, with differences of at most 2-4 when $\rho_{\mathrm{infl}} \sim 10^4 M_\odot \mathrm{pc}^{-3}$. Although three-body interactions with stars from the NSCs and bulge primarily drive the hardening, dark matter (DM) plays a dominant role in low-mass clusters. In our lowest mass model, the hardening rate due to DM is twice that due to stars. Based on our initial conditions, half of our models yield bound binaries with high eccentricity ($e > 0.9$). Although the average unbound eccentricity decreases during the initial dynamical friction stage, substantial stochasticity makes predicting the bound binary eccentricity from the initial orbital eccentricity challenging.

- We follow the evolution of the binary into the GW merger stage using semi-analytic methods, assuming that the hardening rate s remains constant and

incorporating GW effects via the Peters (1964c) equations. Using the $s - \rho_{\rm infl}$ relationship derived in this work, we determine the merger timescales of seed binaries with $M_{\rm bin} \approx 1.5 \times 10^5 M_\odot$ in various environments. We find that a combination of high eccentricity and high stellar density is necessary for seeds to merge at high redshifts ($z > 4$). Even with a hard binary eccentricity of ~ 0.9 our models predict that $\rho_{\rm infl}$ should exceed $3 \times 10^4 M_\odot {\rm pc}^{-3}$, achievable only in nucleated environments. We find that 8 out of 12 of our models merge by $z = 4$. Combined with the conclusions from Zhou et al. (2024), which indicate that naked MBH seeds do not sink efficiently to the center, our findings strongly suggest that NSCs are a crucial component in the seed sinking process. We also predict the merger rate of MBH seed binaries using data from `ASTRID` and add a hardening delay time based on different $\rho_{\rm infl}$, finding that detectable mergers at high redshifts would originate from sources embedded in dense nuclei. For $\rho_{\rm infl} = 10^3 M_\odot {\rm pc}^{-3}$, our results predict that the merger rate is 2-3 orders of magnitude lower than when $\rho_{\rm infl} \geq 10^4 M_\odot {\rm pc}^{-3}$.

While our NSC models are idealized and do not account for gas effects, accretion, or MBH spin, this work lays the groundwork for more comprehensive future studies. Our findings highlight the critical role of NSCs in shaping the dynamics and evolution of seed MBHs, demonstrating that the seed sinking problem can be mitigated by the presence of extended stellar systems. With future integration of MBH spins, GW recoil, and enhancements to the FMM kernels, `Taichi` stands out as a powerful tool for addressing the complex problem of MBH mergers in galactic nuclei.

Chapter 5

Effects of dark matter spikes on eccentric intermediate mass ratio inspirals

Diptajyoti Mukherjee[1], A. Miguel Holdago[1], Go Ogiya[3], Hy Trac[1,2]

[1]McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213
[2] NSF AI Planning Institute for Physics of the Future, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[3] Institute for Astronomy, School of Physics, Zhejiang University, Hangzhou 310027, China

Abstract

Recent studies suggest that dark matter (DM) spikes around intermediate-mass black holes could cause observable dephasing in gravitational wave (GW) signals from Intermediate Mass Ratio Inspirals (IMRIs). Previous research primarily used non-self-consistent analytic methods to estimate the impact of DM spikes on eccentric IMRIs. Our study provides the first self-consistent treatment of this phenomenon using N-body simulations, incorporating Post-Newtonian effects up to the 2.5 order for accurate and robust results. Contrary to prior works, which posited that the cumulative effect of two-body encounters (dynamical friction; DF) is the primary mechanism for energy dissipation, we reveal that a three-body effect (slingshot mechanism) plays a more significant role in driving the binary system's energy loss and consequent orbital shrinkage. We find that binaries counter-rotating with respect to the DM spike merge faster, while co-rotating binaries merge slower, contrary to expectations from the DF theory. Using Fokker-Planck methods, we also assess the presence and detectability of spikes in realistic environments. When interacting with surrounding materials, DM spikes can have shallower slopes and lower densities than previously considered, leading to smaller signals and lower detection prospects via

dephasing. Our results suggest that 'deshifting' rather than dephasing might be a more optimistic signature, as it is more robust even in low-density environments.

5.1 Introduction

The nature of dark matter (DM) remains one of the most pressing mysteries in modern astrophysics. While its existence has been inferred through a range of indirect observations such as galaxy rotational curves (e.g., Rubin and Ford, 1970; Rubin et al., 1978, 1980; Persic et al., 1996), and gravitational lensing of galaxy clusters (e.g., Tyson et al., 1990; Hammer, 1991; Fort et al., 1992; Le Fevre et al., 1994), its properties and interactions remain largely unknown. Cosmological simulations suggest that DM resides in galactic halos and is distributed according to the Navarro-Frenk-White profile (Navarro et al., 1996; Bertone, 2010, see second reference for a review). Galaxies, and therefore galactic halos, often contain massive black holes (MBHs) at the center. Gondolo and Silk (1999) proposed that the adiabatic growth of these MBHs would modify the dark matter profile near them and create extremely dense density spikes. The density profile of these DM spikes $\rho_{\rm DM}$ is parameterized as a power-law profile and is given as (e.g., Gondolo and Silk, 1999; Eda et al., 2013; Kavanagh et al., 2020)

$$\rho_{\rm DM}(r) = \rho_{\rm sp}\left(\frac{r_{\rm sp}}{r}\right)^{\gamma_{\rm sp}} \tag{5.1}$$

where r denotes the distance from the central MBH, $\rho_{\rm sp}$ is a normalization factor for the density profile, $\gamma_{\rm sp}$ characterizes the power-law profile of the spike and $r_{\rm sp}$ is a characteristic radius of the spike.Usually $\gamma_{\rm sp}$ is taken to be between 2 to 2.5. For a central MBH with a mass of $10^3 M_\odot$, under these parameters, $r_{\rm sp} \approx 0.5$pc and the density 10^{-6} pc from the MBH is $\sim 10^{15} M_\odot {\rm pc}^{-3}$.

Although the authors showed that such spikes would typically form around Supermassive Black Holes (SMBHs) with $M_{BH} \geq 10^6 M_\odot$, where M_{BH} is the mass of the black hole (BH), later work (e.g., Ullio et al., 2001; Merritt et al., 2002; Merritt, 2004; Bertone and Merritt, 2005) found that spikes could be depleted via different astrophysical processes such as galaxy mergers, scattering with nearby stars, and off center seed BHs. While SMBHs are expected to undergo many mergers, the same cannot be said about Intermediate Mass Black Holes. DM spikes around Intermediate Mass Black Holes (IMBHs) with $10^3 \leq M_{BH} \leq 10^6$ are expected to survive as they undergo fewer mergers (Zhao and Silk, 2005; Bertone et al., 2005). Additionally, Ferrer et al. (2017) showed that spinning IMBHs could form denser spikes. Thus, IMBHs are expected to be the predominant source for black holes surrounded by DM spikes.

The advent of gravitational wave (GW) astronomy using LIGO-Virgo interferometers (e.g., Abbott et al., 2017, 2020c,b,a) and pulsar timing array (PTA) (e.g., Mingarelli et al., 2017; Kelley et al., 2018; Agazie et al., 2023a,b) has opened up new avenues to detect and explore the properties of DM. In particular, it has been

suggested that the effects of a DM spike could be imprinted on the GW signal of an Intermediate Mass Ratio Inspiral (IMRI) and has drawn a lot of interest (e.g., Eda et al., 2013; Eda et al., 2015; Yue and Han, 2018; Yue et al., 2019; Kavanagh et al., 2020; Becker et al., 2022; Dai et al., 2022). A visual representation of such a system can be seen in Figure 5.1. IMRIs, formed from the inspiral of compact objects such as white dwarfs or neutron stars, or stellar mass black holes into IMBHs, will be detectable by future space based mHz GW detectors like LISA (Amaro-Seoane et al., 2017) and TianQin (Luo et al., 2016) or deciHz GW detectors like DECIGO (Kawamura et al., 2019) or MAGIS (Abe et al., 2021). GWs emanating from such systems are expected to be in the LISA and DECIGO band for periods of months to years allowing environmental effects, including DM spikes, to play a major role in modifying the signal. Since matched filtering relies on a careful determination of the simulated signal to a few cycles over hundreds to thousands of cycles, it is imperative to take into account the environmental effects while numerically calculating the simulated GW signal (e.g., Eda et al., 2013; Macedo et al., 2013; Zwick et al., 2022; Coogan et al., 2022; Baumann et al., 2022; Zwick et al., 2023; Cole et al., 2023a,b)

Eda et al. (2013) and Eda et al. (2015) first proposed that the gravitational effects of the spike could leave an imprint on the GW signal of an IMRI. The spike particles would get scattered by the inspiraling object leading to an extra drag force experienced by the inspiraling object. This drag force has been attributed to dynamical friction (DF; Chandrasekhar, 1943) on the inspiraling object. Although the DM mass contained in the spike is quite small relative to the mass of the central IMBH and the inspiraling object, it has been shown that the gravitational drag can lead to a phase shift in the GW signal over thousands of cycles. Recent studies have shown that dephasing due to DF can lead to anywhere between $10^3 - 10^7$ fewer cycles which would be above the signal to noise detector theshold for LISA and could be detected and distinguished as an imprint of the DM spike surrounding the IMBH (e.g., Eda et al., 2015; Kavanagh et al., 2020; Becker et al., 2022; Dai et al., 2022). Different models of DM can lead to different spike parameters and as such change the amount of dephasing itself, which can be detected and distinguished. As such, GW astronomy provides a unique pathway to ascertain the presence and nature of DM itself. Detection of even one DM spike modified signal would be effective towards eliminating other theories of gravity such as modified Newtonian dynamics and can help constrain the particle nature of DM (Hannuksela et al., 2020).

A careful determination of the the expected dephasing, therefore, is vital. While initial studies (e.g., Eda et al., 2013; Eda et al., 2015) used a static DM background, Kavanagh et al. (2020) showed that the inspiral of the IMRI can inject energy into the spike itself, often comparable to the binding energy of the spike. As such the feedback of the IMRI onto the spike cannot be ignored. Through a semi-analytic framework called `HaloFeedback`, Kavanagh et al. (2020) showed that the including the back reaction from the IMRI can lead upto $100\times$ reduction in expected amount of dephasing. Therefore, it is necessary to self-consistently follow the dynamics of the

binary in order to calculate the amount of dephasing. While `HaloFeedback` relies on the assumption that the binary is on a circular orbit at all times, studies (e.g., Yue and Han, 2018; Yue et al., 2019; Cardoso et al., 2021; Dai et al., 2022) showed that such an assumption might not hold in reality. A self-consistent framework has only been developed for binaries on circular orbits and studies including eccentric binaries neglect the backreaction from the binary to the spike (e.g., Becker et al., 2022; Dai et al., 2022). The feedback can not only affect the evolution of the semi-major axis but also the eccentricity which has a major impact on inspiral times due to GW emission. The rate of circularization of the IMRI has also recently been suggested as a signature of the spike (Becker et al., 2022) which would require a self-consistent framework that takes into account the effects of the feedback from the binary onto the spike and vice versa to be determined accurately.

Additionally, recent work has shown that the analytic Chandrasekhar approximation may lead to inconsistent evolution of semi-major axis and eccentricity due to the lack of inclusion of drag force from fast moving particles, which can be significant in certain cases (Dosopoulou, 2023). As it is difficult to create a self-consistent analytic/semi-analytic framework to include these effects, we have to resort to N-body simulations.

N-body simulations, although expensive, are considered to be the gold standard of dynamical modelling. Previous approaches have rarely relied on N-body simulations as they are extremely computationally expensive. Kavanagh et al. (2020) report that a simulation of 150 orbits of a 1 $M_\odot$ - 100 $M_\odot$ binary takes about 3 days using the N-body code `GADGET-2` (Springel, 2005). This is very computationally expensive to study long term effects of the IMRI on the spike over tens of thousands of orbits. Therefore, in order to study the secular effects of the IMRI on the spike, different strategies are required.

In our study we describe a novel N-body code that is over 100 times faster than traditional N-body codes for simulating IMRIs embedded in DM spikes. Using our code, we present self-consistent results from eccentric IMRIs embdedded in DM spikes.

In addition, we also study the effect of rotation in the DM spike which has not been done before. We argue that the inclusion of rotation is important since conservation of angular momentum would dictate that DM spikes should rotate upon formation from rotating galactic halos. Spinning IMBHs can also transfer angular momentum to the surrounding spike, making it rotate (Ferrer et al., 2017). We systematically study the effects of the different post-Newtonian (PN) terms including precession and radiation reaction terms upto $2.5PN$.

We begin by describing our computational methods in Section 5.2, followed by the models of the IMRIs embedded in DM spikes in Section 5.3. The results are then described in Section 5.4, followed by discussions of our results and assumptions and conclusions in Sections 5.5 and 5.6 respectively.

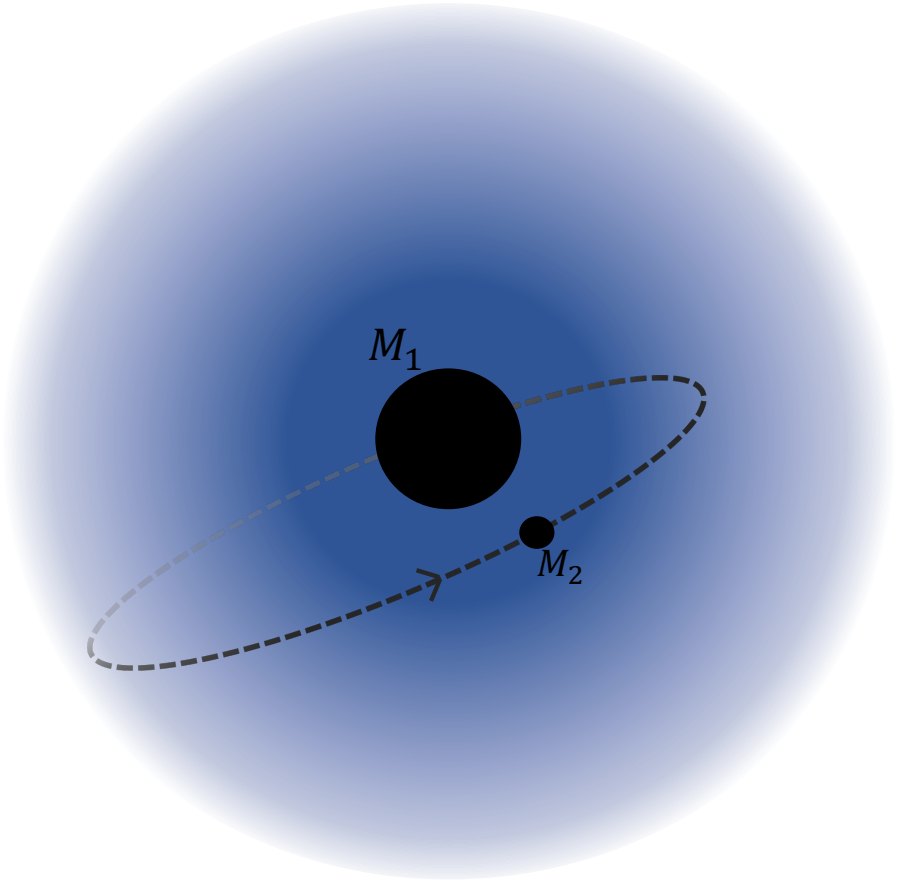

Figure 5.1: A visual representation of an eccentric IMRI embedded in a DM spike. The mass of the central BH is M_1 and that of the inspiraling object is M_2. This figure has been inspired by Figure 1 of Kavanagh et al. (2020).

5.2 Computational Methods

The numerical evolution of the IMRI using N-body methods is an extremely computationally challenging problem owing to the enormous density of the spike near the central IMBH and the $\mathcal{O}(N^2)$ pairwise force calculations. An accurate evolution requires fine tuned time-steps for the particles near the central IMBH (primary) and inspiraling object (secondary) to resolve the scattering effects accurately. Previous studies have all relied on analytic/semi-analytic methods for long term simulations of the inspiraling IMRI. To the best of our knowledge, only Kavanagh et al. (2020) performed N-body simulations of the binary embedded in a spike. However, the simulations were stopped after a period of ~ 100 orbits. To understand realistic effects of the spike on the IMRI (and vice-versa) and to calibrate semi-analytic methods, we need long term N-body simulations. Here we present a novel N-body code [1] that is specifically tuned for simulations of IMRIs in DM spikes.

A close analysis of the N-body simulation reveals that the bulk of computational time is spent calculating the forces of DM particles near the primary. Since the mass of DM in the region of interest is very small compared to the mass of the primary and the secondary, the self-gravity of the spike can be neglected. This allows us to safely neglect the interactions between the DM particles themselves. Full force calculations are only required for the primary and the secondary. Neglecting the self interaction of the spike effectively reduces the number of force calculations to $\mathcal{O}(N)$, speeding up the simulations massively. The mean relative force accuracy between a full $\mathcal{O}(N^2)$ calculation and our approximate method is $\leq 10^{-5}$ in the region of interest as shown in Figure 5.2, on par with the force accuracies obtained by current Barnes-Hut tree (Barnes and Hut, 1986) or Fast Multipole Method (FMM) (Greengard and Rokhlin, 1987; Cheng et al., 1999; Dehnen, 2002; Zhu, 2021c) based N-body codes. To verify the accuracy of our method, we compare the evolution of a $100M_\odot - 1M_\odot$ binary in a DM spike with the FMM based code `Taichi` (Zhu, 2021c; Mukherjee et al., 2021a, 2023) and this method and find negligible differences after ~ 1500 orbits.

To evolve the particles, we use a 2nd order hierarchical Hamiltonian splitting based integration scheme `HOLD` in a Drift-Kick-Drift (DKD) fashion with symmetrized timesteps (Pelupessy et al., 2012a). The usage of symmetrized timesteps ensures that there is no secular drift in energy and renders better energy conservation than other timestepping schemes (e.g., Makino et al., 2006a; Pelupessy et al., 2012a; Mukherjee et al., 2021a). For more information on the integration scheme, we refer the reader to Pelupessy et al. (2012a). The timesteps are controlled by a timestepping parameter η which is proportional to the analytically calculated timestep. In our simulations, we set $\eta = 0.025$. This results in a relative energy error of $\leq 10^{-10}$ per orbit, sufficient to follow the dynamics over few hundreds of thousands orbits.

To account for relativistic effects, we add PN terms to the equations of motions for the primary and secondary. The PN equations of motion can be added to the

[1] `https://github.com/dipto4/falcon_dm/`

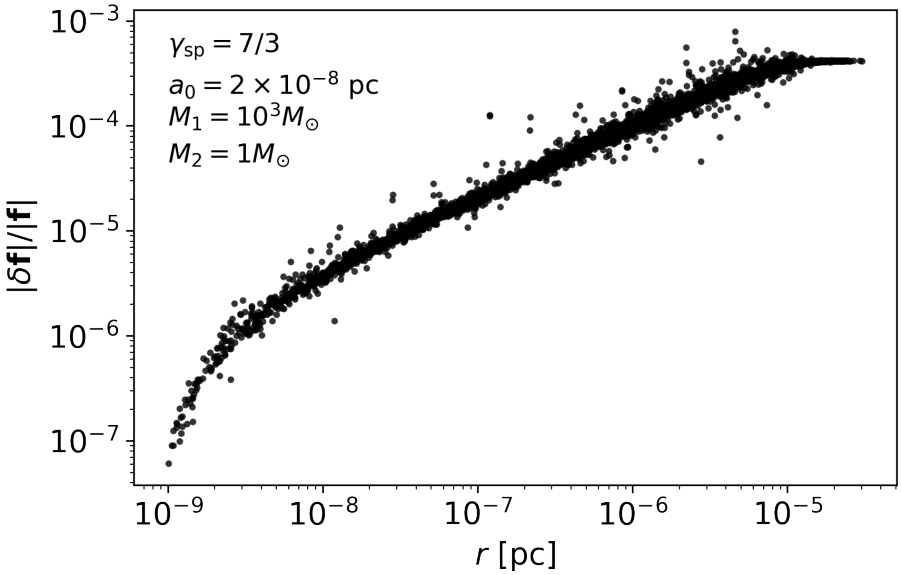

Figure 5.2: The relative force accuracy $|\delta \mathbf{f}|/|\mathbf{f}|$ of DM particles as a function of their separation r from the central IMBH with mass $M_1 = 10^3 M_\odot$. The spike follows a $\gamma_{\rm sp} = 7/3$ density profile. The inspiraling object, whose mass is $M_2 = 1 M_\odot$, is situated at an initial semi-major axis of $a_0 = 2 \times 10^{-8}$ pc. We find that in the region of interest, the force accuracy is $\leq 10^{-5}$, which is comparable to the the force accuracy obtained in tree/FMM based codes.

standard Newtonian equation of motion as follows:

$$\mathbf{a} = \mathbf{a}_{\rm Newtonian} + \frac{1}{c^2}\mathbf{a}_{1PN} + \frac{1}{c^4}\mathbf{a}_{2PN} + \frac{1}{c^5}\mathbf{a}_{2.5PN} + ... \tag{5.2}$$

where $\mathbf{a}_{nPN}$ represents the $n^{\rm th}$ PN term. We include up to $2.5PN$ terms in our calculations following equation 203 from Blanchet (2014). While the $1PN$ and $2PN$ terms only lead to precession of the binary's orbit, the $2.5PN$ term leads to GW radiation resulting in shrinkage of the orbit. The DM particles experience the Newtonian potential of the primary and the secondary. In principle this is not fully self-consistent since the DM particles will also experience PN effects in interactions with other DM particles. One can use the Einstein-Infield-Hoffman equations (e.g., Will, 2014; Portegies Zwart et al., 2022) to self-consistently simulate the PN evolution of all the particles but this is beyond the scope of this study.

A problem arises when adding velocity dependent forces like PN precession and radiation reaction terms since a leapfrog-like explicit splitting of the Hamiltonian is not achievable with velocity dependent forces. In such a scenario, the simple DKD integration scheme cannot be utilized. An implicit method suggested by Mikkola and Aarseth (2002) is popular but is quite inefficient. The implicit scheme requires multiple iterations to solve for the force. For highly non-linear vector fields, it proves to be quite computationally expensive. Hellström and Mikkola (2010) show that using a clever mathematical trick one can create an explicit DKD-like scheme by extending the phase space of variables by introducing an auxiliary velocity $\mathbf{w}$. An updated integration between step t and $t+1$ using the auxiliary velocity $\mathbf{w}$ can now be written as:

$$\mathbf{x}_{t+1/2} = \mathbf{x}_t + \frac{h}{2}\mathbf{v}_t \tag{5.3}$$

$$\mathbf{w}_{t+1/2} = \mathbf{w}_t + \frac{h}{2}\mathbf{f}\left(\mathbf{x}_{t+1/2}, \mathbf{v}_t\right) \tag{5.4}$$

$$\mathbf{v}_{t+1} = \mathbf{v}_t + h\mathbf{f}\left(\mathbf{x}_{t+1/2}, \mathbf{w}_{t+1/2}\right) \tag{5.5}$$

$$\mathbf{w}_{t+1} = \mathbf{w}_{t+1/2} + \frac{h}{2}\mathbf{f}\left(\mathbf{x}_{t+1/2}, \mathbf{v}_{t+1}\right) \tag{5.6}$$

$$\mathbf{x}_{t+1} = \mathbf{x}_{t+1/2} + \frac{h}{2}\mathbf{v}_{t+1} \tag{5.7}$$

where $\mathbf{x}$ is the position, $\mathbf{v}$ is the velocity, $\mathbf{f}$ is the force and h is the timestep. We note that $\mathbf{w}_0 = \mathbf{v}_0$.

To test the integration scheme, we simulate the evolution of a $1000M_\odot - 1M_\odot$ binary in vacuum and compare the evolution of the eccentricity and semi-major axis to those derived using Peters' equations (Peters, 1964b) and find that the results are consistent with one another. We also compare the evolution of the binary against that from the publicly available regularization code `SpaceHub` (Wang et al., 2021a) which also includes PN terms upto 2.5PN and find no differences.

For three-body simulations, we utilize the 15th order Gauss-Radau integrator `IAS15` (Rein and Spiegel, 2015) from the `rebound` package (Rein and Liu, 2012). `IAS15` can handle close encounters and is extremely accurate allowing for a closer and in-depth analysis into the dynamics of the scattering process of the binary with a DM particle. The energy is conserved to machine precision.

All simulations are performed on the Vera computing cluster utilizing `AMD Epyc 7742` nodes. The majority of the full N-body simulations take about 100-120 core hours to finish to completion using a single core. The interactions between the secondary and DM particles are set to have zero softening while we set a relatively conservative softening value of 10^{-10} pc between the primary and DM particles in our simulations. This is equal to the Schwartzchild radius for the $10^3 M_\odot$ IMBH and one-tenth that for the $10^4 M_\odot$ IMBH. For the three-body simulations, all interactions are unsoftened. A set of simulations were run with softening between the DM particles and the secondary and no major differences were noticed between the simulations that used softening and those that did not. A small value of softening was necessary to prevent some particles near the primary from taking extremely small timesteps. The interaction between the primary and the secondary object is not softened.

5.3 Models

We are interested in scenarios wherein the IMRI is visible in the LISA/DECIGO band for a duration of ~ 5 years. To understand the effects of varying eccentricity, primary and secondary masses, and density, we generate physically motivated models corresponding to IMRIs in vacuum where the GW signal is in the LISA/DECIGO band for ~ 5 years and above the LISA/DECIGO signal to noise threshold. We define the mass-ratio of the secondary to the primary q as $q \equiv \frac{M_2}{M_1}$ where M_1 is the mass of the primary and M_2 is the mass of the secondary. We choose $M_1 = 10^3 M_\odot, 10^4 M_\odot$ and $M_2 = 1 M_\odot, 10 M_\odot$, representing three different scenarios with mass ratios $q = 10^{-2}, 10^{-3}, 10^{-4}$ as depicted in Figure 5.3. $M_2 = 1 M_\odot$ is representative of a scenario the inspiraling object is similar to a white dwarf of a neutron star whereas $M_2 = 10 M_\odot$ represents a scenario where the inspiraling object is a stellar mass black hole. The initial properties of the binary are characterized by its initial semi-major axis a_0 and eccentricity e_0. We set $a_0 = 2 \times 10^{-8}$ pc for $q = 10^{-2}, 10^{-3}$ models and $a_0 = 5 \times 10^{-8}$ for the $q = 10^{-4}$ model. We also set $e_0 = 0.4$ for $q = 10^{-2}$ model, $e_0 = 0.75$ for $q = 10^{-3}$ model, and $e_0 = 0.65$ for $q = 10^{-4}$ model. Figure 5.3 provides a representation of the strain versus frequency curves of the models superimposed over the strain-frequency curves for LISA and DECIGO.

5.3.1 Non-rotating models

To generate the N-body realizations of the DM spike, we utilize the galactic modeling toolkit `Agama` (Vasiliev, 2019a). Following previous studies (e.g., Eda et al., 2015;

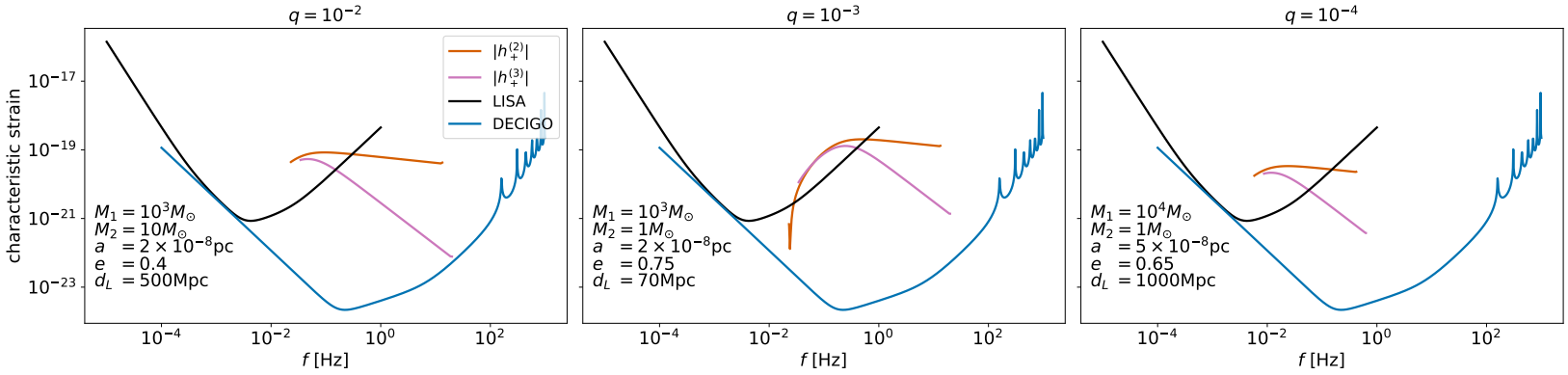

Figure 5.3: The characteristic strain of the IMRIs used as our initial conditions as a function of the frequency f in vacuum for different mass-ratio q models. The central IMBH has a mass of M_1 while the inspiraling object has a mass of M_2. The binary is defined by its initial semi-major axis a and eccentricity e. The luminosity distance is denoted as d_L. All of the IMRIs presented here have a merger time of ~ 5 years. Since eccentric binaries radiate in multiple harmonics, we have plotted the strain from the second and third harmonics. In all of the scenarios, we find that the second harmonic has an equivalent or higher strain than the third harmonic. Even higher harmonics have lower strains and are harder to detect using LISA but would be detectable using DECIGO.

M_1 $[M_\odot]$	M_2 $[M_\odot]$	q	a_0 [pc], e_0	$\gamma_{\rm sp}, r_{\rm sp}$[pc]	$m_{\rm DM}$ $[M_\odot]$	PN terms used	$t_{\rm final}$	$N_{\rm sims}$ (for each config)
10^3	10.0	10^{-2}	$2\times10^{-8}, 0.4$	$7/3, 0.54$	5×10^{-5}	2.5PN	merger	5
10^3	1.0	10^{-3}	$2\times10^{-8}, 0.75$	$7/3, 0.54$	5×10^{-5}	2.5PN, all terms up to 2.5PN	merger	5
10^3	1.0	10^{-3}	$2\times10^{-8}, 0.75$	$9/4, 0.54$	10^{-5}	2.5PN	merger	5
10^3	1.0	10^{-3}	$2\times10^{-8}, 0.75$	$3/2, 0.46$	10^{-11}	2.5PN	merger	5
10^4	1.0	10^{-4}	$5\times10^{-8}, 0.65$	$3/2, 0.98$	2×10^{-10}	2.5PN	merger	5
10^4	1.0	10^{-4}	$5\times10^{-8}, 0.65$	$7/3, 0.54$	5×10^{-5}	2.5PN	merger	5
10^4	1.0	10^{-4}	$5\times10^{-8}, 0.65$	$7/3, 1.17$	5×10^{-5}	2.5PN	2.5 yr	5

Table 5.1: A summary of the initial conditions for our non-rotating models. The rotating models have similar initial conditions but are either in a net prograde motion or retrograde motion with respect to the motion of the IMRI. For more information on how to generate the initial conditions, we refer the reader to section **??**.

Kavanagh et al., 2020), we use equation 5.1 as our density profile. This profile is valid for all for all radii $r > r_{\rm ISCO}$ where $r_{\rm ISCO}$ is the innermost stable circular orbit (ISCO) for the central IMBH. The density profile is taken to be 0 where $r < r_{\rm ISCO}$. Furthermore, $r_{\rm sp}$ is not considered to be a free parameter but is, instead, calculated as

$$r_{\rm sp} \approx \left(\frac{(3-\gamma_{\rm sp}) 0.2^{3-\gamma_{\rm sp}} M_1}{2\pi \rho_{\rm sp}} \right)^{1/3}. \tag{5.8}$$

For all our simulations we set $\rho_{\rm sp} = 226 M_\odot/{\rm pc}^3$ following Kavanagh et al. (2020). Using equation 5.8, we find $r_{\rm sp} \approx 0.54$ pc when $M_1 = 10^3 M_\odot$ and $r_{\rm sp} \approx 1.17$ pc when $M_1 = 10^4 M_\odot$. Additionally, for most of our simulations we set $\gamma_{\rm sp} = 7/3$. Such a spike profile can arise out of adiabatic growth of an IMBH in an NFW halo (Eda et al., 2015). However, this is not universal and $\gamma_{\rm sp}$ depends on how the DM spike originates. For example, primordial BHs have been shown to produce a $\gamma_{\rm sp} = 9/4$ spike around them (Boudaud et al., 2021). To understand how the dephasing changes as $\gamma_{\rm sp}$ changes, we run a set of simulations with $\gamma_{\rm sp} = 9/4$. Furthermore, to understand how the dephasing changes at a fixed density profile for different q, we perform a set of simulations with $M_1 = 10^4 M_\odot, r_{\rm sp} = 0.54$pc and $\gamma_{\rm sp} = 7/3$. However, it is hard to robustly predict the existence of such dense spikes. In realistic environments where effects of stars cannot be neglected, DM "crests" are more likely. Such "crests" are formed due to the scattering of DM particles by the more massive more massive stellar particles and have been shown to produce a $\gamma_{\rm sp} = 1.5$ profile (Merritt et al., 2007b). To understand if such profiles produce any observable dephasing effects, we perform a set of simulations with $\gamma_{\rm sp} = 1.5$, $M_1 = 10^3, 10^4 M_\odot$ and $M_2 = 1 M_\odot$. The density spikes considered in our simulations have been plotted in Figure 5.4. To generate models with finite total mass, we use an exponential truncation function with a truncation radius of 10^{-5} pc for $q = 10^{-2}, 10^{-3}$ models and 10^{-6} pc for the $q = 10^{-4}$ model. We verify that our choice of truncation radius does not affect the mass profile of the spike in the region of interest. We also compare our initial N-body profiles to those from Kavanagh et al. (2020) and find that the density profiles in the region of interest and the velocity profiles match. Additionally, we ensure that the density profile remains stable for the duration of the simulations.

The mass of the DM particle, $m_{\rm DM}$, is chosen carefully. Since $M_1 \gg M_2$, the dynamics of inspiral is dependent on the mass ratio of the secondary to the mass of the DM particle (e.g., Merritt, 2013a, chapter 8 and references therein). We define the mass-ratio of the secondary to the DM particle as $q_{\rm DM} \equiv \frac{m_{\rm DM}}{M_2}$. Three-body simulations show that using $q_{\rm DM} \leq 10^{-2}$, the results are convergent. The energy, angular momentum, and ejection time distributions are consistent between all of the models where $q_{\rm DM} \leq 10^{-2}$. Such a low mass ratio is desirable since it reduces any spurious three-body effects leading to sudden jumps in the evolution of the semi-major axis or eccentricity and ensures that the evolution is smooth. Accordingly, based on resolution and computational expenses, we choose $m_{\rm DM} = 5 \times 10^{-5} M_\odot$ for the

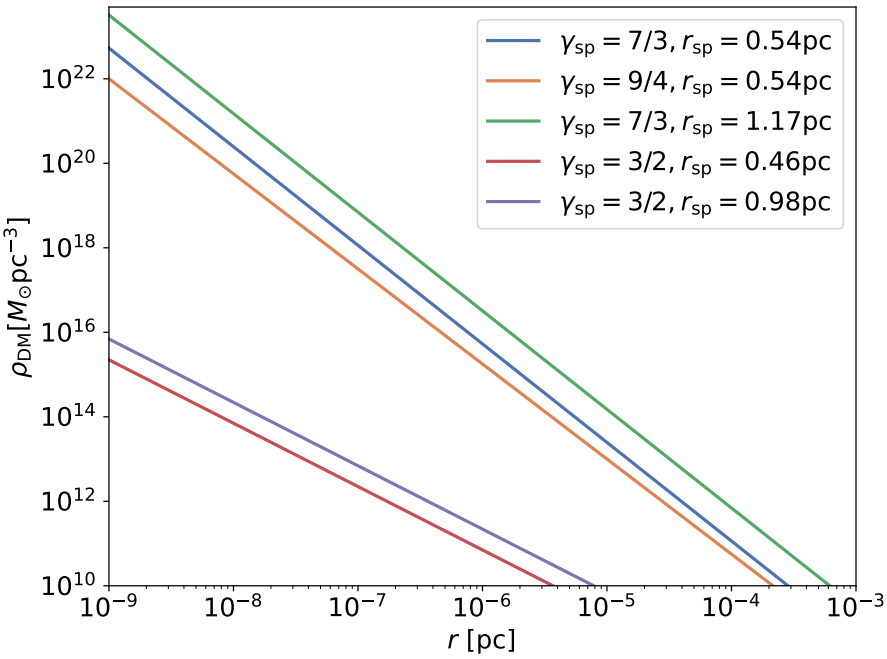

Figure 5.4: The density of the DM spike $\rho_{\rm DM}$ as a function of the distance from the central IMBH r. We use these density profiles to generate our N-body initial conditions.

simulations using $\gamma_{\rm sp} = 7/3$ and $10^{-5} M_\odot$ when $\gamma_{\rm sp} = 9/4$. This results in $\approx 8\text{k} - 10\text{k}$ particles in the simulation. Extra care is taken to ensure that the region surrounding the secondary have a sufficient number of particles to resolve the scattering process. Since the scattering process is quite stochastic, results from an individual N-body simulation are unreliable. Accordingly, five independent realizations of every model are generated and simulated.

For most of the simulations, only the $2.5PN$ term is added to the equation of motion of the binary. This is done in order to provide a direct comparison to previous studies such as Kavanagh et al. (2020) and Becker et al. (2022) where the effects of relativistic precession are neglected. In general, we do not expect the dephasing to be affected by any precession effects. However, the net precession can itself be affected by the Newtonian precession of the binary in the potential of the spike in addition to the relativistic precession, which can also be a signature of the spike (Dai et al., 2022). It can, also, potentially affect the exchange of angular momentum between the spike to the IMRI. In order to understand if relativistic precession affects our results, we run a set of simulations with all PN terms enabled up to $2.5PN$ for $q = 10^{-3}$ models. We leave a more systematic study of precession effects to future work. A summary of the initial conditions for the non-rotating models can be found in Table 5.1.

We evolve the $q = 10^{-2}$, and 10^{-3} models to merger. This is defined as the time when $a < r_{\rm ISCO}$. The simulations where $q = 10^{-4}$ are extremely computationally expensive. We only achieve an evolution to 2.5 yr in 15-20 days of computing time. This results in about 2×10^6 orbits of the secondary around the primary. We calculate the dephasing cycles by extrapolating the rest of the evolution in vacuum to save on computational resources. To verify if the extrapolation produces valid results, we run a set of lower resolution simulations with 1024 DM particles to completion and compare our extrapolated results to the results from our lower resolution simulations. The extrapolated results can be seen as a lower limit on the number of dephasing cycles. We compare our results against the vacuum evolution computed using the Peters (1964b) analytic formula and also against those calculated using the Chandrasekhar DF assuming there is no backreaction on to the spike from the IMRI. This is done using the publicly available code `IMRIPy` (Becker et al., 2022; Becker and Sagunski, 2023; Becker, 2023). The `IMRIPy` results are denoted as "Static DF" in the results section. Unfortunately, we could not use `HaloFeedback` (Kavanagh et al., 2020) for our purposes since it can only simulate circular binaries.

5.3.2 Rotating models

As mentioned in the introduction, we aim to understand the effect of rotation in the DM spike and its effect on the IMRI. If the DM halo surrounding the IMBH is rotating, angular momentum conservation will dictate that DM spike should rotate as well. Additionally, spinning IMBHs will transfer angular momentum from the

IMBH to the spike, torquing the spike up or down depending on the direction of the primitive rotation of the spike (Ferrer et al., 2017). In a rotating spike, exchange of angular momentum between the IMRI and the spike will be enhanced compared to the non-rotating models and the eccentricity of the IMRI can be significantly affected. We aim to understand this effect in our study. To the best of our knowledge, there is no literature available quantifying the level of rotation in the spike and its correlation to the galactic halo rotation or the spin of the central IMBH. Therefore, we caution the reader that our rotating models are somewhat ad hoc and therefore exploratory. However, they are still useful in understanding the qualitative effects a spike that is in a counter-rotating motion (retrograde motion) or co-rotating motion (prograde motion) with respect to the binary.

To include rotation, we follow the Lynden-Bell trick (Lynden-Bell, 1960). This has been motivated by several studies which use the method to generate rotating models of galactic nuclei (e.g., Holley-Bockelmann and Khan, 2015; Khan et al., 2020; Khan and Holley-Bockelmann, 2021; Varisco et al., 2021) to examine the effects of rotation upon the dynamics of MBH binaries. In this method, we flip the z-component of the angular momentum of the DM particles, $L_{z,DM}$, to generate prograde or retrograde models. We note that the z-component of the angular momentum of the IMRI, L_z, is always positive. Accordingly, to generate co-rotating or prograde models, we reverse the direction of the x and y components of the velocity for all DM particles which satisfy $L_{z,DM} < 0$. To generate counter-rotating or retrograde models, we do the same, except with particles with $L_{z,DM} > 0$. This leaves the density and velocity profiles of the spike unchanged. In principle, this is not a fully self-consistent method of generating rotating models. Rotation is expected to flatten out the models to some extent so it is somewhat ad hoc to include rotation in spherical spikes. One can construct a distribution function $f(E, L_z)$ as a function of the energy E and z-component of the angular momentum L_z as is done in Wang et al. (2014) to generate self-consistent rotating models which are flattened. However, this is beyond the scope of this work. Irrespective of that, our models are able to qualitatively understand the effect of rotation on the IMRI. In future studies, we plan on investigating self-consistent models of rotating DM spikes to understand how the geometry of the spike affects the dynamics.

We denote the ratio of retrograde particles to the total number of particles as $\mathcal{F}$. In our retrograde simulations, $\mathcal{F} = 1$, indicating that *all* particles are moving in a retrograde motion with respect to the binary, and in our prograde simulation, $\mathcal{F} = 0$, which is opposite of the previous scenario. Realistic spikes are expected to have $0 \leq \mathcal{F} \leq 1$. Since the simulations are quite expensive even with our fast N-body code, we restrict our study only to the $\mathcal{F} = 0, 1$ cases. This helps us put limits on the dephasing effect and compare the rotating models to fully isotropic non-rotating models. We note that the Lynden-Bell method is not the same as introducing rigid body rotation into the DM spike. The latter generates much stronger rotation whereas our method introduces a weaker level of rotation.

We generate prograde and retrograde rotating versions of all of the models presented in Table 5.1 and create five independent realizations for each configuration. All models were evolved to merger, except in the case where $M_1 = 10^4 M_\odot$ as explained in the section above.

5.4 Results

Owing to the inherent stochasticity present due to the discrete nature of N-body simulations, we note that the results presented here are calculated after taking the average of the quantities across all five independent realizations. In all of the calculated values, we noticed a maximum of 1 percent difference between results from individual simulations This does not affect our overall results and indicates the robustness of our simulations. We first present the results from the non-rotating models before moving on to the rotating models.

5.4.1 Non-rotating models

The mean orbital frequency of the binary f can be written as

$$f = \frac{1}{2\pi}\sqrt{\frac{GM}{a^3}} \tag{5.9}$$

where $M = M_1 + M_2$. We calculate the amount of dephasing by taking the difference between the number of GW cycles completed by the binary with and without the spike. Following Becker et al. (2022), we write the number of GW cycles $N^{(n)}(t_0, t_{\rm final})$ for the $n^{\rm th}$ harmonic between some initial time t_0 and final time $t_{\rm final}$ as follows:

$$N^{(n)}(t_0, t_{\rm final}) = n \int_{t_0}^{t_{\rm final}} f(t) dt \tag{5.10}$$

We can, then, calculate the difference in the number of GW cycles $\Delta N^{(n)}(t_0, t_{\rm final})$, or dephasing, as:

$$\Delta N^{(n)}(t_0, t_{\rm final}) = N^{(n)}_{\rm vacuum} - N^{(n)}_{\rm spike} \tag{5.11}$$

In our calculations, $t_{\rm final}$ is taken to be the time of merger For comparison amongst different models, we choose $n = 2$, as has been the strategy in previous studies. Eccentric binaries radiate in multiple harmonics and the dephasing will be larger for larger n. However, these higher harmonics are harder to detect.

We present a comparison of the evolution of the mean orbital frequency of the binary f, the absolute value of number of dephasing cycles of the 2nd harmonic $|N^{(2)}|$, and the evolution of eccentricity of the binary e as a function of its semi-major axis a using N-body models and `IMRIPy` for the non-rotating models in $\gamma_{\rm sp} = 7/3$

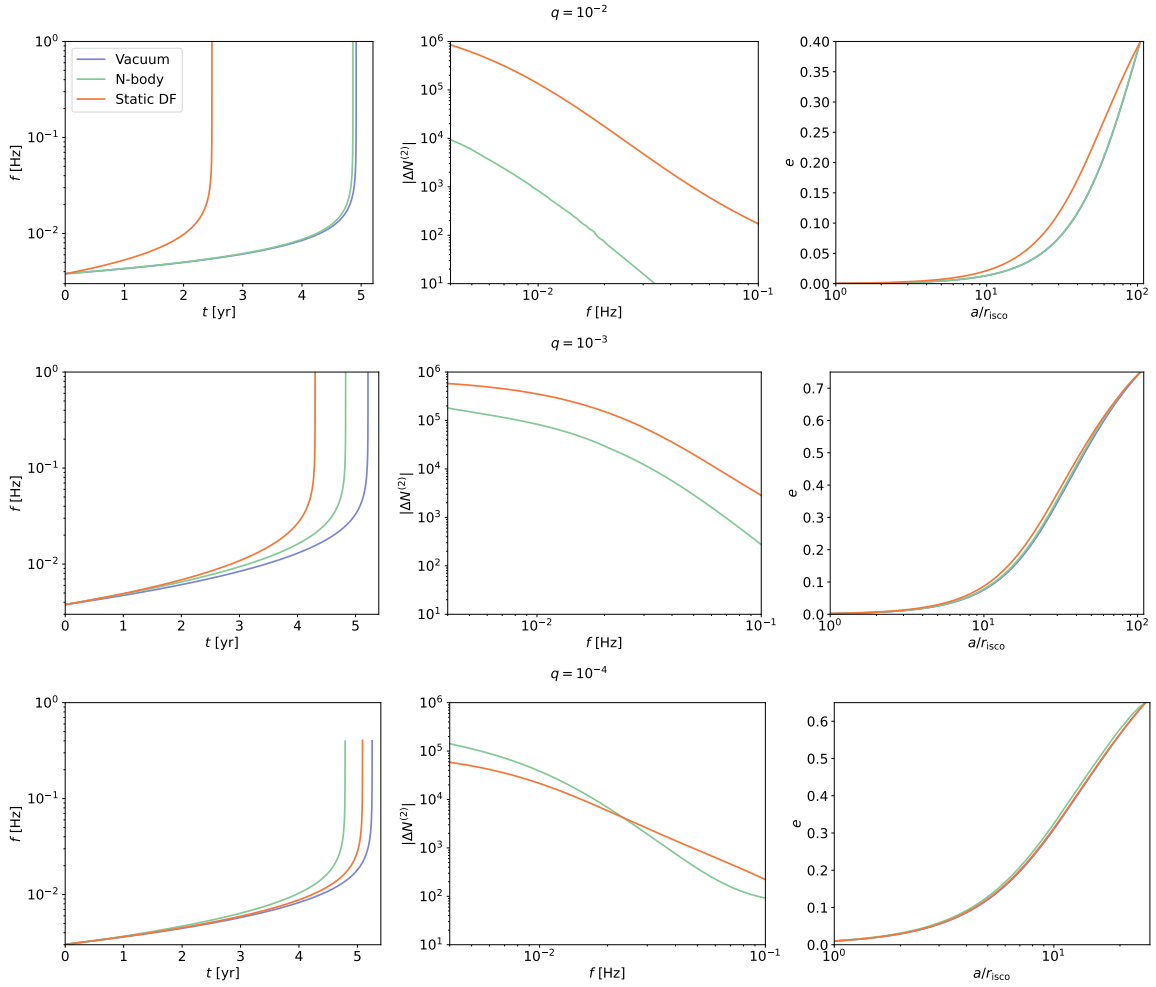

Figure 5.5: A comparison of the binary parameters for different mass ratio models evolving in a $\gamma_{\rm sp} = 7/3$ spike with $r_{\rm sp}$ calculated using equation 5.8. The evolution in vacuum (purple line) is calculated using Peters (1964b) analytic formula, while the evolution calculated using the Chandrasekhar DF formula assuming a static spike (orange line) is calculated using `IMRIPy`. They are compared to the evolution from our N-body simulations (green line). Left column: the mean orbital frequency f as a function of time t in years. Middle column: the estimated number of dephasing cycles of the second harmonic $|\Delta N^{(2)}|$ as a function of the frequency f in Hz. Right column: the eccentricity e as a function of the semi-major axis a. We notice that in higher mass ratio models, the evolution of the binary is similar to that in vacuum. For $q = 10^{-2}$ model, there is a $100\times$ reduction in the estimated number of dephasing cycles compared to the evolution calculated using Chandrasekhar DF formula as the spike has been disrupted in a very short time span. As we decrease the mass ratio, the disruption decreases. We notice that in case of the $q = 10^{-3}$ model, the dephasing is only reduced by $3\times$ compared to the evolution calculated using `IMRIPy`. For $q = 10^{-4}$ model, we find that dephasing is a factor of 3 larger than what we obtain using the Chandrasekhar formula, with little to no disruption of the spike. This signals that the Chandrasekhar DF might be insufficient to explain the evolution of the binary in DM spikes.

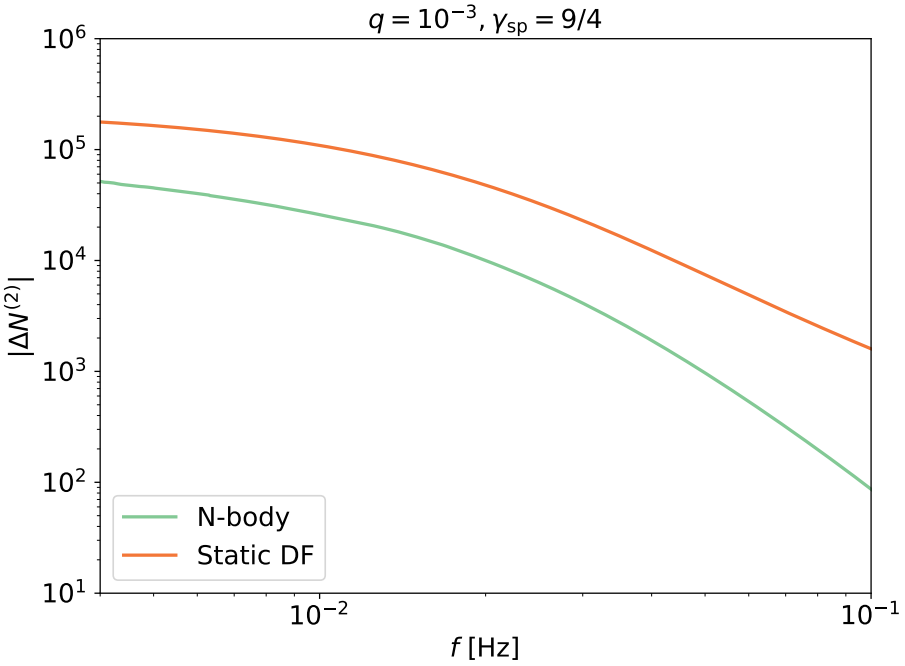

Figure 5.6: The dephasing of the second harmonic $|\Delta N^{(2)}|$ as a function of the binary frequency f in Hz for a $q = 10^{-3}$ binary embedded in a $\gamma_{sp} = 9/4$ spike. The color scheme is the same as that used in Figure 5.5. We notice that similar to the $\gamma_{sp} = 7/3$ case, the amount of dephasing in our N-body models is reduced by $\sim 3\times$ compared to the DF models.

spike with r_{sp} calculated using equation 5.8 in Figure 5.5. Examining the $q = 10^{-2}$ model, our N-body simulations predict that the binary merges 19 days earlier if it is embedded in a DM spike than in vacuum. It takes about $\approx 10^4$ fewer GW cycles than in vacuum to merge. Comparing the N-body model to the DF model with a static spike, we find a $100\times$ reduction in the number of dephasing. The binary disrupts the spike almost completely within the first 0.1 years of inspiral leading to a drastic reduction in the amount of dephasing. This is unsurprising as the mass ratio of the binary is large which injects a substantial amount of energy ejecting DM particles around the secondary leading to the disruption of the spike. As such, the circularization rate of the binary is similar to that in vacuum as is evident from the $a - e$ plot.

As we lower the mass ratio to $q = 10^{-3}$, the binary merges earlier. It takes about 141 fewer days to merge in the $\gamma_{sp} = 7/3$ N-body model than its vacuum counterpart. This creates a larger amount of dephasing. In a $\gamma_{sp} = 7/3$ spike, the dephasing with respect to vacuum is $\approx 2 \times 10^5$, only $3\times$ lower than what is predicted by the DF with static spike model. Although not directly comparable, we highlight the fact that this

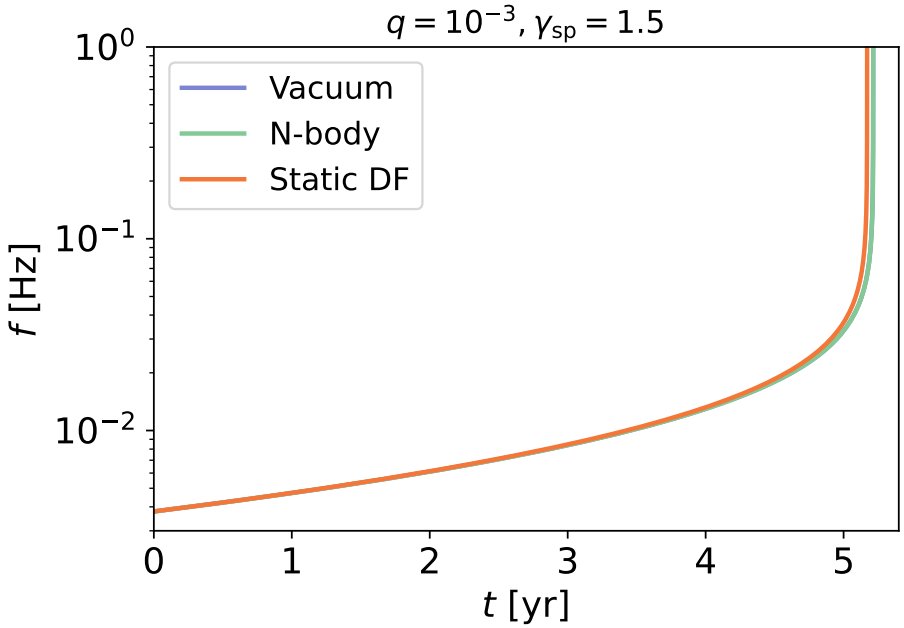

Figure 5.7: Binary frequency f as a function of time t for the $q = 10^{-3}$, $\gamma_{sp} = 1.5$ model. We find that the low density of the spike results in no discernable differences between the inspiral in vacuum and that in the spike. This results in $\leq \mathcal{O}(10)$ cycles which might not be detectable.

is much larger than the 100× reduction for the same mass ratio found by Kavanagh et al. (2020) in the semi-analytic `HaloFeedback` models but with the secondary on circular orbit. Since the self gravity of the spike is minimal, we should expect a $\sim 3\times$ reduction in the amount of dephasing compared to the static DF models irrespective of the density profile. To verify that, we compare the dephasing cycles in the $\gamma = 9/4$ model with $q = 10^{-3}$ in Figure 5.6. We find that $|\Delta N^{(2)}| \approx 1.8 \times 10^5$ for the static DF models whereas $|\Delta N^{(2)}| \approx 5.2 \times 10^4$, consistent with the $\sim 3\times$ reduction we expected. The rate of circularization is quite similar between the vacuum, static DF, and the N-body models. Nevertheless, we find that the eccentricity of the binary as a function of its semi-major axis for the N-body models lies approximately in between the static DF and the vacuum cases.

For the $\gamma_{sp} = 1.5, q = 10^{-3}$ case representing perturbed DM "crests", our findings are less optimistic. We note that such crests can also be formed due to adiabatic growth in an off-center IMBH (Zhao and Silk, 2005). In Figure 5.7 we plot the evolution of the binary frequency f as a function of time for this model. We find that the frequency evolution in the N-body models practically overlaps with that from the vacuum model. Unsurprisingly, the calculated dephasing is $\leq \mathcal{O}(10)$ which is larger than the limits of uncertainty from our N-body simulations. Such a small dephasing

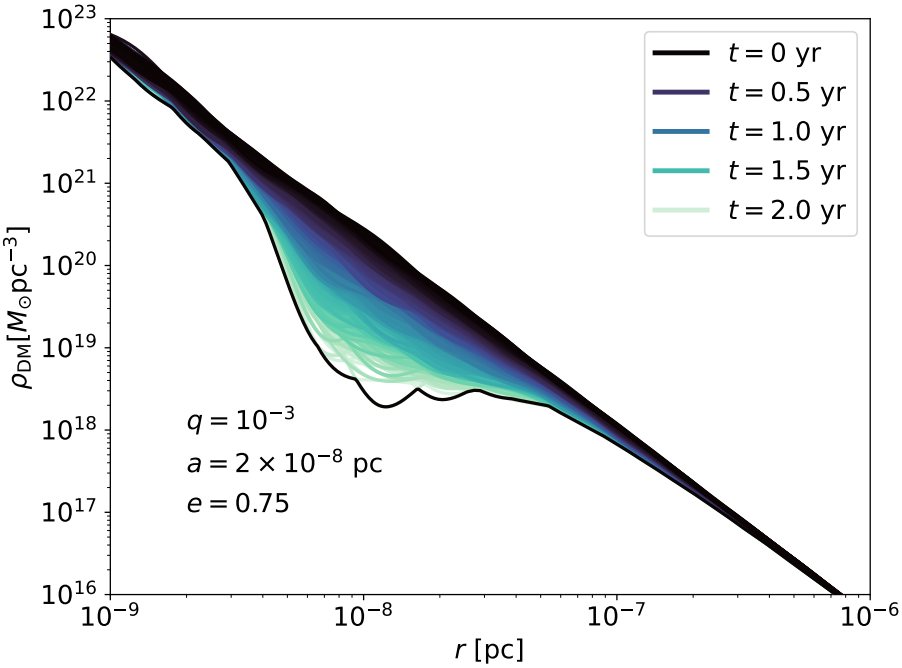

Figure 5.8: The density of the DM spike $\rho_{\rm DM}$ as a function of the distance r from the primary in pc for the $q = 10^{-3}$ model in a $\gamma_{\rm sp} = 7/3$ spike. The different colors represent the evolution of the density profile over time with darker colors representing earlier times and lighter colors representing later times. We notice that the inspiral of the IMRI leads to a significant disruption in the spike near the semi-major axis a of the IMRI. The density of DM in that region is reduced by a factor of 100 compared to the original density after an inspiral time of 2 years. This is qualitatively consistent with the findings from Kavanagh et al. (2020). However, we find that the disruption to the spike due to the IMRI in our N-body is much slower than what is found by Kavanagh et al. (2020) who find that the spike is significantly disrupted within 0.1 years.

would also be much harder to detect than those from the denser spike models. Since the formation of the DM crests are more robust than that of DM spikes, our findings pose an important question of whether the spike models are overly optimistic. This is discussed further in the section 5.5.3.

We examine the evolution of the density profile in the $q = 10^{-3}$, $\gamma_{\rm sp} = 7/3$ model over time due to the effect of the binary in Figure 5.8. The binary injects energy into the spike by scattering the DM particles, thereby reducing the density of the spike. It preferentially interacts and scatters particles near itself, so we expect a drop in the density profile near the binary over time. The local efficiency of scattering will depend on the orbital parameters of the binary and density profile near the

secondary. During a 2 year time span, we notice that the binary has drastic effects on the density of the spike. There is up to a factor of 100 reduction in the density of DM particles near the initial semi-major axis of $a = 2 \times 10^{-8}$. The density of DM particles reduces from $\sim 10^{20} M_{\odot}\mathrm{pc}^{-3}$ to $\sim 10^{18} M_{\odot}\mathrm{pc}^{-3}$. The binary effectively carves out a flat core near its semi-major axis. Since the secondary spends a larger time near its apoapsis than its periapsis, we expect the scattering to be most effective near the apoapsis carving out a core. The apoapsis of the secondary is 3.75×10^{-8} pc. We notice that the flat core extends from $\sim 10^{-8}$ pc to 4×10^{-8} pc consistent with the hypothesis. The impact further away, and near the periapsis is smaller. This has important implications for the survival and detectability of DM spikes and is discussed at length in section 5.5.3. Similar findings were noted in Kavanagh et al. (2020) in case of circular binaries. However, Kavanagh et al. (2020) find that the disruption happens quite quickly, effectively within 0.1 yr from the beginning of the simulation for $q = 10^{-3}$, whereas in our case the disruption is much slower, by almost $20\times$. This leads to a much larger dephasing effect compared to the `HaloFeedback` models. Notably, the evolution of the density profile does not show the presence of a wake or a density "bump" behind the secondary, which is present in the self-consistent `HaloFeedback` models. According to the authors, the "bump" is caused due to scattering of DM particles near the secondary which give rise to the DF effect. The absence of the "bump" and the longer spike disruption time suggests that DF theory is inconsistent with the results from our N-body simulations. Additionally, we find that the effect of softening is minimal on the feedback time and effects on the density profile. When a softening of $\approx 10^{-10}$ pc is used between the secondary and DM particles, we find that the feedback at $r < 10^{-8}$ pc is decreased but the differences between the density profiles in that region between non-softened and the softened sims never vary substantially.

For our $q = 10^{-4}$ non-rotating model with $r_{\mathrm{sp}} = 1.17$ pc, examining Figure 5.5, we notice a very surprising result. While the binary in the static DF model merges 60.3 days faster then the vacuum case, the binary in the N-body model merges 169 days faster. This leads to dephasing effects that are almost $3\times$ larger than the DF model. Whereas in the static DF case we find that $|\Delta N^{(2)}| \approx 6 \times 10^4$, our N-body simulation suggests that $|\Delta N^{(2)}| \approx 1.5 \times 10^5$. Since the results are extrapolated beyond 2.5 years using the Peters (1964b) equation, we compared them to those obtained from set of lower resolution simulations which are run to completion. We find that the extrapolated results are consistent with those from the lower resolution simulations. The extrapolated results fall within a standard deviation of the results from the lower resolution simulations. For the models where we fix $r_{\mathrm{sp}} = 0.54$ pc, similar to the $q = 10^{-2}$ and 10^{-3} models, we notice a decrease in the amount of dephasing compared to the $r_{\mathrm{sp}} = 1.17$ pc model. We present the dephasing in $q = 10^{-4}$, $r_{\mathrm{sp}} = 0.54$ pc model in Figure 5.9 and find that the N-body and the `IMRIPy` simulations predict a similar amount of dephasing, with $|\Delta N^{(2)}| = 3 \times 10^4$. This is $\approx 6\times$ lower than that obtained in the $r_{\mathrm{sp}} = 1.17$ pc models which is unsurprising given the lower density in

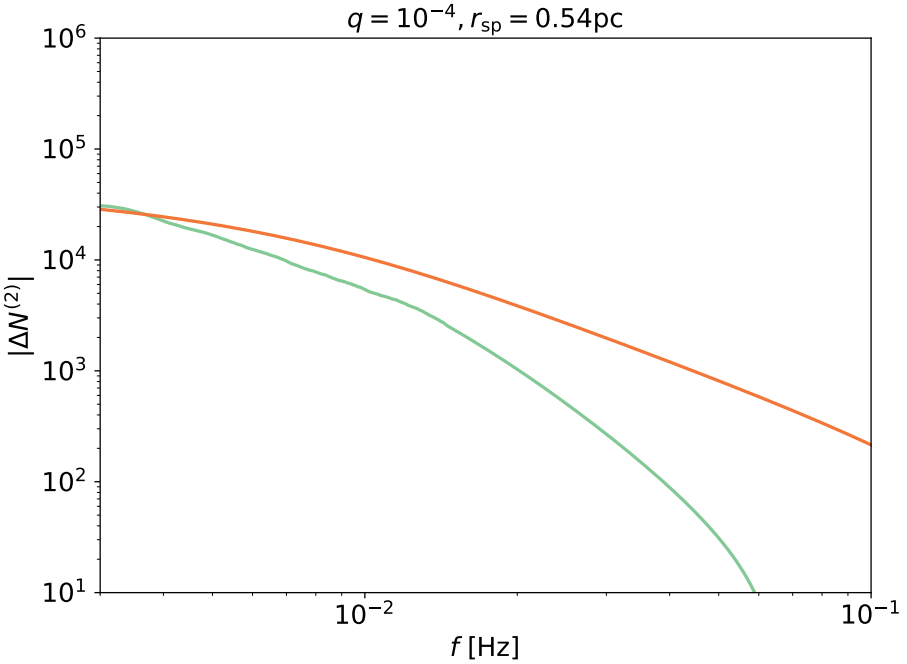

Figure 5.9: Similar to Figure 5.6 but for $q = 10^{-4}$ binary embedded in a $\gamma_{\rm sp} = 7/3$, $r_{\rm sp} = 0.54$pc spike. We notice that the net dephasing predicted by the N-body simulation is comparable to that predicted by the `IMRIPy simulation`, although at higher frequencies the dephasing falls off faster in the N-body models. The density in the region of interest for this particular model is about 6 times lower than that for the $q = 10^{-4}$, $r_{\rm sp} = 1.17$pc model and we notice a similar decrease in the amount of dephasing in this scenario compared to that model.

the $r_{\rm sp} = 0.54$ pc models. The density in the region of interest is 6.1× lower in the model with $r_{\rm sp} = 0.54$ pc compared to the model with $r_{\rm sp} = 1.17$ pc, which explains the factor of 6 reduction in the dephasing. This, along with the results from the $q = 10^{-3}$ models, strongly suggests that the dephasing is proportional to the local density of the profile at the position of the secondary. Similar to our findings from the $q = 10^{-3}$ model, we find the spike plays a very little role in dephasing if $\gamma_{\rm sp} = 1.5$ even when $M_1 = 10^4 M_\odot$. The dephasing is again $\leq \mathcal{O}(10)$, and below our uncertainty thresholds.

The results obtained for the $r_{\rm sp} = 1.17$ pc case for the $q = 10^{-4}$ binary is a factor of 15 greater than what was found by Kavanagh et al. (2020) in their `HaloFeedback` simulations for similar models on circular orbits. We also compare the spike density profiles at the beginning and end of our simulation and find that there is no degradation in the density profile of the spike, while the reduction of the DM density is observed in the `HaloFeedback` models even with such low mass ratios. This, along with our previous results, suggests that previous analytic/semi-analytic calculations using DF underestimated the total number of dephasing cycles by a factor of one to ten (or even larger) for lower-mass ratio binaries. This is especially evident while comparing the N-body simulations to the `HaloFeedback` simulations, where we find a difference of ten to hundred times. We use three-body simulations in section 5.4.3 to show that three-body scattering provides a better description of the dynamics of the binary in the spike and could explain the inconsistency observed in this section, while showing that the impact of DF is subdominant.

5.4.2 Rotating models

We now move on to the analysis of the binary inspiral in the rotating models. From DF theory, we would expect the inspiral of the binary in counter-rotating or retrograde models to be slower than that in co-rotating or prograde models. This is caused due to the fact that in the prograde models, the relative velocity of the DM particles is smaller than that in the retrograde models. Since the force of DF is inversely proportional to the relative velocity of the interacting particles, we would expect the DF force to be larger in the prograde scenario, resulting in a faster inspiral.

Examining the properties of the binary over time in both the prograde models and retrograde models in Figure 5.10 we notice a very surprising result. We find that in all of the cases, the retrograde models merge faster than the prograde and even the non-rotating models. Even more astonishing is the fact that the prograde models are slower to inspiral than the vacuum models. This is the opposite of what we would have expected from DF. Examining the $q = 10^{-2}$ models, we find that the expected amount of dephasing increases by almost 4× compared to its non-rotating counterpart. Since we are plotting the abolute value of the number of dephasing cycles with respect to vacuum, we find that the prograde model has a larger amount of dephasing, almost 1.5× that of the non-rotating model. However, we want to stress

that the prograde model actually take longer to inspiral than the vacuum model. So it actually takes 1.5×10^4 more GW cycles than the vaccum model to merge.

As we lower the mass ratio, we notice the same trend. In the $q = 10^{-3}$ models, we find that $|\Delta N^{(2)}| \approx 7 \times 10^5$ in the retrograde model, whereas $|\Delta N^{(2)}| \approx 2 \times 10^5$ in the non-rotating model, almost $3.5\times$ lower. Similar to the 10^{-2} model, we find that the prograde model takes 4.5×10^5 more GW cycles to merge than the vacuum model. Interestingly, we also notice that the retrograde model eccentrifies quickly in the beginning, before circularizing later due to GW effects. The opposite is observed in prograde models. There is an accelerated circularization in the beginning, followed by a period of circularization led by the emission of GW.

As we lower the mass-ratio to $q = 10^{-4}$, we notice that the eccentrification and circularization effects increase. Examining the eccentricity as a function of the semi-major axis, we find that the evolution in the retrograde and prograde models diverge quickly. The binary eccentrifies quickly, reaching a maximum eccentricity of 0.67 before circularizing. The opposite happens in the prograde models where the binary circularizes quickly, within the first one year time span, reaching an eccentricity of 0.61. This eccentrification and circularization lead to accelerated or deccelerated inspiral due to GW emission since the GW emission is very sensitive to the binary eccentricity. This leads to the larger dephasing we see in retrograde models. We find that the retrograde model merges about 550 days earlier compared to the vacuum model, whereas the prograde model takes 484 days longer. In the retrograde models, this produces a dephasing effect that is 2.5 times as large as the non-rotating models, with the binary taking 5×10^5 fewer cycles to merge compared to the vacuum models. The opposite effect is noted in the prograde models with the binary taking 3.5×10^5 more cycles to merge. Interestingly, in all of our simulations we find that the ratio of dephasing cycles in retrograde rotating to non-rotating models is between $2.5 - 3$. Similarly for prograde models, we find that the ratio of dephasing cycles in rotating to non-rotating models is about $1.5 - 2$, independent of the density profile and mass ratio. This is likely to be an artifact of our chosen initial conditions rather than a fundamental property of rotation itself.

We also compare the density profiles of the non-rotating to the rotating models over time to understand if the feedback from the binary to the spike changes upon the inclusion of rotation. Surprisingly, we find that the density profiles are consistent among the rotating and non-rotating models. Although rotation changes the dynamics of the binary significantly, the effect of binary on the spike is consistent among non-rotating and rotating models.

5.4.3 Three-body simulations

The results obtained in the previous section suggest that DF theory is insufficient at explaining the long term dynamics of the binary. In fact, this is not a very surprising result as it has been known that upon the formation of a hard binary, three- body

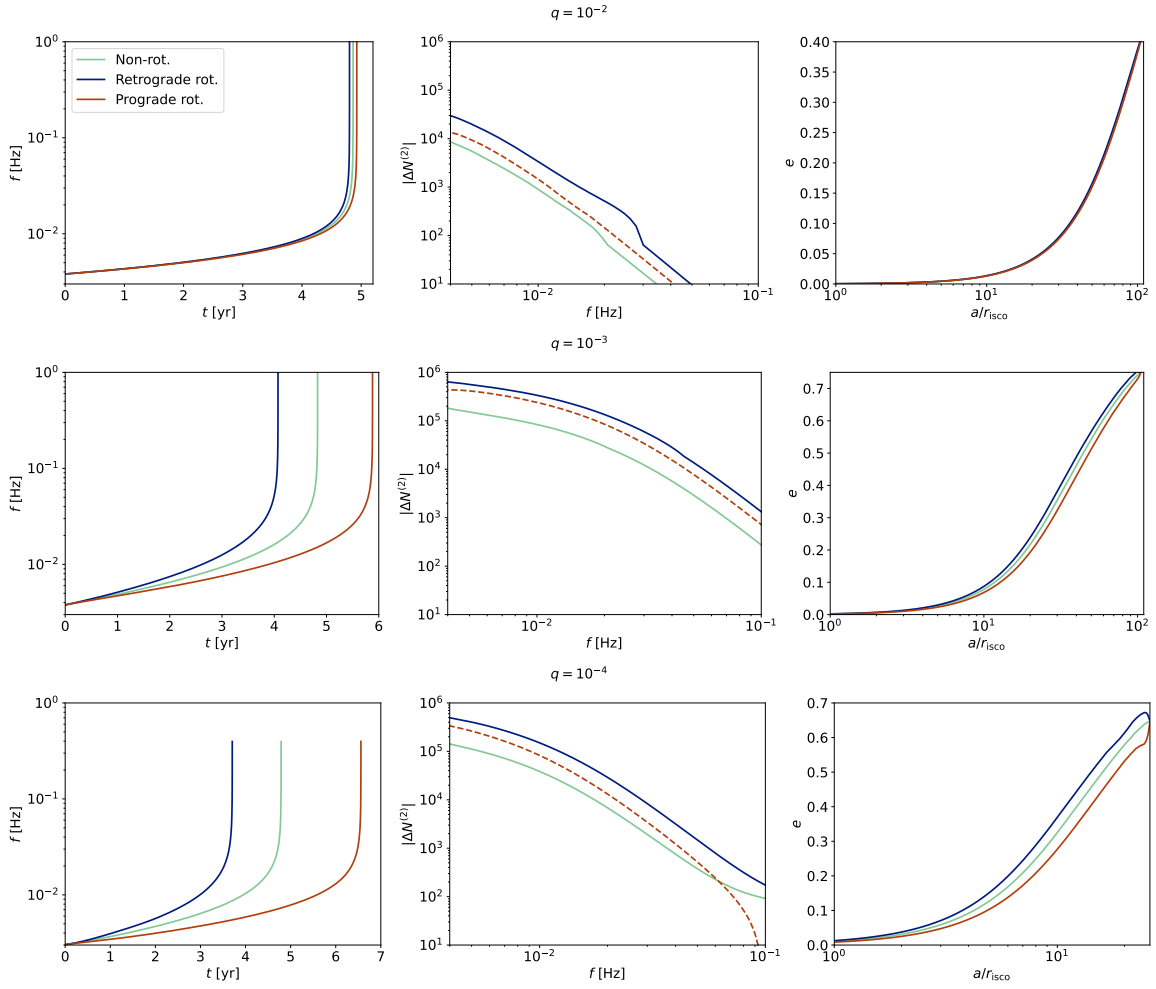

Figure 5.10: A comparison of the binary evolution parameters for different mass ratio models inspiraling in a $\gamma_{\rm sp} = 7/3$ spike similar to Figure 5.5 but also including the evolution from rotating models. We notice that the in all of the retrograde models, the binary merges faster than in the prograde and even the non-rotating models. This leads to a major enhancement in the number of dephasing cycles, by as much as $3\times$ that of the non-rotating model in the case of $q = 10^{-4}$. Since we are plotting the absolute value of the number of dephasing cycles, the dephasing in case of the prograde models appears to be positive. However, the prograde models actually merge slower than even the vacuum models and $\Delta N^{(2)}$ is actually *negative* (indicated by the dashed lines) . Thus the number of GW cycles in prograde models is larger than that in vacuum. We also find that in the retrograde models, the binary circularization rate is slower than that in the prograde models. In the latter scenario, the circularization is enhanced leading to a slower inspiral. As the mass ratio decreases, the effect of rotation becomes more prominent suggesting that transfer of angular momentum between the spike and the IMRI contributes majorly and cannot be ignored.

scattering becomes more effective than DF at dissipating energy, similar to SMBH binaries (e.g., Begelman et al., 1980; Merritt, 2013a). A binary is said to become a hard binary when the separation between the primary and secondary falls below the hard binary radius $a_{\rm h}$ which is defined as

$$a_{\rm h} = \frac{q}{(1+q)^2}\frac{d_{\rm infl}}{4}, \tag{5.12}$$

where $d_{\rm infl}$ is the influence radius of the primary and is defined as the radius that encloses twice the mass of the primary (e.g., Mukherjee et al., 2023). For our models, we find that $a_{\rm h}$ ranges from 6.75×10^{-3}pc in the $q = 10^{-2}$ model to 1.46×10^{-4}pc in the $q = 10^{-4}$ model. Therefore, all of our binaries are in the hard binary limit at the beginning of our simulations.

In the three-body scattering scenario, the binary undergoes a complicated three-body interaction in which the incoming particle interacts with the binary multiple times. This results in the ejection of the particle eventually, leading to shrinkage of the binary's orbit. This process has been called the *slingshot mechanism* and is fundamentally different from Chandrasekhar DF which is a cumulative effect of two-body encounters.

To understand which of the two above-mentioned processes dominate, we can analyze the N-body simulations and compare the efficiency of three-body scattering to that of dynamical friction. The DM particles are orbiting the primary and are bound to it initially. This allows us to obtain the semi-major axis $a_{\rm DM}$ and eccentricity $e_{\rm DM}$ of each particle. From our $q = 10^{-3}$ model with $\gamma_{\rm sp} = 7/3$, we select a subset of DM particles which satisfy $0.5a_0 \leq a_{\rm DM} \leq 1.5a_0$, where a_0 is the initial semi-major axis of the binary. In principle, all particles interact with the binary, but particles with semi-major axis close to the binary's interact strongly. We, then, calculate the initial and final energies of the selected particles at the beginning and end of our simulations to calculate the change in energy of the DM particle ($\Delta E_{\rm DM}$). We find that all of the selected particles are ejected from the spike by the end of the simulation.

The selected particles are evolved from their initial positions along with the binary individually using `IAS15` for the same duration as the full N-body simulations. The energy of each particle is recorded at the beginning and end of the simulation. The difference in energies provides an estimate of the change in energy due to three-body effects. On the other hand, calculating the energy dissipation due to dynamical friction from the N-body simulations is somewhat non-trivial. According to Chandrasekhar (1943), as M_2 moves through the medium of DM particles, it experiences a number of two-body encounters that change the velocity of the secondary in a direction parallel and opposite to the initial velocity of the secondary. Each particle contributes a net change in the velocity of the secondary giving rise to dynamical friction force over time. We can estimate the dynamical friction force due to each individual particle using the method described in Ma et al. (2023). The energy change

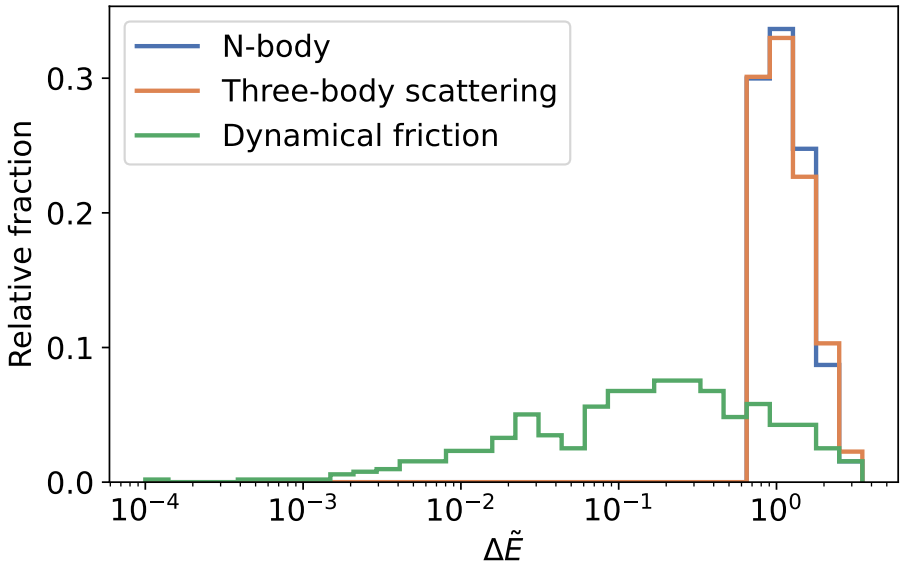

Figure 5.11: The distribution of the normalized energy change $\Delta\tilde{E}$ of strongly interacting DM particles from our N-body simulations compared to that from three-body scattering simulations and dynamical friction approximation. We find that the distribution of energy change of strongly interacting particles is consistent between the N-body simulations and the three-body simulations but not with the calculated values using the dynamical friction approximation. This confirms our hypothesis that three-body scattering, and not dynamical friction is responsible for dissipating energy from the binary.

due to dynamical friction attributed to the $i^{\rm th}$ DM particle $\dot{E}_{\rm i,df}$ can be written as

$$\dot{E}_{\rm i,df} = M_2 \mathbf{a}_{\rm i,df} \cdot \mathbf{v}_2 \tag{5.13}$$

where $\mathbf{a}_{\rm i,df}$ can be calculated using equation 9 from Ma et al. (2023) and $\mathbf{v}_2$ is the velocity of the secondary. We calculate $\dot{E}_{\rm i,df}$ using the saved snapshots and integrate over time to find the net energy change due to dynamical friction. The results from our simulations are stored at a fine enough time resolution that this method is able to provide a good approximation of the dynamical friction energy loss.

We present the relative distribution of the energy change of the selected particles from the N-body simulation, and compare it to the energy change due to three-body effects and dynamical friction in Figure 5.11. For clarity, we present the normalized change in energy $\Delta\tilde{E}$ where we normalize the change in energy with respect to the binding energy of the DM particle with $a_{\rm DM} = a_0$. We find that the distribution of the change in energy from the N-body simulations matches that from the three-body simulations but is inconsistent with the energy change due to dynamical friction.The N-body and three-body simulations show that each selected particle experiences, on

average, a normalized energy change of 1.0-1.5. In the dynamical friction scenario, the average energy loss per particle is about 10× lower. Additionally, we find that, akin to the N-body simulations, all particles are ejected from the system in the three-body simulations. From the estimates of the dynamical friction force, we find that only 10-20 per cent of the particles would be ejected from the system entirely, consistent with the findings from Kavanagh et al. (2020), but is in tension with the findings from the N-body simulation. This confirms our hypothesis that three-body scattering, and not dynamical friction is the predominant method of energy dissipation in our simulations. Although not presented here, we find a similar story across all of the models used in this study. Interestingly, as the mass ratio is decreased, the fraction of energy loss due to the three-body scattering increases. This is consistent with Merritt (2013a) where the three-body scattering efficiency is proportional to q^{-1}. We find a similar distribution in the change in energy of the particles from both the three-body and full N-body simulations in our $q = 10^{-4}$ models as well. Our results also highlight that our N-body simulations are robust, accurate, and consistent with results from the extremely accurate `IAS15` integrator where the net energy is conserved to machine precision.

What causes the counter intuitive results from our rotating simulations? Merritt et al. (2009), Iwasawa et al. (2011), and Sesana et al. (2011) hold clues that are able to shed some light on this mystery. During a three-body encounter, the binary exchanges energy and angular momentum with the particle in a complicated fashion. Iwasawa et al. (2011), in mergers of MBH binaries in galactic nuclei, noted that counter-rotating stars are much more effective in extracting angular momentum from the binary during the three-body scattering phase. As explained by Merritt et al. (2009) and Sesana et al. (2011), this is caused due to the torquing mechanism of the binary's potential on to the particle which leads to a secular evolution of the particle's eccentricity and inclination. This mechanism converts the counter-rotating particles to co-rotating which are then preferentially ejected by the binary (Iwasawa et al., 2011). This results in a larger change in angular momentum of the particle and as such, counter-rotating particles are able to extract angular momentum from the binary more efficiently compared to co-rotating particles. This process becomes more efficient for more eccentric binaries.

To understand if the mechanism mentioned above is able to explain the counter-intuitive results from the previous sections and to provide a better description of the dynamics of the binary, we perform three-body simulations. This is done to understand the transfer of energy and angular momentum during the scattering of a DM particle by the binary. We follow the similar steps as in Sesana et al. (2011) to set up our simulation with some differences.

We first generate a $N = 1M$ particle realization using `Agama` containg a primary with mass $M_1 = 10^3 M_\odot$ and $\gamma_{\rm sp} = 7/3$ and then use the Lynden-Bell trick to generate prograde and retrograde models. We then place the secondary with mass $M_2 = 10 M_\odot$ at $a_0 = 2 \times 10^{-8}$ pc with varying eccentricity. We ensure that the binary lies in the

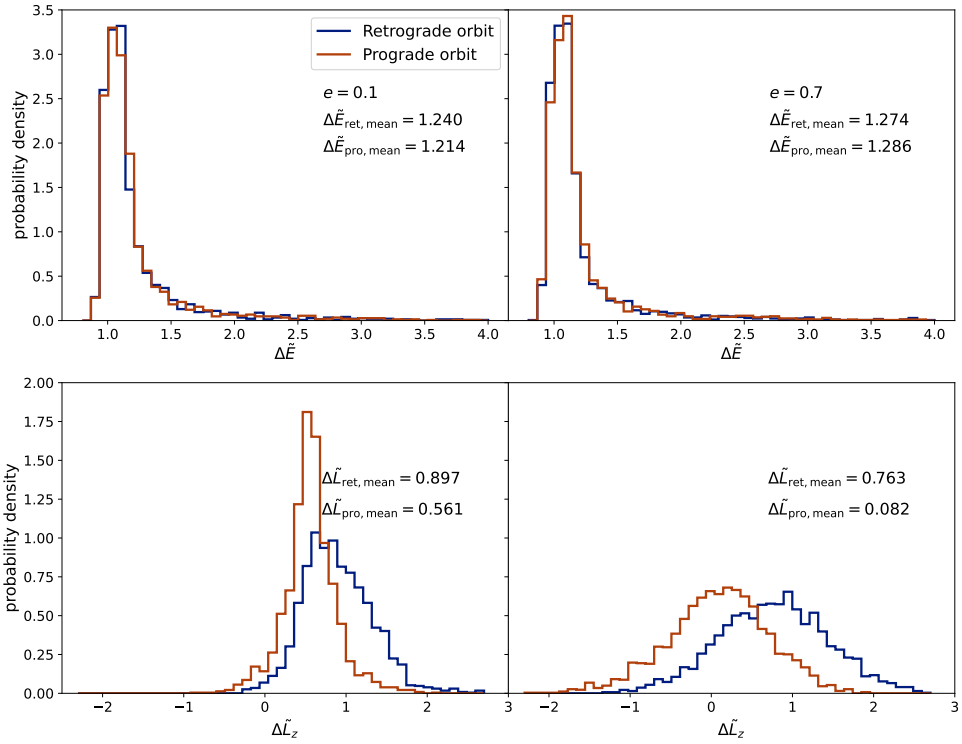

Figure 5.12: Top: probability distribution of the normalized energy change $\Delta\tilde{E}$ of 2500 DM particles from the three-body simulations performed until the particle is ejected by interactions with the binary for both prograde and retrograde rotation models with two different binary eccentricities e. We notice that in both prograde and retrograde models, the distribution of the normalized energy is similar and does not change with the binary eccentricity. Bottom: similar to top but the probability distribution of the normalized angular momentum $\Delta\tilde{L}_z$ of 2500 DM particles after ejection. We notice that there are major differences in the distribution of $\Delta\tilde{L}_z$ between prograde and retrograde models. In particular, in the retrograde models, the angular momentum of the particle is larger than that in the prograde models, especially for more eccentric binaries. This suggests that retrograde particles are more efficient at "stealing" angular momentum from the binary, especially at larger e.

$x-y$ plane. From the generated initial model, we only select DM particles which have $a_{\rm DM} \approx a_0$. This results in ~ 2500 DM particles being selected. We, then, evolve each particle with the binary individually, running 2500 three-body simulations using `IAS15`, until the DM particle is completely ejected from the system. We record the initial and final values of the energy and angular momentum of the particle and the binary. We note that the simulations are purely Newtonian.

The eccentricity of the binary can be written as follows:

$$e = \sqrt{1 - \frac{2EL^2}{GM^2\mu^3}} = \sqrt{1 - \frac{2EL_z^2}{GM^2\mu^3}} \tag{5.14}$$

where E is the binding energy of the binary, L is the angular momentum of the binary, $M = M_1 + M_2$, and $\mu = \frac{M_1 M_2}{M_1+M_2}$. The last equality follows from the fact that our binary is in the $x-y$ plane. As such, $L = L_z$. Differentiating equation 5.14 as is done in Sesana et al. (2011) provides us with the change in eccentricity Δe, which we find to be:

$$\Delta e = -\frac{1-e^2}{2e}\left(\frac{\Delta E}{E} + \frac{2\Delta L_z}{L_z}\right) = \frac{1-e^2}{2e}\chi \tag{5.15}$$

where ΔE is the change in E, ΔL_z is the change in L_z and χ, called the eccentrification parameter, has been defined using the last equality. We find that for $\chi > 0$, the binary becomes more eccentric after the encounter with the DM particle, and for $\chi < 0$, the binary becomes more circular. Since the energy and angular momentum in the simulation are conserved to machine precision, we expect that the change in energy and angular momentum of the binary are equal and opposite to the change in energy and angular momentum of the DM particle. Therefore,

$$\Delta E_{DM} = -\Delta E \tag{5.16}$$

$$\Delta L_{z,DM} = -\Delta L_z \tag{5.17}$$

For the sake of clarity, we use the normalized change in energy $\Delta\tilde{E}$ (as defined previously), change in angular momentum $\Delta\tilde{L_z}$, and eccentrification parameter $\tilde{\chi}$ in our calculations which are defined as follows:

$$\Delta\tilde{L_z} = \frac{\Delta L_{z,DM}}{L^c_{z,DM}} \tag{5.18}$$

$$\tilde{\chi} = \chi\frac{M_2}{m_{DM}} \tag{5.19}$$

where $L^c_{z,DM}$ is the angular momentum of a DM particle on a circular orbit with $a_{DM} = a_0$, and E_{DM} is the binding energy of the DM particle.

To understand how the retrograde and prograde families contribute differently to the change in energy and angular momentum of the binary, we plot the distribution

of $\Delta\tilde{E}$ and $\Delta\tilde{L_z}$ for a binary with $e = 0.1, 0.7$ in Figure 5.12. Quite surprisingly, we find that the energy exchanged during the scattering process is similar between the retrograde and prograde models with relative differences of at most $1-2$ percent between them. This is in tension with the DF theory where we would expect the retrograde family to have a much lower final energy than the prograde family owing to the larger relative velocity in the former case. $\Delta\tilde{E}$ is always positive signaling that *all* particles contribute to dissipating energy from the binary, even fast moving particles with velocities larger than the velocity of the secondary, in contrast with the assumptions made in previous studies (e.g., Kavanagh et al., 2020; Becker et al., 2022). Additionally, we note that the distribution of energy exchanged does not change substantially as the eccentricity is increased. Although not presented here, we find that the mean energy exchanged $\Delta\tilde{E}_{\rm mean}$ remains almost constant across all values of e, consistent with the findings from Sesana et al. (2011).

Contrary to the distribution of $\Delta\tilde{E}$, we find major differences between the retrograde and prograde models while comparing the distribution of $\Delta\tilde{L_z}$. We find that the distribution of angular momentum exchanged always skews towards lower values for prograde models compared to the retrograde models. The mean angular momentum exchanged is almost 37% lower for the prograde family than the retrograde family when $e = 0.1$. As the eccentricity increases, the differences become even more drastic. At $e = 0.7$, the mean angular momentum exchanged by the retgrograde family is almost $10\times$ larger than the prograde family. Thus, we find that counter-rotating particles are more efficient at stealing angular momentum from the binary than co-rotating particles, especially for more eccentric binaries.

We calculate the mean values of $\Delta\tilde{E}$ and $\Delta\tilde{L_z}$ across different binary eccentricities to calculate the eccentrification parameter using equation 5.15. We present the normalized eccentrification parameter $\tilde{\chi}$ as a function of e in Figure 5.13 for both the retrograde and prograde models. Consistent with Sesana et al. (2011), we find that $\tilde{\chi}$ is *always positive* in retograde models while it is *always negative* in prograde models. Thus, in our retrograde models, the binary always undergoes eccentrification as a result of interaction with the spike, whereas the binary undergoes circularization in the prograde scenario. This is consistent with findings from Fokker-Planck models of binary evolution in rotating nuclei. According to Rasskazov and Merritt (2017) equation 84a,

$$\langle \Delta e \rangle = \frac{\rho G a K H}{\sigma} \tag{5.20}$$

where $\langle \Delta e \rangle$ is the mean change in eccentricity, H is the dimensionless binary hardening rate, and K is the dimensionless eccentricity growth rate (Quinlan, 1996; Sesana et al., 2006). Using Fokker-Planck methods, Rasskazov and Merritt (2017) find that, in rotating nuclei,

$$K = 1.5e(1-e^2)\left(0.15 - (2\eta - 1)\cos(\theta)\right) \tag{5.21}$$

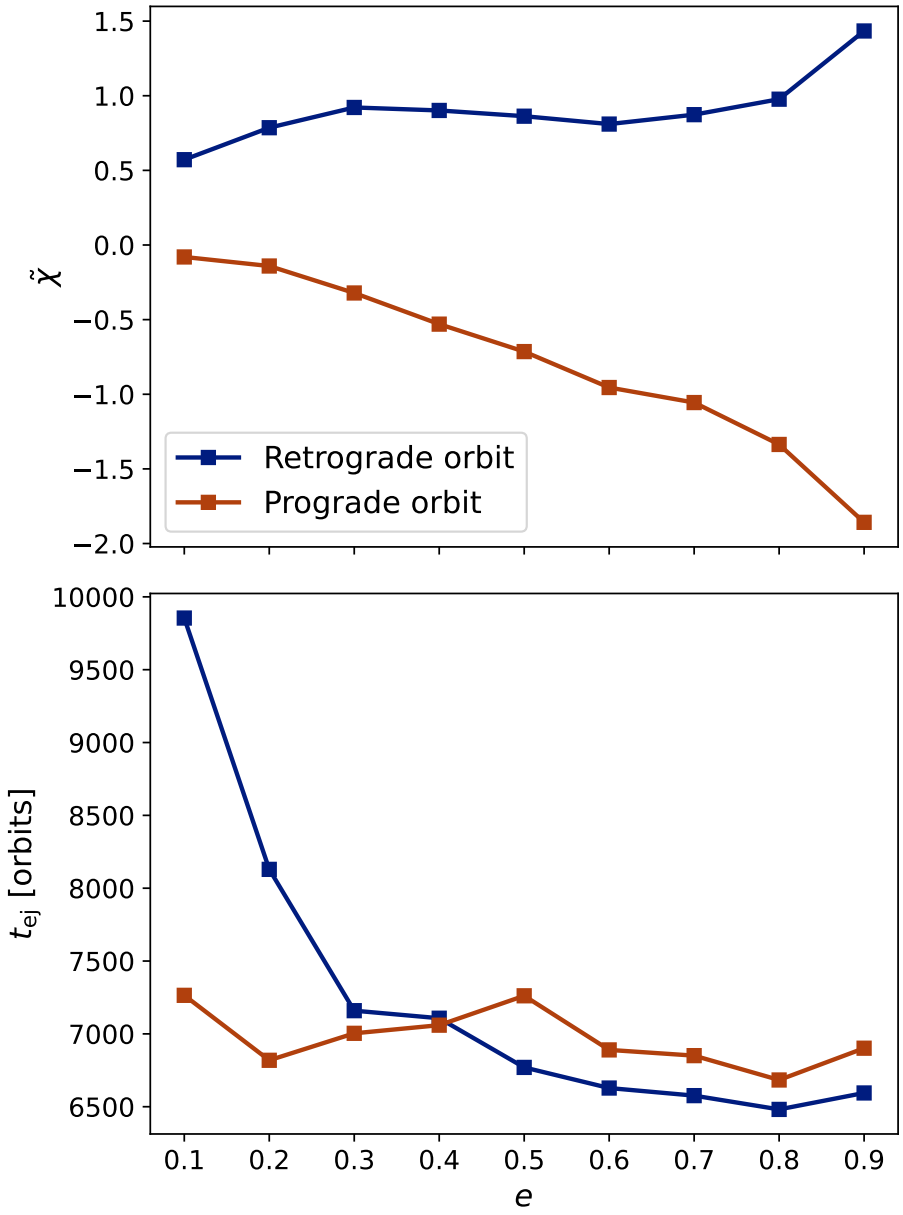

Figure 5.13: Top: the normalized eccentrification parameter $\tilde{\chi}$ as a function of the eccentricity e of the binary. Bottom: the mean time of ejection of the DM particle $t_{\rm ej}$ in units of the number of orbits of the binary as function of the eccentricity of the binary. We notice that for retrograde models $\tilde{\chi}$ is always positive and always negative for prograde models. This results in an eccentrification of the binary in case of retrograde rotation and circularization of the binary in case of prograde rotation. We also notice that the time of ejection of the DM particle is almost uniform for the prograde case across different values of e. For retrograde scenario, for larger values of e, $t_{\rm ej}$ is similar to that in the prograde case. However, it rapidly increases as e decreases.

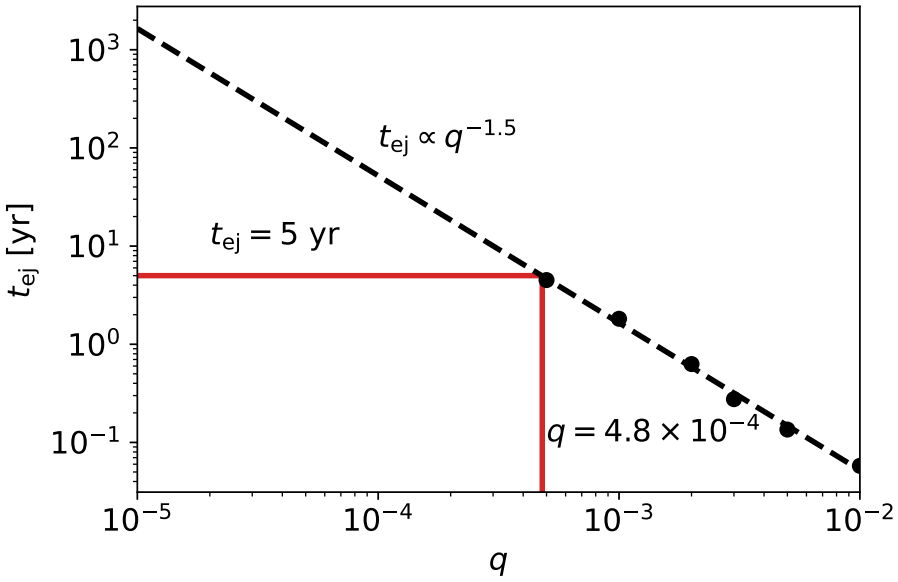

Figure 5.14: The mean ejection time $t_{\rm ej}$ as a function of the binary mass ratio q. The dots represent the values calculated from the three-body simulations whereas the dashed line represents the best fit to the datapoints. We find that $t_{\rm ej} \propto q^{-1.5}$. This implies that for any $q < 4.8 \times 10^{-4}$, $t_{\rm ej} > 5$ yr, larger than the inspiral time of a LISA detectable IMRI. For $q = 10^{-4}$ or lower, the ejection time is much larger implying that the backreaction on to the halo is expected to be minimal. This is consistent with the findings from our full N-body simulations for $q = 10^{-4}$ where we did not find any substantial feedback on the DM density profile.

where θ is the orbital inclination of the binary and η is the fraction of prograde to retrograde particles (as opposed to $\mathcal{F}$ which is the ratio of retrograde to prograde particles). Since our binary inclination is 0, $\cos(\theta) = 1$. For a retrograde rotation, $\eta = 0$ and we find, $K > 0$ so the binary eccentrifies as a result of interactions with DM particles. In the prograde case, $\eta = 1$ and $K < 0$ and the binary circularizes due to three body interactions. This explains why the binary eccentrifies or circularizes quickly at the beginning of our full N-body simulations. The rate of circularization is lower in the retrograde models than in the non-rotating and the prograde models, resulting in a faster inspiral as the effects of GW radiation are stronger at larger eccentricities. The inspiral is the slowest in prograde models as the binary circularizes faster than in non-rotating and even vacuum scenarios. While not presented here, we verified that our results are valid across the mass-ratios considered in this study.

One can also use the three-body simulations to understand how long it takes for the three-body interactions to disrupt the spike significantly. For a fixed mass ratio of $q = 10^{-2}$, looking at Figure 5.13, we find that for lower eccentricity binaries, the

mean ejection time $t_{\rm ej}$ is larger for particles on retrograde orbits than prograde orbits. As the eccentricity is increased, the ejection time becomes similar for both prograde and retrograde orbits. Since the particles are preferentially ejected when they are co-rotating with the binary, we postulate that difference between the prograde and the retrograde models across eccentricity is caused due to the fact that a less eccentric binary exterts weaker torques on the particle leading to a longer secular timescale over which the binary converts the particle from retrograde to prograde. For moderately or highly eccentric binaries, the conversion from retrograde to prograde rotation happens quickly, leading to a similar ejection time between the two models.

We can also study the ejection time as a function of the mass ratio. We take a non-rotating version of the model generated above and run multiple simulations with different mass ratio binaries with $a_0 = 2 \times 10^{-8}$pc, $e_0 = 0.7$ and plot the mean ejection time $t_{\rm ej}$ as a function of the mass ratio q in Figure 5.14. We find that the relationship between the ejection time and the mass-ratio can be described by a power-law. We find that $t_{\rm ej} \propto q^{-1.5}$. The proportionality constant is a function of the central IMBH mass, semi-major axis, and eccentricity of the binary.

As pointed in section 3.2, we caution the reader, that our models are somewhat unphysical as $\mathcal{F}$, the fraction of counter-rotating particles to the total number of particles, is 0 (prograde models) or 1 (retrograde models). The eccentrification or circularization sensitively depends on this fraction $\mathcal{F}$. Sesana et al. (2011) find that when this fraction is greater than 0.5, the binary eccentrifies as a result of three-body interactions. This suggests that in our non-rotating models, the binary undergoes mild eccentrification as a result of interactions with the spike. A systematic study of the change in the rate of circularization as a function of $\mathcal{F}$ is left for a future study.

The three-body simulations are easier to parallelize, and faster to run than the N-body simulations which can provide the foundation for a more extensive parameter space study in the future. We can also use the three-body simulations to derive the distribution of the energy and angular momentum changes as a function of the binary semi-major axis and eccentricity which can later be used in semi-analytic models like `HaloFeedback`. Future work might also involve considering the binary's angular momentum to be tilted relative to the angular momentum of the DM spike.

5.4.4 Precession effects

We present a brief analysis of the effects of relativistic precession on the results from both our rotating and non-rotating models. Relativistic precession is included by using the PN1 and PN2 terms in a set of rotating and non rotating $q = 10^{-3}$ models in $\gamma_{\rm sp} = 7/3$ spikes. We plot the mean orbital frequency of the binary f as a function of the inspiral time t in Figure 5.15 for the rotating and the non rotating models. Comparing the evolution of the non rotating model in Figure 5.15 to that in Figure 5.5, we find that there are minimal differences in the evolution with and without relativistic precession. On the other hand, we immediately notice that the inclusion

Model	$\lvert\Delta N^{(2)}\rvert$
Non rotating	2.4×10^5
Retrograde rotation	2.6×10^5
Prograde rotation	1.9×10^5

Table 5.2: An estimate of the dephasing in the second harmonic $|\Delta N^{(2)}|$ of the GW signal for $q = 10^{-3}$ models in $\gamma_{\rm sp} = 7/3$ spike when relativistic precession effects are included. We find that the net dephasing for the non-rotating model is consistent among the non-precessing simulations and the precessing simulations. However, the effects of rotation are significantly dampened upon the inclusion of rotation leading to lower estimates of dephasing than before.

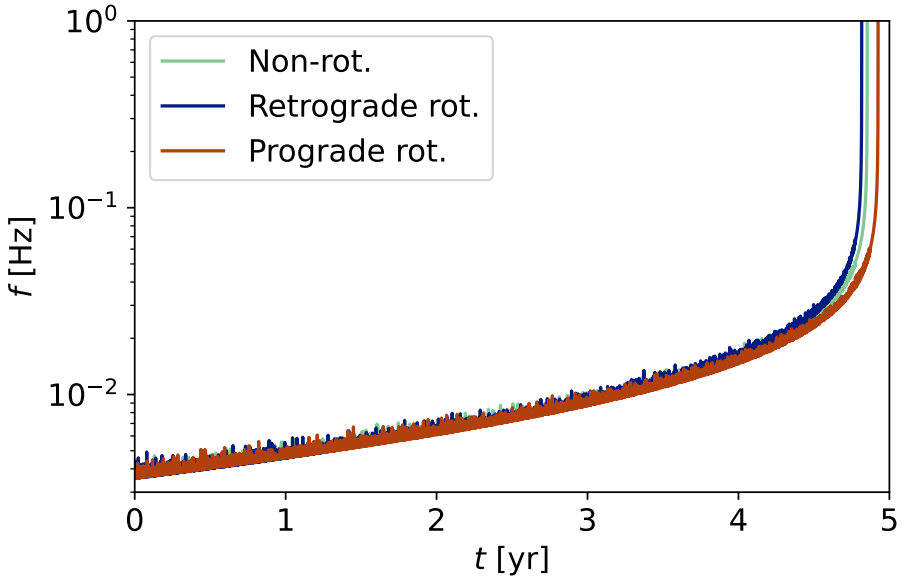

Figure 5.15: Evolution of the mean orbital frequency of the binary f in Hz as a function of time t in years for $q = 10^{-3}$ in a $\gamma_{\rm sp} = 7/3$ spike. We find that precession dampens the effects of rotating spikes substantially. In the non-precessing scenario the difference in dephasing between the non-rotating and rotating models was $\mathcal{O}(10^5)$ GW cycles, whereas in the precessing scenario that difference drops to $\mathcal{O}(10^4)$ GW cycles. Nevertheless, we find that the prograde rotation model takes longer to merge than the non-rotating model while the retrograde rotation model merges faster. This indicates that the results of our rotating models are robust, at least qualitatively.

of precession severely dampens the effect of the rotational effects that were evident in the non precessing simulations. Although, we find the same qualitative effects as in the non precessing simulations, i.e., retrograde model merges faster than the prograde and non rotating model, the results indicate that precession reduces the effects of rotation. Whereas in the non-precessing retrograde model, the binary takes 433.4 fewer days to merge compared to the inspiral in vacuum, in the precessing retrograde model, the binary takes 145.4 fewer days to merge. In the non-precessing prograde models, we found that the binary actually took 227.8 days longer to merge than in vacuum but we find that upon the inclusion of precession it merges 105 days earlier than in vacuum. We note that this is still $\sim$ 38 days slower than the non rotating model. The changes in the inspiral time are also reflected in the amount of dephasing over the 5 year inspral timespan. We present the number of dephasing cycles of the second harmonic $|\Delta N^{(2)}|$ over the course of the full inspiral in Table 5.2. The amount of dephasing for the non rotating model is consistent among both sets of simulations, i.e. precessing and non precessing, but we notice differences in the rotating models. In the non-precessing retrograde model $|\Delta N^{(2)}| \approx 7 \times 10^5$ but in the precessing retrograde model that drops to 2.4×10^5. We note that in the non precessing prograde models, we required 5×10^5 more cycles compared to vacuum for the binary to merge but due to the decreased effect of rotation, the binary merges faster, taking about 1.9×10^5 fewer cycles. The differences in dephasing between the non rotating and rotating models in the non-precessing case was $\mathcal{O}(10^5)$ but upon the inclusion of relativistic precession, that difference drops to $\mathcal{O}(10^4)$.

The reason for the drastic changes in the rotating models upon the inclusion of precession is unclear. Sesana et al. (2011) noted that in self-consistent N-body simulations where Newtonian precession of an extended gravitational system played a role, the effects of rotation were dampened. Here we observe a similar effect but with relativistic precession indicating the similarities between the two scenarios. Notably, the effect of the eccentric binary on the secular evolution of a particle can be described through the lens of the eccentric Kozai-Lidov effect (e.g., Merritt et al., 2009; Merritt, 2013a). Recent simulations of hierarchical MBH triplets have found that inclusion of PN effects can diminish or even extinguish the Kozai-Lidov evolution of the inner binary (Tanikawa and Umemura, 2011; Bonetti et al., 2018; Mannerkoski et al., 2021; Koehn et al., 2023). These simulations may hold clues to understanding the effect of precession and how it affects our results. However, a more thorough analysis requires future work.

Although the effects of rotation are decreased in the $q = 10^{-3}$ models, we expect rotation to still play a significant role in lower mass ratio binaries where the mass enclosed within the orbit of the binary would be larger. In such a case, the Newtonian precession of the spike could also play a part and the net change in precession could be used as an indicator for the presence of DM. However verifying this requires simulations that are beyond the scope of this current study.

5.5 Discussion

5.5.1 Accretion effects

The secondary is also expected to accrete from the spike during the inspiral. For the parameters considered in our simulations, the accretion effect is expected to be quite minimal, especially when $M_2 = 1M_\odot$. For the stellar mass BH scenario with $M_2 = 10M_\odot$, we can estimate the rate of accretion assuming that the secondary undergoes Bondi-Hoyle accretion (Bondi and Hoyle, 1944; Edgar, 2004; Macedo et al., 2013; Mach and Odrzywołek, 2021). Following Yue and Han (2018), the change in mass of the secondary over time ($\frac{dM_2}{dt}$) can be written as

$$\frac{dM_2}{dt} = \frac{16\pi G^2 M_2^2 \rho_{\rm DM}}{c^2 v}\left(1 + \frac{v^2}{c^2}\right) \tag{5.22}$$

where v is the velocity of the secondary. Assuming that $\rho_{\rm DM} \sim 10^{20} M_\odot {\rm pc}^{-3}$ near the secondary, and $v \sim 2 \times 10^4 {\rm kms}^{-1}$, we find that $\frac{dM_2}{dt} \sim 0.005 M_\odot {\rm yr}^{-1}$. Since the spike gets disrupted within the first 0.1 yr, we expect the accretion onto the secondary to be of the order of $10^{-4} M_\odot$ having minimal effects on our results. One can imagine that in lower mass ratio scenarios where the spike is not disrupted as much, the accretion effects might be larger, but still subdominant compared to dephasing due to three-body scattering.

We note that $\frac{dM_2}{dt} \propto \frac{1}{v}$. Since the relative velocity between the DM particles is lower in prograde rotating models than isotropic or retrograde models, the accretion effects could be larger. An estimate of the difference is beyond the scope of this current study but will be considered in the future.

Nichols et al. (2023) recently explored a self-consistent treatment of accretion in IMRIs with stellar mass BHs as the secondary and found that inclusion of accretion can lead to difference of $100-1000$ GW cycles compared to the models where accretion is not included. Although this represents about < 1 per cent difference in the number of dephasing cycles in our $q = 10^{-3}$ and 10^{-4} models, future studies will need to account for accretion to generate LISA waveforms since matched filtering requires a proper waveform determination over a few cycles. More importantly, the change in eccentricity due to accretion would also need to be studied considering the large effect eccentricity plays in GW radiation dominated binary evolution, as demonstrated in this work.

5.5.2 Dark matter annihilation and EM signatures

In addition to modifying GW signals of IMRIs, DM spikes are considered to be a source of gamma radiation as a result of DM annihilation. Thus, they provide strong indirect signatures of weakly interacting DM particles and allow us to probe DM microphysics (e.g., Gondolo and Silk, 1999; Ullio et al., 2001; Fields et al., 2014;

$M_1[M_\odot]$	$\langle\sigma v\rangle[\mathrm{cm^3 s^{-1}}]$	$\rho_{\rm pl}[M_\odot \mathrm{pc}^{-3}]$	$r_{\rm pl}[\mathrm{pc}]$
10^3	10^{-27}	8.4×10^{14}	2.2×10^{-6}
10^4			4.8×10^{-6}
10^3	10^{-30}	8.4×10^{17}	1.1×10^{-7}
10^4			2.5×10^{-7}
10^3	10^{-33}	8.4×10^{20}	5.9×10^{-9}
10^4			1.3×10^{-8}

Table 5.3: The annihilation radius $r_{\rm pl}$ and annihilation plateau density $\rho_{\rm pl}$ under different DM cross sections $\langle\sigma v\rangle$ for a $\gamma_{\rm sp} = 7/3$ spike with a central BH with mass M_1.

Shapiro and Shelton, 2016; Lacroix, 2018). In our work we do not consider the effect of DM annihilation on the spike profile. Our models are, therefore, representative of the non-annihilating DM case. Nevertheless, we provide estimates on how the spike profile can be changed and discuss how that would affect our results in case of annihilating DM.

For self-annihilating DM, the spike profile is significantly depleted due to the interactions between DM particles. It has been suggested that near the central MBH, a flat plateau, or core forms as a result of DM annihilation (Gondolo and Silk, 1999, but see also Vasiliev (2007); Shapiro and Shelton (2016)). The density of this plateau $\rho_{\rm pl}$ is given as

$$\rho_{\rm pl} = \frac{m_\chi}{\langle\sigma v\rangle T} \tag{5.23}$$

where m_χ is the mass of the DM particle, $\langle\sigma v\rangle$ is the interaction cross section, and T represents the age of the MBH. The interaction cross-section of the DM particles is considered to be a constant in standard thermal weakly interacting particle (WIMP) models. Assuming a scenario with $m_\chi = 1\mathrm{TeV}$, and $T \gtrsim 10^6\mathrm{yr}$ we can calculate the density of the plateau and the radius of the plateau ($r_{\rm pl}$) by setting $\rho_{\rm DM}(r_{\rm pl}) = \rho_{\rm pl}(r_{\rm pl})$ for various values of $\langle\sigma v\rangle$. We present this information in Table 5.3. We find that unless the DM annihilation cross section is very small, the density of the plateau is lower than the density of the spike where the binary is situated. The lower density would lead to a smaller amount of dephasing. Since the dephasing is proportional to the density of DM near the binary, as argued before, when $\langle\sigma v\rangle = 10^{-27}$ and $10^{-30}\mathrm{cm^3 s^{-1}}$, the dephasing would be reduced by factors of $50 - 5 \times 10^5$. In such a case, the dephasing would be significantly reduced, and non-detectable in the $\langle\sigma v\rangle = 10^{-27}\mathrm{cm^3 s^{-1}}$ case. However, it should be noted that, according to MAGIC Collaboration (2016) the upper limit on the cross-section is $\sim 10^{-25}\mathrm{cm^3 s^{-1}}$ and the annihilation cross-section in reality could be much lower. In such a scenario, one could observe GW dephasing in tandem with an EM signature.

We note that the above mentioned case is an optimistic scenario. Under a different scenario, we consider $T \sim 10^{10}$yr, typical for the ages of nearby galaxies. Using more conservative values of $m_\chi = 35$GeV and $T = 10^{10}$yr leads to a plateau density of $2.9 \times 10^9 M_\odot \mathrm{pc}^{-3}$ (assuming that $\langle \sigma v \rangle = 10^{-27} \mathrm{cm}^3 \mathrm{s}^{-1}$) which would leave no imprints on the GW signal. It should be noted, however, that over such a long duration, the system is not expected to be isolated.

Alternately, one could use the detection of GW signals from DM spikes to put upper-limits on the annihilation cross section of DM. Setting $\rho_{\rm pl} = \rho_{\rm DM}$, we get

$$\langle \sigma v \rangle \leq \frac{m_\chi}{\rho_{\rm sp} (\frac{r_{\rm sp}}{r_{\rm pl}})^{\gamma_{\rm sp}} T} \tag{5.24}$$

where we used equation 7 for $\rho_{\rm DM}$. For $M_1 = 10^3 M_\odot$ with a $\gamma_{\rm sp} = 7/3$ spike, $r_{\rm sp} \approx 0.5$ pc. As in our previous approximations, we take $\rho_{\rm sp} = 226 M_\odot \mathrm{pc}^{-3}$. We assume that the annihilation radius $r_{\rm pl} \sim a = 2 \times 10^{-8}$ pc, the semi-major axis of the binary. Since the dephasing due to three-body scattering depends on the local density as mentioned in previous sections, we should expect to obtain a similar amount of dephasing in the annihilation scenario as in the non-annihilating scenario. We, then, obtain the following upper-limit on the cross section:

$$\langle \sigma v \rangle \lesssim 1.71 \times 10^{-32} \mathrm{cm}^3 \mathrm{s}^{-1} \left(\frac{m_\chi}{1\mathrm{TeV}} \right) \left(\frac{10^6 \mathrm{yr}}{T} \right). \tag{5.25}$$

Similar analysis can be performed with different spike parameters and central IMBH masses. We point to the reader, however, that a lower amount of dephasing than expected can arise from different DM properties that lead to a larger annhilation, or from dynamical factors such as rotation, as pointed in this study. In such a scenario, a different signature, possibly electromagnetic would be needed to break the degeneracy. In any case, as Hannuksela et al. (2020) points out, any detection of DM spike using GWs will be in strong tension with current thermal WIMP models and place constraints on the mass of the particle.

Other models can suggest a lower cross section allowing for the co-existence of EM signals along with GW signals from the spike (e.g., Shelton et al., 2015). However, Hannuksela et al. (2020) report that in such a case, even in the most optimistic scenario with $M_1 = 10^6 M_\odot$, the electromagnetic counterpart will be detectable only up to a distance of 90 Mpc by detectors such as ASTROGAM (de Angelis et al., 2018), FERMI, or CTA. Only a few IMRI events would happen in such a small volume and an EM counterpart would require a large fraction of IMRIs to be embedded in DM spikes. Since the luminosity of the gamma radiation L is proportional to the squared density of the DM spike profile $\rho_{\rm DM}^2$, the EM signals from IMBHs are expected to be weaker and the prospects for finding EM counterparts are more pessimistic.

5.5.3 Implications for binary inspiral in realistic environments

As we have seen in the previous sections, the amount of dephasing is sensitive to the density profile near the IMBH. Furthermore, the actions of the inspiraling compact object can have a drastic effect on the spike, completely unbinding it. This implores us to ask an important question: how realistic are the spikes considered in this study and previous studies? How do realistic spikes affect the dephasing of an inspiraling compact object? While a comprehensive study to understand spike profile in realistic environments is beyond the scope of this work, we present a brief analysis of the impact of surrounding environments and past inspirals on the spike density profile.

In a non-isolated environment interactions between the spike and the surrounding material can reduce the density of the spike compared to isolated adiabatic growth models. Previous works have only considered the effect of stars surrounding the spike (e.g., Merritt et al., 2007b). Using the Fokker-Planck code `Phaseflow` (Vasiliev, 2017) we study, for the first time, the effect of stellar mass BHs along with stars on the final DM spike profile. Here we only consider the case where the central SMBH has a mass of $10^6 M_\odot$ and the total mass of the stellar mass BH particles is 1% of that of the total mass of stars in the galactic nucleus. The total stellar mass in the nucleus is set to be $10^7 M_\odot$. We note that this excludes the bulge mass. Our two-component model is consistent with the expectation from the Kroupa initial mass function (IMF) (Kroupa, 2001) The stellar mass BH particles are given masses 10 times larger than that of the star particles. The DM halo has a Hernquist profile initially representing a $\gamma = 1$ slope in the center and the stars and BHs in the nuclei are given a shallow $\gamma = 0.5$ cusp.

In Figure 5.16, we plot the slope of the DM component at different times and find that in an equilibrium state the spike reaches a $\gamma = 1.5$ profile near the MBH. This is similar to what is observed in a stars only model. However, due to mass segregation in the presence of a two component mass spectrum, the rate of growth of the spike is enhanced. Comparing our two component mass model to that from a single component model, we find that presence of two mass species accelerates the growth by a factor of 4. In fact, this is quite sensitive to the IMF and the initial density profile of the nuclei. A top-heavy IMF (e.g., Chabrier, 2005) results in an even faster growth. This can have major implications for the time taken by the spike to regrow after it has been disrupted. A detailed study of regrowth time after disruption will be considered in a future work.

A major caveat of the above result is that we do not consider the effect of stellar evolution on relaxation. Stellar evolution can lead to mass loss and kicks to compact objects over time which reduces their population and therefore affects the relaxation timescale. We consider the effect of stellar evolution in an approximate manner by evolving a population of stars drawn from a Kroupa IMF with different metallicities and in different environments using SSE (Hurley et al., 2000). In a dense environment

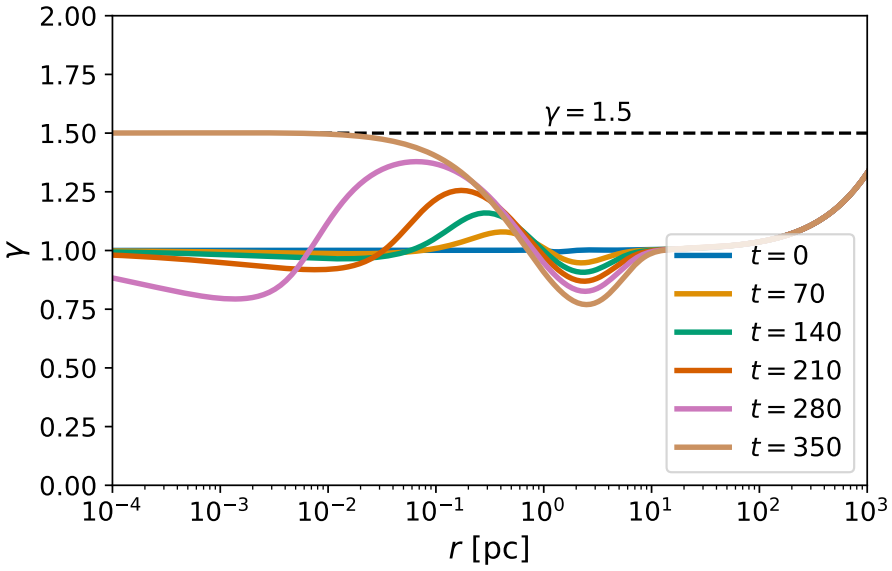

Figure 5.16: Slope of the DM profile γ as a function of distance r from the central MBH at various times. The time is presented in N-body units with 1 time unit corresponding to ≈ 14.9 Myr. The initial profile of the DM surrounding the MBH is a $\gamma = 1$ model, representing a Hernquist type halo. The DM profile is embedded in a nuclei consisting of stars and stellar mass BHs, drawn from a Kroupa IMF, and the subsequent evolution is performed using a Fokker-Planck model. We find that, similar to the stars only model, the spike reaches a $\gamma = 1.5$ profile at the center, near the MBH. However, unlike the stars only model, two body relaxation is enhanced due to the presence of a two component mass function, leading to an accelerated growth.

such as a galactic nucleus where the escape velocity is higher, we find the mass-ratio of BH particles approaches $M_{\rm BH}/M_* \sim 10^{-3}$ of the mass of stellar particles after 5 Gyr of evolution when the metallicity $Z = 0.1Z_\odot$ where $Z_\odot$ is the solar metallicity. This drops to $M_{\rm BH}/M_* \sim 10^{-4}$ when $Z = Z_\odot$. The decrease in mass is due to stellar evolution mass loss and random kicks imparted by SSE onto compact objects during their formation. In a globular cluster like environment with a total mass of $10^6 M_\odot$, the escape velocity is lower and we find heavier compact objects are hardly retained ($M_{\rm BH}/M_* \sim 0$). This affects the timescale of growth of the DM spike. Using Fokker-Planck models with the evolved profile, we find that when $M_{\rm BH}/M_* \sim 10^{-3}$, the timescale of spike growth increases by 40% compared to the two-component model without stellar evolution where $M_{\rm BH}/M_* \sim 10^{-2}$. When $M_{\rm BH}/M_* \sim 10^{-4}$, the timescales increases by a factor of 2 compared to the non-stellar evolved model. This highlights the importance of inclusion of a mass-species which accelerates the growth of a spike even when the population of the heavier mass species is much smaller than lower mass species.

Once a spike has been disrupted, the regrowth happens on timescales that are on the order of collisional relaxation time within the sphere of influence of the MBH. The relaxation time is affected by the mass species surrounding the MBH. In a single component model, relaxation takes longer than when a mass spectrum is present, as evident from our results above. We can estimate the relaxation time $t_{\rm relax}$ in a stellar only environment as follows (e.g., Babak et al., 2017; Becker, 2024) :

$$t_{\rm relax} = \frac{5}{\ln(\Lambda)} \left(\frac{\sigma}{10 {\rm kms}^{-1}}\right) \left(\frac{r_{\rm infl}}{1 {\rm pc}}\right)^2 {\rm Gyr} \tag{5.26}$$

where $\ln(\Lambda)$ is the Coulomb logarithm,σ is the velocity dispersion, and $r_{\rm infl}$ is the influence radius of the MBH. σ can be estimated from the well known $M - \sigma$ relationship (e.g., Gültekin et al., 2009; Kormendy and Ho, 2013c) or from density profiles of galactic nuclei. Since we focus on IMBHs in this work, we refer to the density profiles of known dwarfs from Nguyen et al. (2017, 2018) to calculate σ and $r_{\rm infl}$. Taking $\ln(\Lambda) \sim 10$, $\sigma \sim 50 {\rm kms}^{-1}$, and $r_{\rm infl} \sim 0.1$ pc ,we find $t_{\rm relax} \lesssim 0.025$ Gyr for a $10^4 M_\odot$ IMBH. This is representative of relaxation time in a high density environment. On the other hand, in a low density environment, $r_{\rm infl}$ is going to be larger. Assuming $r_{\rm infl} \sim 0.5$ pc in that case, we find $t_{\rm relax} \lesssim 0.6$ Gyr. The inspirals happen on timescales over which two-body relaxation depletes the cusp by driving the compact objects on loss cone orbits. The depletion timescale $t_{\rm d}$ is estimated as follows (e.g., Babak et al., 2017; Becker, 2024):

$$t_{\rm d} = \frac{20}{1+R} \left(\frac{m}{10 M_\odot}\right)^{-1} \left(\frac{M_{\rm MBH}}{10^6 M_\odot}\right)^{1.19} {\rm Gyr} \tag{5.27}$$

where R is the ratio of plunges to inspirals, and m is the characteristic mass of the compact object. Taking $R = 0$ to find the upper limit and $m = 10 M_\odot$ representing BHs , we find that $t_d \lesssim 0.1$ Gyr. Thus, the time between inspirals is comparable to

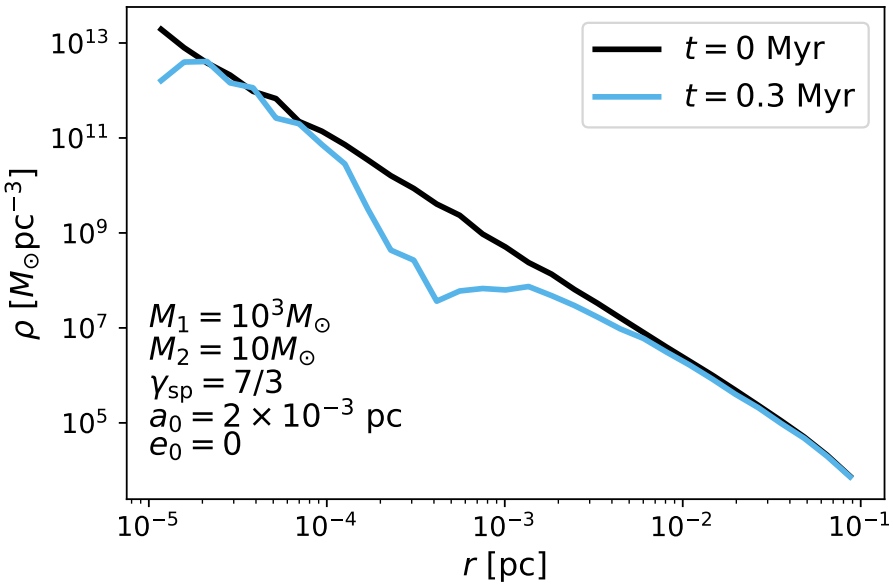

Figure 5.17: Density of a pristine $\gamma_{\mathrm{sp}} = 7/3$ DM spike ρ as a $10M_\odot$ BH is in. Even at larger separations, we find that the inspiraling object forms a core extremely rapidly. This suggests that dense spikes of the form of $\gamma_{\mathrm{sp}} = 7/3$ can only be present if there were no prior inspirals.

the two body relaxation time in high density environments and is shorter in the low density environments. As such, we expect the spike to not exist in its equilibrium state in low density environments. However, under the presence of a mass spectrum the relaxation time decreases in which case $t_d \sim t_{\mathrm{relax}}$ which can lead to the spike existing in the equilibrium $\gamma = 1.5$ state, even in low density environments. This highlights the importance of examining the spike growth embedded in a realistic mass spectrum. We note one caveat of this calculation is that relaxation times after depletion of a spike a larger than the ones calculated from the above equation. Therefore, our relaxation timescale should be considered as a lower limit.

A big question also lingers regarding the starting point of the compact object in the simulations. In a realistic scenario, a compact object would be unbound initially and become bound over time due to the effect of dynamical friction, two-body relaxation, and hardening in the stellar and DM environment. In a stellar dominated environment the compact object is going to be driven to inspiral orbits mainly due to the effects of two-body relaxation. In such a case, we would expect the initial inspiral to cause minimal disruption to the spike. However, in such a case, the spike profile would have a $\gamma_{\mathrm{sp}} = 1.5$ slope which we showed produces minimal dephasing effects for the systems considered. This was also highlighted in Becker (2024) where the author found that GW signals from the spikes in this case would be hardly detectable. On the other

hand, in an isolated environment with no erosion, the compact object is going to inspiral due to the effect of dynamical friction and three body hardening from the DM environment. In such a case, the formation and inspiral itself would produce a flat core that can massively reduce dephasing effects. As an example, we produce the impact on the density profile from a simulation with a $\gamma_{\rm sp} = 7/3$ spike where the compact object now starts at an initial semi-major axis of $a_0 = 2 \times 10^{-3}$ pc in Figure 5.17. Within 0.1 Myr, we find a core has formed with a density of $\sim 10^7 M_\odot {\rm pc}^{-3}$. According to Merritt et al. (2007c), the formation of a hard binary is accompanied by an ejection of mass comparable to the mass of the binary. Therefore in DM only environments, we expect the density to be even lower and the core to be larger. The relaxation time of the spike in this isolated environment is so long that we do not expect a regrowth within a Hubble time. This indicates that the simulations with the $\gamma_{\rm sp} = 7/3$ models represent optimistic scenarios where the spike *exists in isolation and has not undergone a previous inspiral*. This is a major caveat of our work and that of previous works. Additionally, on account of their lower masses, IMBHs can be off-center where the adiabatic growth of the spike can be diminished. In such a scenario, a $\gamma = 1.5$ is typically formed, as noted by Zhao and Silk (2005). All of the above noted issues are pertinent and we emphasize the need to understand spike growth in realistic environments. We leave this as a future study.

5.5.4 Prospects for GW detection

As shown in Figure 5.3, IMRI sources are promising for multiband GW astronomy with LISA and potential decihertz detectors. Several formation scenarios have been proposed for the formation of IMRIs in DM spikes, including host sites such as the nuclear star clusters of dwarf galaxies (Yue et al., 2019) and merger remnants of elliptical galaxies (Vázquez-Aceves et al., 2023). The merger rate of such systems with and without DM spikes, however, is still uncertain. More detailed population synthesis of these sources with either semi-analytic models, cosmological simulations, or some combination thereof would provide more insight into what merger rates might be possible. Conversely, the GW detection of IMRI sources and constraints on potential DM environments would provide tests of astrophysical population models in addition to DM physics.

In light of the critical examination of realistic spike profiles from the previous section and Becker (2024), the prospects for dephasing due to DM spikes on IMRIs appear to be somewhat not optimistic. However, we emphasize that a full parameter space study is required before we can conclusively determine whether spikes would have tangible dephasing impact on GWs from IMRIs. On the other hand, Dai et al. (2022, 2023) and Becker (2024) noted that inspirals often happen on highly eccentric orbits with larger semi-major axes. In such a case, the DM spike can lead to a periapse precession in addition to the relativistic precession. The Newtonian precession of the spike is opposite in direction to the relativistic precession and the net effect can be

quantified as a *deshifting* index (Becker, 2024), which is a measure of the change in GW cycles due to change in pericenter precession from the DM spike. Dai et al. (2023) noted that deshifting can be present even in low density environments. Our preliminary analysis suggests that deshifting may be a more optimistic signature than dephasing. Unfortunately, for the set of models considered in this work, we were not able to quantify deshifting. In the future, we plan on studying an extended parameter space where we study dephasing as well as deshifting across spike models.

A potential space-based decihertz GW detector provides a significant advantage for boosting the IMRI detection rate as well as breaking parameter degeneracies that might be encountered with just LISA data. In addtion, if the inspiral phase of IMRIs is at a sub-threshold level in LISA data, the resolved merger phase in a space-based decihertz GW detector can provide priors to search for the subthreshold inspiral phase in archival LISA data. This is similar to how LIGO and 3G/XG ground-based detectors like Einstein Telescope or Cosmic Explorer may provide prior information on stellar-mass BH binary mergers to search for potential inspiral counterparts in archival LISA data (Ewing et al., 2021).

In addition to the proposed tests of dynamical friction and accretion on gravitational waveforms, one can also test for the presence of three-body/loss-cone scattering for a potential IMRI event. One caveat for testing the presence of such effects might be that nonlinear feedback and DM-spike relaxation may complicate waveform parameterization. Though this certainly motivates more numerical simulations of IMRIs in DM spikes in order to explore the parameter space more comprehensively and to inform waveform parameterizations that are interpretable.

5.6 Conclusions

IMRIs are considered among the most crucial sources of low-frequency gravitational waves (GWs) that future space-based GW detectors like LISA and DECIGO can potentially detect. Recent studies have suggested that if the intermediate-mass black hole (IMBH) in these systems is surrounded by a dark matter (DM) spike, the gravitational effects of the spike may induce dephasing in the observed GW signal, serving as a distinctive signature of the DM spike. Prior investigations have predominantly employed analytic methods to model the interaction between the spike and the binary, treating the gravitational impact of the spike on the binary through dynamical friction (DF). However, these approaches neglect the influence of the binary on the spike itself, a factor that can significantly affect the dynamics and, consequently, the degree of dephasing.

In our study, we employ N-body methods to delve into the complex dynamics involved in eccentric IMRIs embedded in both non-rotating and rotating DM spikes, evolving the system self-consistently. Our work represents the first attempt at modeling such systems using N-body simulations. The simulations reveal that, contrary to previous assumptions, the primary mode of dissipating energy from the system

is not DF, a cumulative effect of two-body encounters, but rather three-body scattering, similar to stellar loss-cone scattering in SMBH/MBH binaries. We note that during the submission of this work, Kavanagh et al. (2024) also finalized their work on N-body simulations employing methods similar to ours. They qualitatively find similar behavior of the IMRI but find that the interactions can be described using a semi-analytic method combining dynamical friction loss with the time dependent potential of the binary. While their simulations are run for a shorter duration than ours, the combined effect which they propose is of three-body nature and similar to what we describe in this work.

The main conclusions of our work can be summarized as follows:

- The evolution of the non-rotating models reveals that in larger mass ratio cases, the binary is highly efficient at rapidly degrading the spike, resulting in minimal dephasing effects. Depending on the mass-ratio of the binary, our dense spike models with $\gamma_{\rm sp} = 7/3, 9/4$ show that dephasing of $\mathcal{O}(10^4) - \mathcal{O}(10^5)$ can be expected. However, in lower density $\gamma_{\rm sp} = 3/2$ spikes produced in a collisional environment, we find minimal dephasing of $\leq \mathcal{O}(10)$.

- We find that all of our simulations with dense spikes ($\gamma_{\rm sp} = 7/3, 9/4$) predict that the dephasing is much larger, by factors of 10-100, than that predicted by the self-consistent semi-analytic method `HaloFeedback` for similar binaries in circular orbits. This indicates that DF theory is unable to fully capture the dynamics of the binary.

- In our rotating models with dense $\gamma_{\rm sp} = 7/3$ spikes, we find that, spikes that counter-rotate with the binary lead to faster inspirals compared to spikes that co-rotate with the binary, which lead to slower inspirals, even slower than inspirals in vacuum. This leads to dephasing effects that are $2.5 - 3.5$ times higher in the retrograde/counter-rotating cases than their non-rotating counterparts.

- We use three-body simulations to investigate the nature of the interaction of the binary with spike particles and find that the binary primarily dissipates energy via three-body interactions rather than DF, a two-body effect. The impact of DF in all of our simulations is sub-dominant shedding some light on the discrepancy between our results and those obtained in previous studies. Additionally, three-body simulations reveal that in rotating spikes, particles on retrograde motion eccentrify the binary whereas particles on prograde motion circularize the binary. This indicates that rotation in spikes cannot be neglected and should be the subject of further investigations in the future.

- We conduct simulations to investigate the impact of relativistic precession on dephasing in our $q = 10^{-3}$ models. Our findings indicate that precession has minimal to no effect on dephasing in the non-rotating models but significantly diminishes the effects of rotation.

- By examining our initial models with Fokker-Planck methods, we assess the presence and detectability of spikes in realistic environments. Our results suggest that non-isolated environments have DM spikes with shallower slopes than previously considered, leading to smaller signals and lower detection prospects via dephasing. We show that the densest DM spike models with $\gamma_{\rm sp} = 7/3$ represent scenarios where the spike is isolated and has not undergone any previous inspirals, as any inspiral leads to the formation of a low-density core where the regeneration of the spike takes much longer than the Hubble time. In the presence of lower density $\gamma_{\rm sp} = 3/2$ spikes, present in non-isolated environments, dephasing is minimal and non-detectable. High stellar and stellar-mass black hole densities can accelerate the growth of $\gamma_{\rm sp} = 3/2$ spikes, but a more extensive parameter space study is needed to determine which E/IMRIs embedded in such spikes can produce detectable dephasing signatures. Our preliminary analysis, coupled with recent studies, suggests that "deshifting" rather than dephasing might be a more optimistic signature, as it is more robust even in low-density environments (Dai et al., 2022; Becker, 2024).

Although our work uses idealized models, it provides a foundational exploration that paves the way for more comprehensive analyses in the future. Despite critical uncertainties, particularly regarding the growth, retention, and population of DM spikes around IMBHs, our research highlights the promise and potential benefits of delving deeper into this intricate problem. We demonstrate the need to consider more realistic environments, as detection prospects depend sensitively on spike density, which is influenced by the surrounding environment. Our work also underscores the importance of multiband GW astronomy and the need for a dedicated decihertz detector. By enhancing our approach to include accretion onto the inspiraling object and incorporating higher-order post-Newtonian terms, our N-body code emerges as an effective method for simulating the evolution of IMRIs embedded in DM spikes.

5.7 Appendix

5.7.1 Comparison to Taichi, PH4, and HaloFeedback

To verify the accuracy of our method, we run a set of five simulations with $M_1 = 100M_\odot$, $M_2 = 1M_\odot$, $a_0 = 3 \times 10^{-8}$ pc and $e_0 = 0$. We compare the results to those obtained from N-body codes `Taichi` (Zhu, 2021c; Mukherjee et al., 2021a, 2023), and `ph4` from `AMUSE` (Portegies Zwart et al., 2013; Portegies Zwart and McMillan, 2018) where the self gravity of the spike is not neglected. `Taichi` is run in the direct force summation mode with the fourth order Hamiltonian splitting integrator `HHS-FSI` while. The simulations are run for a duration of 1500 orbits of the secondary around the primary. Additionally, we compare our results to those obtained from `HaloFeedback` (Kavanagh et al., 2020) as well since our binary is in a circular orbit

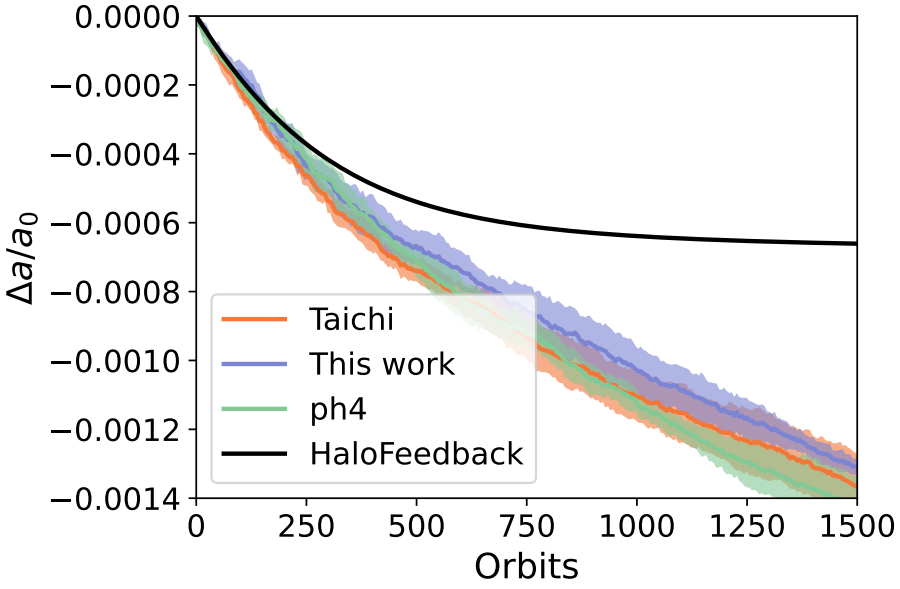

Figure 5.18: The evolution of the relative change in the semi-major axis $\Delta a/a_0$ as a function of the number of orbits of the binary. The solid lines indicate the average of five independent simulations whereas the shaded region indicates the standard deviation. We find that the results from method presented in this work (purple line) agree with that from `Taichi` (orange line), `ph4` (green line), N-body codes where the self-gravity of the spike is not neglected. This indicates that the effect of the DM-DM interactions is minimal and can be safely ignored, validating our method. Also presented is the evolution of the same system using `HaloFeedback` (black line). While the results from `HaloFeedback` and the N-body codes agree over the first 100-150 orbits, they diverge after that. The spike is able to dissipate more energy from the binary in the N-body models compared to the `HaloFeedback` models since the impact of DF is subdominant compared to three-body effects.

initially. We note that we do not include the relativistic terms in the above-mentioned simulations.

As a test of the validity our method, we present the evolution of the relative change in the semi-major axis $\Delta a/a_0$ from the N-body code used in this work, `Taichi`, `ph4`, and `HaloFeedback` in Figure 5.18. The solid lines represent the mean values obtained from the five simulations whereas the shaded regions represent the standard deviation. We find that the evolution using our approximate force N-body method is consistent with that obtained from `Taichi` and `ph4` where the DM-DM interactions are not neglected. This indicates the effect of the self gravity of the spike is minimal and can be safely neglected. However, we find major differences between the results from the N-body codes and `HaloFeedback`. The three methods produce consistent results for the first 150 orbits after which the `HaloFeedback` results deviate because the binary enters the hard binary phase in which three-body interactions play a dominant role. The binary stalls at $\Delta a/a_0 \approx -5\times10^{-4}$ in the `HaloFeedback models` whereas a longer evolution indicates that the binary actually stalls at $\Delta a/a_0 \gtrsim -2\times10^{-3}$ in the N-body simulations. This is almost $4\times$ larger than what is predicted by `HaloFeedback`. As a result of this discrepancy, the dephasing obtained from the `HaloFeedback` models should be smaller than those obtained from the N-body models, a fact discussed in the results section. The discrepancy arises from the fact that DF is subdominant to three-body scattering as we showed in section 4.3. Moreover, the three-body effects occur over longer timescales, indicating that semi-analytic methods should be validated against long term, secular N-body simulations.

Chapter 6

A hierarchical Hamiltonian splitting N-body code that runs entirely on GPUs

Diptajyoti Mukherjee[1], **Hy Trac**[1]

[1]McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213

Abstract

Solving the post-million particle N-body problem presents a formidable computational challenge. The advent of highly parallel hardware, such as Graphics Processing Units (GPUs), has garnered significant attention within the astrophysics community as a means to address this challenge. While GPUs can offer impressive speedups over traditional CPUs, developing a fully-fledged N-body code optimized for GPUs remains non-trivial. In particular, for discrete GPUs, data transfers between the CPU and GPU can become a major performance bottleneck. To mitigate this, we develop a fully GPU-resident code for collisional dynamics. Our implementation employs a GPU-resident version of the second-order Hamiltonian splitting HOLD integrator, incorporating individual symmetrized timesteps and zero softening. We detail the algorithmic modifications required to ensure that all components of the HOLD integrator remain resident on the GPU, completely eliminating CPU-GPU data transfers during the integration process. Compared to a highly optimized and parallelized CPU version of the same algorithm, our GPU implementation achieves a $4\times$ speedup for $N \sim 10^6$. This work lays the foundation for constructing a fully GPU-resident collisional N-body code based on the fast multipole method.

6.1 Introduction

The computational intensity of both direct and approximate N-body methods naturally motivates the use of highly parallel hardware to handle the computational challenges they pose. Traditionally, large-scale N-body simulations have relied on distributed-memory parallelism (`MPI`) or shared-memory parallelism (`OpenMP`), both of which are limited to CPU-based architectures. However, the advent of `CUDA` in 2007 significantly streamlined the development of Graphics Processing Unit (GPU) applications, making high-performance GPU computing more accessible to the broader computational astrophysics community (e.g., Bédorf and Portegies Zwart, 2012).

GPUs offer dramatic speedups compared to conventional CPUs by harnessing thousands of cores on a single chip, making them especially well-suited to problems such as N-body simulations, which are naturally parallelizable. While CPUs rely on a relatively small number of cores optimized for serial performance, GPUs achieve high throughput by scheduling many light-weight threads that can operate concurrently. When one *warp* (a group of GPU threads that execute in parallel) is stalled waiting for data to be fetched from memory, the scheduler can seamlessly switch execution to another warp that is ready to perform numerical operations. This strategy, known as *latency hiding*, allows the GPU to achieve high floating-point throughput, provided there is enough parallel work to keep the hardware busy.

A critical consideration for exploiting GPU performance is *arithmetic intensity* which is defined as the the ratio of floating-point operations to memory transfers. The classical N-body problem stands out in this respect because, while each particle needs to load its state from memory, the force calculation itself involves $\mathcal{O}(N)$ or even $\mathcal{O}(N^2)$ arithmetic operations, depending on the method. Moreover, GPUs feature a configurable on-chip cache known as *shared memory*, which allows for tailored data reuse strategies and further boosts performance. This combination of abundant parallelism, high arithmetic intensity, and flexible memory hierarchies makes GPUs an ideal platform for modern N-body codes.

Numerous GPU-accelerated N-body frameworks have already emerged—`HiGPUs` (Capuzzo-Dolcetta et al., 2013), `Phi-GPU` (Berczik et al., 2011; Khan et al., 2011), `Bonsai` (Bédorf et al., 2012), and `FROST` (Rantala et al., 2021a, 2022), among others—often running on multi-GPU clusters to tackle large-scale simulations of gravitationally interacting particles. Such frameworks have been instrumental in performing simulations with $N \gtrsim 10^6$, crucial for many dense stellar systems. Even many legacy N-body codes such as `NBODY6` (Aarseth, 1999; Wang et al., 2015) which utilize force decomposition into a short and a long range part have been modified to utilize GPUs for the long range force calculation and have benefited from improved performance.

GPUs broadly fall into two categories: discrete and integrated. Discrete GPUs, commonly employed in high-performance computing clusters and servers, maintain their own dedicated memory-often referred to as global memory or VRAM. Because these GPUs lack direct access to the system's main memory (DRAM), simulation

data must be explicitly transferred between the CPU and GPU memory before computations can begin. In contrast, integrated GPUs share the same memory space as the CPU, eliminating the overhead of CPU–GPU data transfers but often providing lower overall memory bandwidth. While discrete GPUs have traditionally dominated large-scale scientific computing, emerging architectures such as AMD's MI300X and Apple's M-series system-on-chip are popularizing integrated designs in exascale supercomputers, potentially bridging the gap between performance and ease of memory access.

A major performance bottleneck in many GPU-based codes that utilize discrete GPUs lies in the memory transfer between CPU and GPU. As mentioned above, in discrete GPU systems, the GPU cannot directly access host memory (DRAM), requiring data to be copied back and forth. This overhead can significantly hamper speedups relative to CPU-based implementations, especially if the data has to be transferred on a high frequency. Fortin and Touche (2019) note that the fast multipole method (FMM) algorithm (Greengard and Rokhlin, 1987; Cheng et al., 1999) utilizing the dual tree walk (Dehnen, 2000, 2002) has a higher performance per cost on integrated GPUs versus discrete GPUs. Consequently, a promising approach is to offload the entire integration process to the GPU, thereby minimizing or eliminating CPU–GPU data transfers. The success of frameworks like `Bonsai` and `HiGPUs` demonstrates the viability of this approach. `Bonsai` uses a BH tree approach and a second order symmetrized leapfrog integrator (Hut et al., 1995) and is primarily used for collisionless simulations, while `HiGPUs` uses a fourth order Hermite integrator with block timesteps and is primarily used in collisional simulations.

Motivated by these advances, we seek to understand whether a hierarchical Hamiltonian splitting based N-body method for collisional dynamics would be viable exclusively on GPUs and compare its performance against a highly optimized CPU version of the same algorithm. To this end, we implement a version of the HOLD (Pelupessy et al., 2012b) integrator that runs *exclusively on GPUs* [1], requiring no data transfer between CPU and GPU during the entire simulation. Our code is fully collisional, and features an individual, adaptive time-stepping scheme based on symmetrized timesteps. Unlike `HiGPUs`, our code utilizes zero force softening. By designing the algorithm to remain on the GPU throughout the integration, we aim to optimize performance while preserving the accuracy and flexibility offered by Hamiltonian splitting methods for large-scale N-body simulations.

We begin by introducing the standard force calculation algorithm on GPUs using CUDA in section 6.2. This is followed by a discussion of the HOLD integrator on GPUs in section 6.3 and results in section 6.4. We then discuss our results and challenges with implementing the HOLD algorithm in section 6.5 which is followed by conclusions in section 6.6.

[1] https://github.com/dipto4/parallel-hhs/

6.2 Force calculation on GPUs

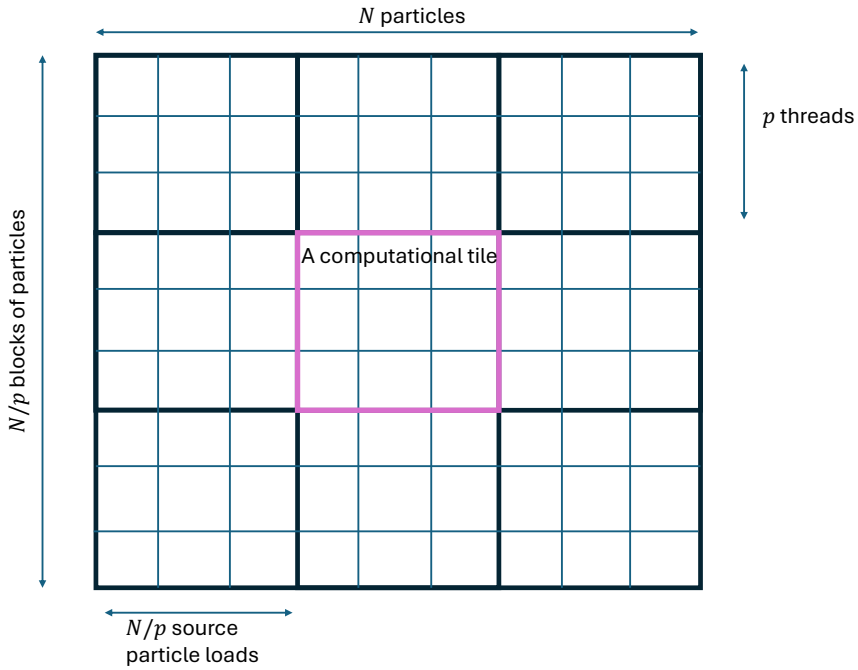

Figure 6.1: Visualization of the Nguyen (2007) algorithm for force calculation on GPUs. Each big square represents a *block* that contains a group of p *threads*. Each small square represents a *thread*. A thread within each block reads N/p particles from the global memory and stores it on the on-chip shared memory for faster access. Each thread in the block then calculates the force between a particular particle and the group of N/p particles. Once the calculation for the particular tile of N/p particles is over, another tile of N/p particles is loaded until the forces from all N particles have been calculated. This leads to an efficient force calculation algorithm.

The most computationally expensive task in an N-body code is typically the pairwise force calculation. Fortunately, this component is also highly parallelizable, making it an ideal target for GPU offloading. Many modern N-body implementations therefore focus their efforts on accelerating this force-evaluation kernel, drawing on the approach outlined by Nguyen (2007) for efficient GPU usage.

On GPU architectures, threads are organized into *blocks*, each of which has access to a dedicated pool of *shared memory* and a synchronization mechanism. Shared memory sits closer to the GPU's compute units and is roughly two orders of magnitude faster to access compared to global *(VRAM)* memory. A common strategy to leverage this high bandwidth is:

1. A grid of thread blocks is created. Each thread within a thread block handles

the force calculation for a single *sink* particle. Each thread, known as a *sink thread*, loads the position and velocity information of one *sink* particle into its private registers. Registers are the fastest storage available to a thread.

2. Each block of p threads collectively loads a subset of N/p *source* particles from global memory into shared memory.

3. Every sink thread in the block computes the interaction between its assigned sink particle and the N/p source particles currently resident in shared memory.

4. Once all threads have finished processing the forces from the current batch of N/p particles, the next batch is loaded into shared memory, and the process repeats.

By iterating in this fashion, the total number of costly global memory accesses is reduced, as particle data is fetched once per batch and then reused by each thread in the block. This approach takes full advantage of the GPU's high parallel throughput and shared-memory bandwidth, significantly increasing performance over naïve, global-memory-based methods. A visualization of this algorithm is provided in Figure 6.1.

6.3 HOLD on GPUs

Although offloading just the pairwise force calculation to the GPU can significantly boost performance, memory transfers between the CPU and GPU can become a significant bottleneck. In Figure 6.2, we compare the time spent on computation against the time devoted to data transfers. For $N \lesssim 10^4$, more than 20% of the runtime is typically consumed by copying particle data back and forth. At smaller particle counts on the order of 10^3, these transfers can account for over half of the total simulation time.

Further insights come from Figure 6.3, which illustrates the distribution of particles across integration levels in the HOLD scheme (see Section 1.1.2 for an introduction to the HOLD scheme). These levels are governed by a timestep parameter η, and we observe that for both "long" ($\eta = 0.2$) and "short" ($\eta = 0.02$) timesteps, at least 95% of the integration levels involve ≤ 8192 particles. In other words, most integration steps process a relatively small number of particles, yet each step still requires sending data to (and potentially retrieving data from) the GPU. Consequently, frequent CPU-GPU transfers can overshadow the gains achieved by GPU-accelerated force computations.

These observations underscore the efficiency gains possible when offloading every stage of the integration process to the GPU. By avoiding repetitive data transfers for small subgroups of particles, a fully GPU-resident approach can drastically reduce overhead and exploit the parallel resources of modern GPUs more effectively.

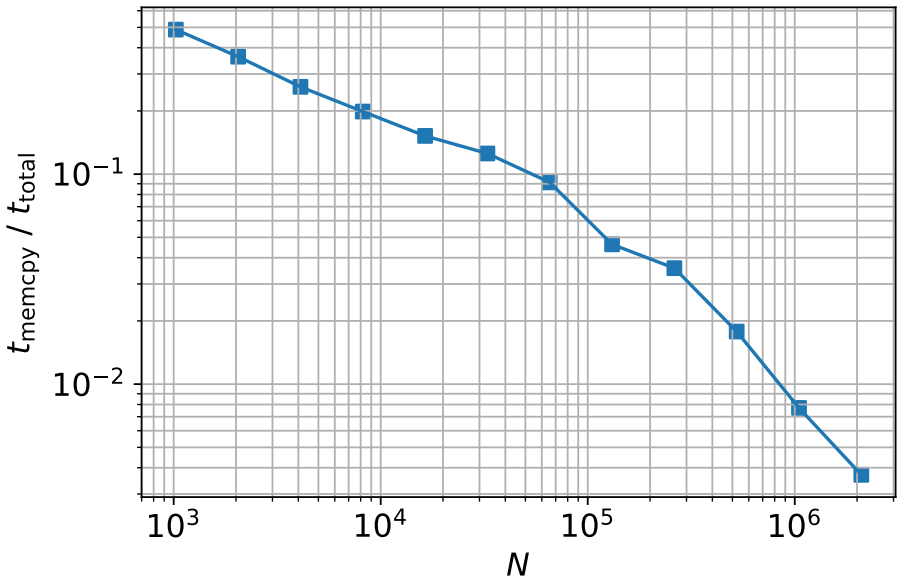

Figure 6.2: Ratio of the host to GPU and GPU to host memory transfer time t_{memcpy} to the total time (force calculation + memory transfer) t_{total}. We find that when $N = 10^3$, $\sim 50\%$ of the total time is spent in memory transfers. Even when $N = 10^4$, 20% of the time in spent in memory transfer leading to inefficiencies. At much higher particle numbers, the memory transfer time becomes subdominant. At $N = 10^6$, roughly 1% of the time is spent in memory transfers.

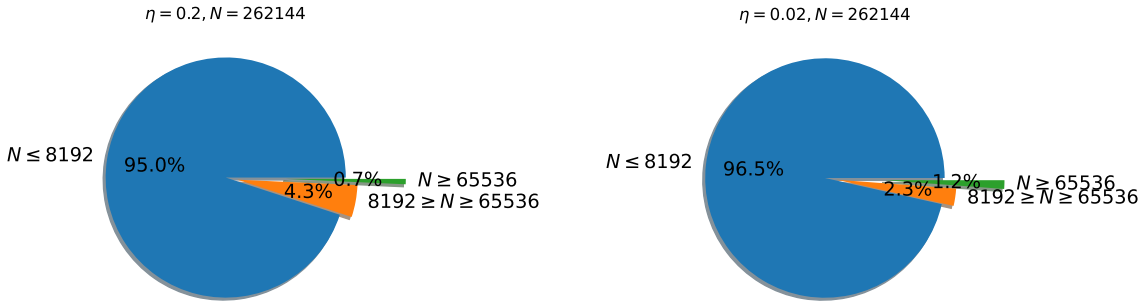

Figure 6.3: Piechart of the number of particles per level of the HHS splitting algorithm using $N = 262144$ particles and integrating to $t_{\mathrm{final}} = 0.0625$. We notice that over 95% of the levels using both a high timestep parameter ($\eta = 0.2$) and a low timestep parameter ($\eta = 0.02$) involve particle numbers ≤ 8192. At $N \leq 8192$, memory transfers consume a significant portion of the total computational time. Performing memory transfers at every level would thus lead to performance bottlenecks.

Following Listing 1.1 in Chapter 1, we note that the main components of the HOLD integration scheme are the `split`, `drift`, `find_timesteps`, and `kick` routines. In this section we outline the different components of the HOLD integration scheme and the changes made to adapt them to GPUs. Further references to these routines can also be found on the github repository.

`split` This routine subdivides the system into *slow* and *fast* subsystems based on a prescribed threshold timestep. Particles whose timesteps exceed the threshold form the slow subsystem, while those whose timesteps lie below or equal to the threshold form the fast subsystem. This division is a crucial part of the HOLD integrator. On a CPU, it is straightforward to rearrange particles with a single thread, but such an approach is inefficient on a GPU due to its parallel execution model.

To exploit GPU parallelism, we apply an algorithm called *stream compaction* (Nguyen, 2007), which selects elements matching a given condition by using the *Blelloch scan* (also known as exclusive scan). Although a detailed derivation of the Blelloch scan is beyond our scope, we refer interested readers to Nguyen (2007) for a thorough explanation.

In our `split` routine, we create two buffer arrays of particles: one for slow particles and one for fast particles. First, we check each particle's timestep relative to the threshold and build a *predicate array* containing 1 if the particle's timestep is below or equal to the threshold and 0 otherwise. Next, we perform an *exclusive scan* on this predicate array. In essence, the exclusive scan computes, for each position, the sum of all preceding predicate values. If the predicate at a given position is 1, the corresponding particle is placed into the fast buffer at the index specified by the scanned value. A similar process fills the slow buffer with particles whose timesteps exceed the threshold. Figure 6.4 provides a visual overview of this algorithm, illustrating how stream compaction enables efficient, fully parallel partitioning of particles into slow and fast subsystems on the GPU.

`drift` This routine updates each particle's position according to the equation

$$\mathbf{x}_{\frac{1}{2},i} = \mathbf{x}_{0,i} + 0.5\, dt\, \mathbf{v}_{0,i}, \tag{6.1}$$

where $\mathbf{x}_{0,i}$ and $\mathbf{v}_{0,i}$ denote the position and velocity, respectively, of the i^{th} particle at the initial step. The operation is performed independently for each particle, making it straightforward to parallelize: every GPU thread can simply compute the new position for a single particle, thereby distributing the workload efficiently across the device.

`find_timesteps` This routine calculates and updates each particle's timestep using a symmetrized approach. We adopt the step-size criteria from Pelupessy et al.

(2012a), which are constrained by both the freefall time,

$$\tau_{\text{freefall}} = \eta\sqrt{\frac{r_{ij}}{a_{ij}}}, \tag{6.2}$$

and the flyby time,

$$\tau_{\text{flyby}} = \eta\frac{r_{ij}}{v_{ij}}. \tag{6.3}$$

To incorporate time-symmetrization, we use the time derivatives of these expressions to construct a first-order accurate estimate of the future timestep:

$$\tau_{\text{sym}} = \tau\left(1 - \tfrac{1}{2}\tfrac{d\tau}{dt}\right)^{-1}. \tag{6.4}$$

Because both timestep criteria depend on interparticle distances and velocities, the overall calculation scales as $\mathcal{O}(N^2)$. A procedure similar to the one described in Section 6.2 can be employed with minimal modification. Note that the actual timestep for particle i is taken to be the minimum of its freefall and flyby timesteps.

`kick` This routine updates each particle's velocity according to

$$\mathbf{v}_{1,i} = \mathbf{v}_{0,i} + dt\,\mathbf{a}_{0,i}, \tag{6.5}$$

where $\mathbf{v}_{0,i}$ and $\mathbf{a}_{0,i}$ represent the velocity and acceleration, respectively, of the i^{th} particle at the initial step. The velocity "kick" relies on pairwise force calculations, which are performed using the $\mathcal{O}(N^2)$ method outlined in the previous section. Once the accelerations have been computed, each GPU thread can independently update the velocity of a single particle.

`update_system` This routine is unnecessary in the CPU version because slow and fast particles are not stored in separate arrays. However, in the GPU version, particles are split into distinct slow and fast buffers, requiring a mechanism to propagate updates between integration levels. The HOLD algorithm works recursively to update the fast subsystem and its interactions with the slow subsystem at each integration level. Although the positions and velocities for the particles in the current level are updated, these changes must be reflected in the parent system at the previous level for correct integration. To accomplish this, each particle in the slow and fast buffers is assigned a *parent ID* that indicates its corresponding index in the parent array. After a full DKD step, this routine is invoked to retrieve these parent IDs and update the parent system, ensuring that all relevant particle states remain consistent across integration levels.

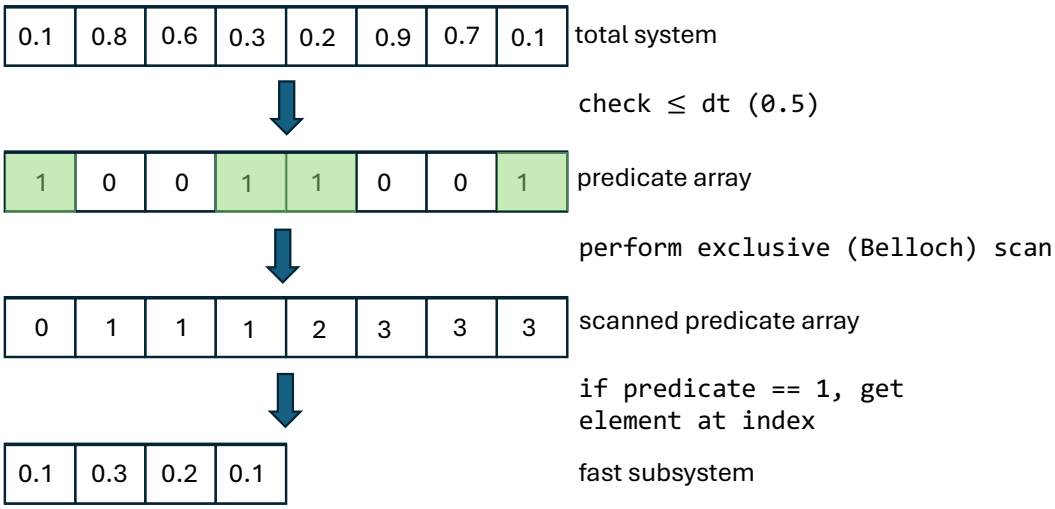

Figure 6.4: The splitting algorithm as described in section 6.3.1. We utilize a technique called stream compaction (Nguyen, 2007) to select particles that have only have timesteps smaller than a threshold timestep dt. To do so, we first check each particle's timestep relative to the threshold and build a predicate array containing 1 if the particle's timestep is below or equal to the threshold (dt) and 0 otherwise. Then an exclusive scan is performed to get a scanned predicate array which is then used to extract particles that belong to the fast and slow subsystems. The algorithm helps split the total system into two subsystems fast and slow as required by the HOLD algorithm. The fast subsystem contains particles with timesteps smaller than dt whereas the slow subsystem contains particles with timesteps larger than dt.

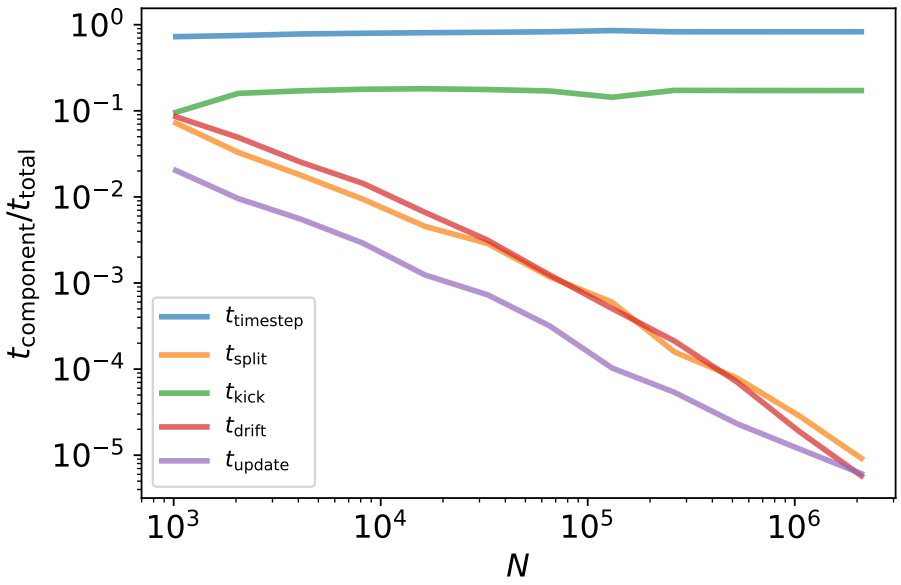

Figure 6.5: The ratio of the time taken per GPU routine to the total time taken as a function of the particle number N. We find that the `find_timesteps` routine is consistently the most expensive routine, taking up $\sim 80\%$ of the total time, given the $\mathcal{O}(N^2)$ scaling and the branching inefficiency in the GPU kernel. The `kick` routine is the next most expensive routine as it also involves pairwise calculations, taking up about $\sim 10 - 15\%$ of the total time. All other routines are subdominant at large N.

6.4 Results

All tests were performed on an A100 GPU on the Vera computing cluster using `fp32` precision. All kernels utilized a blocksize of 128 threads per block. For the CPU runs, we utilized CPU threads on an `Epyc 7742` node on the Vera computing cluster. Both pieces of code were compiled with `-O3` optimization flag.

6.4.1 Cost per module

Figure 6.5 shows the ratio of each routine's runtime to the overall simulation time. We find that *timestep determination* is by far the most expensive component, accounting for approximately 83% of the total runtime. This is not surprising given the $\mathcal{O}(N^2)$ complexity of evaluating pairwise interactions in conjunction with conditional expressions that introduce branching. Because GPUs are generally less efficient when dealing with branch-heavy kernels, adopting a simpler timestep criterion (e.g., the Aarseth condition) or reducing the number of conditionals could alleviate this bottleneck.

The next most time-consuming operation is the *kick* routine, which contributes 10-15% of the total runtime. Consequently, the combination of timestepping and kick routines dominates the computational budget. For smaller systems ($N \sim 10^3$), the *drift* and *split* modules can each constitute up to 5-10% of the overall runtime, but as N grows, these contributions become negligible relative to the cost of timestepping and kicking. Finally, the *update system* routine remains consistently efficient, never exceeding 1% of the total execution time.

6.4.2 Scaling and comparison with CPU version

To assess how our HOLD GPU implementation scales with the number of particles N, we perform a series of simulations using a Plummer model with an isotropic velocity distribution. We adopt code units such that $G = M = a = 1$, where M is the total mass of the system and a is its scale radius. The timestep parameter is fixed to $\eta = 0.2$, and each run proceeds until a final time of $t_{\text{final}} = 0.0625$.

As a reference, we compare our GPU code against a highly optimized CPU version of the HOLD integrator, also implemented in Taichi with direct force summation. The CPU version leverages OpenMP plus AVX2 vector instructions and runs on 64 threads. Our results indicate that the GPU code outperforms the CPU version for all tested N. For $N > 10^5$, the wall-clock time t_{wall} scales approximately as $\propto N^2$, but the GPU implementation consistently maintains a speedup factor of roughly $4\times$ over the CPU-based HOLD code. In comparison to a purely serial implementation (not shown here), our GPU approach can be over $200\times$ faster.

To verify correctness, we measure the relative difference in particle positions after t_{final}, finding a discrepancy on the order of 10^{-6}. This level of precision is consistent

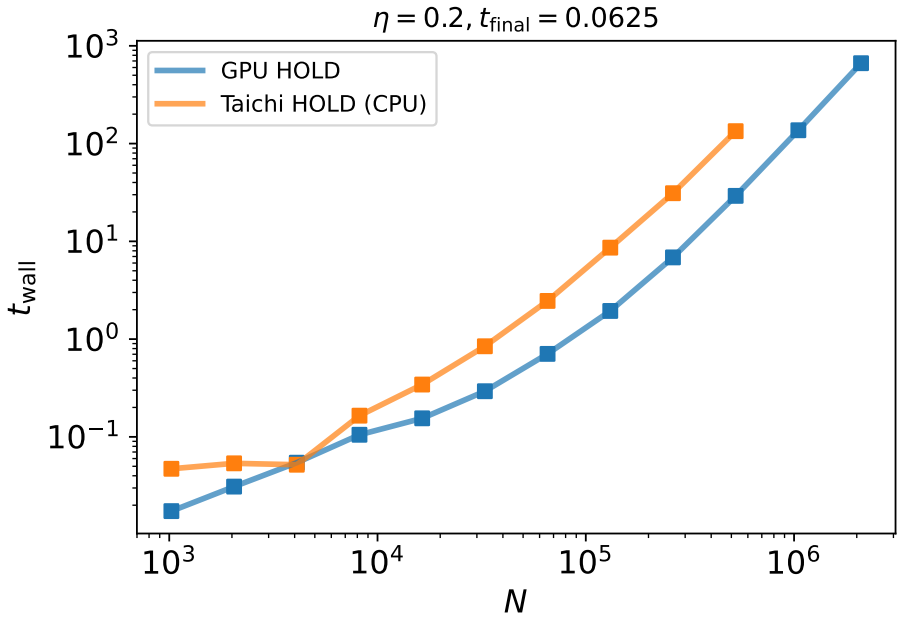

Figure 6.6: Wall-clock time t_{wall} taken to integrate to a final time of $t_{\text{final}} = 0.0625$ for our GPU version of HOLD (blue) and a highly optimized CPU version of HOLD in the N-body code `Taichi`. The latter was run with `AVX2` support with 64 OpenMP threads. We find that our GPU code performs consistently better than the CPU version with a speed-up of $4\times$ at $N > 10^5$.

with both single-precision arithmetic and the inherent miller instability in collisionless N-body systems. Overall, these tests confirm that our full-GPU HOLD integrator achieves both efficiency and accuracy, comparing favorably with a state-of-the-art CPU-based code.

6.5 Discussion

6.5.1 Performance bottlenecks and potential improvements

The present GPU implementation of the HOLD integrator serves primarily as a proof-of-concept, demonstrating that the entire algorithm can run efficiently and exclusively on GPUs. However, several inefficiencies remain. The most significant arises from the timestep determination process: ideally, we would want the bulk of the computational time to be spent in force evaluations, but the timestepping kernel currently involves multiple branching operations for symmetrization and for selecting the global minimum timestep. Because GPUs handle branch divergence poorly, a simpler criterion (e.g., the Aarseth condition) or an approach without symmetrization could mitigate this issue. Additional optimizations—such as manually unrolling kernels to process multiple particles per thread and reducing shared-memory bank conflicts—could further boost performance.

Another performance bottleneck stems from the way we manage *slow* and *fast* buffers at each integration level. For very small timestep parameters (η), the HOLD scheme produces a large number of integration levels, leading to multiple buffer creations (`cudaMalloc` calls) that both consume run-time and risk exhausting the GPU's available VRAM. Indeed, experiments reveal that for $N > 524{,}288$ and $\eta = 0.02$, the GPU can run out of memory even with 40 GB of VRAM. One straightforward solution is to maintain only two sets of buffers—one for the current level and one for the previous level—by allocating a fixed set of four global arrays (slow and fast subarrays for each of the two levels). We plan to explore this strategy in future work to determine whether it substantially enhances performance.

6.5.2 Future work

Although the current GPU-exclusive HOLD implementation represents a step forward, several promising avenues for further development remain:

1. **Multi-GPU Scaling:** Adapting the code to run on multiple GPUs in parallel would enable simulations with $N \gtrsim 10^7$ particles. A logical approach here is to employ *CUDA-aware MPI* to handle inter-GPU communication efficiently, potentially scaling to clusters or exascale systems.

2. **GPU-Based FMM Solver:** Replacing the current direct summation force solver with a FMM) implementation could drastically reduce computational

cost, particularly for large N. Atkinson (2021) have demonstrated a fully GPU-resident FMM code using a custom task-stealing scheduler, highlighting the feasibility of this approach. However, achieving sufficient accuracy for collisional dynamics may require careful optimization of the multipole acceptance criteria and specialized adjustments for small-scale interactions.

3. **Higher-Order Integrators:** A natural extension of our second-order HOLD scheme would be a fourth-order generalization using *HHS-FSI* (Chin and Chen, 2005; Rantala et al., 2021a). This would preserve the general spirit of Hamiltonian splitting while offering improved energy conservation, particularly for long-term integrations of collisional systems.

Exploring these directions could unlock even greater performance gains and scientific reach for GPU-accelerated N-body simulations, particularly in the realm of dense star clusters and massive black hole dynamics.

6.6 Conclusions

In this work, we have demonstrated the feasibility of a fully GPU-resident N-body integrator based on a second-order Hamiltonian decomposition. Our principal findings are summarized as follows:

- **Memory Transfer Bottlenecks.** By keeping the entire integration process on the GPU, we eliminate most CPU–GPU data transfers. This is especially crucial for the low-N regime, where more than 95% of the integration levels contain fewer than 10^4 particles.
- **Timestep Overhead.** In our implementation, the timestepping routine accounts for over 80% of the total wall-clock time, largely due to its $\mathcal{O}(N^2)$ complexity and the branching operations required to symmetrize the timesteps.
- **Minor Cost of Other Modules.** Remaining modules (e.g., drift and split operations) collectively consume under 5% of the total runtime, indicating that timestepping optimizations would yield the largest performance gains.
- **Comparison with CPU Parallelization.** Our GPU-based HOLD scheme retains $\mathcal{O}(N^2)$ scaling but delivers a roughly $4\times$ speedup relative to a highly optimized CPU version featuring OpenMP and SIMD parallelization.

Several opportunities exist for further optimization, including a more efficient timestepping criterion and reducing branching in kernel code. Moreover, combining our GPU integration framework with recent developments in fully GPU-resident *fast multipole methods* (FMM) could pave the way for a comprehensive, high-accuracy collisional N-body code running *exclusively* on GPUs.

Chapter 7

Conclusion

In this thesis, I develop and test a state-of-the-art fast multipole method (FMM)-based N-body code tailored for collisional dynamics, which is then used to study the evolution of massive black hole binaries (MBHBs) within dense nuclear star clusters (NSCs). Building on this, I construct an approximate N-body code with post-Newtonian corrections to explore intermediate mass ratio inspirals (IMRIs) in dark matter (DM) spikes, probing the detectability of gravitational waves shaped by such dense DM environments. Finally, I present a proof-of-concept Hamiltonian splitting integrator running entirely on GPUs, delivering performance exceeding a parallelized CPU-based implementation.

The collisional N-body code, `Taichi`, achieves $\mathcal{O}(N)$ scaling by splitting forces into short-range (direct summation) and long-range (FMM) components. By tuning the expansion order p and accuracy parameter ϵ, force errors remain well-controlled, and global properties, such as Lagrangian radii, density profiles, and dynamical friction times, agree with both direct-summation results and analytic predictions. While FMM becomes efficient only for $N \gtrsim 10^4$, it provides over $100\times$ speedup compared to direct summation and runs more than $10\times$ faster than `NBODY6++GPU` on a 28-core system.

Building on these developments, we upgrade `Taichi` with a fourth-order forward-symplectic integrator and few-body regularization to resolve hard binaries. Using this, I study how collisional relaxation and mass segregation in NSCs affect MBH binary merger timescales across various mass ratios. With a two-component mass function in N-body simulations, we find that relaxed cusps modify density profiles and lead to unexpected trends, such as lower binary separations in non-relaxed systems for $q > 10^{-2}$ and suppressed eccentricities in relaxed cases. For circular orbits, relaxed cusps slow down coalescence for higher q, but accelerate it for $q \leq 10^{-2}$, highlighting the importance of incorporating realistic mass spectra and dynamically evolved NSCs. For eccentric orbits, MBHBs merge $50 - 100\%$faster in non-relaxed models. Across all cases, the merger times remain below the Hubble time, underscoring the need for high-resolution collisional modeling in MBH merger and waveform predictions.

Motivated by these accelerated mergers, I examine MBH seed dynamics in high-redshift dwarf galaxies using initial conditions from the `ASTRID` cosmological simulation. High-resolution N-body simulations (up to 10^7 particles) with `Taichi` reveal that dense NSCs help MBHs reach sub-pc scales within a few tens of Myr, forming hard binaries much faster than when MBHs lack extended stellar envelopes. The final NSCs match observed nucleated dwarfs in density and size. Hardening rate analysis shows stellar encounters dominate in denser clusters, while DM plays a larger role in lower-mass systems. Bound binary eccentricities reach $e > 0.9$ in half of the models, though stochasticity makes outcomes sensitive to initial conditions. Semi-analytic extrapolations show that high-eccentricity MBHs in dense NSCs can merge before $z = 4$, emphasizing the role of NSCs in resolving the seed-sinking problem. Alongside prior findings that "naked" MBHs do not sink effectively, this highlights the crucial role of NSCs in enabling MBH coalescence.

Recent studies suggest dense DM spikes can form around IMBHs, potentially altering GW signals via phase shifts. I investigate IMRIs in both non-rotating and rotating DM spikes, focusing on how the spike's self-consistent response affects binary evolution. N-body simulations show that three-body scattering—rather than dynamical friction—drives most orbital decay, resulting in $\geq 10^4$ dephasing cycles in steep spikes ($\gamma_{\rm sp} = 7/3, 9/4$) but only ~ 10 in shallower, collisional environments ($\gamma_{\rm sp} = 3/2$). Rotating spikes show enhanced eccentricity growth and $3\times$ more dephasing cycles than non-rotating ones, while prograde motion has the opposite effect. Relativistic precession has minimal impact on non-rotating cases but suppresses rotational effects. A preliminary Fokker–Planck analysis suggests real-world spikes likely have lower densities ($\gamma_{\rm sp} = 3/2$ or less) due to prior inspirals and environmental evolution, thus producing weaker dephasing. However, "deshifting" remains potentially detectable under moderate conditions, reinforcing the value of multi-band GW detectors and the need for simulations with higher-order post-Newtonian effects, accretion, and spike-regeneration.

Lastly, I showcase a second-order Hamiltonian decomposition `HOLD` N-body integrator running entirely on GPUs, eliminating CPU–GPU data transfers and benefiting low-N systems, where over 95% of integration levels involve fewer than 10^4 particles. The timestepping kernel dominates the runtime (over 80%) due to its $\mathcal{O}(N^2)$ complexity and branching overhead, while drift and split modules remain under 5%. Compared to a highly optimized CPU version using `OpenMP` and `SIMD`, the GPU implementation achieves a $4\times$ speedup. Further gains may come from optimizing the timestepping routine and reducing branching, with future extensions incorporating GPU-resident FMM to deliver a fully GPU-based, high-accuracy N-body code for large-scale collisional simulations.

This thesis presents several computational advancements in N-body methods and theoretical modeling, specifically tailored to the study of MBH binaries and IMRIs. It lays the foundation for future investigations into the role of NSCs in shaping the MBH merger landscape, particularly at high redshifts, and underscores the necessity

of realistic DM spike modeling for accurate GW dephasing predictions. Our findings emphasize the pivotal role of NSCs in driving the dynamical evolution of MBHs, especially seed MBHs, demonstrating that the long-standing seed-sinking problem can be significantly alleviated by the presence of extended stellar environments. With future improvements to `Taichi`, such as enhancements to the FMM kernels, full GPU implementation using the framework developed in this thesis, and the incorporation of additional physics, such as post-Newtonian effects and GW recoil, we establish `Taichi` as a powerful and versatile tool for studying dense collisional N-body systems.

Bibliography

Fazeel Mahmood Khan, Peter Berczik, and Andreas Just. Gravitational wave driven mergers and coalescence time of supermassive black holes. A&A, 615:A71, July 2018. doi:10.1051/0004-6361/201730489.

P. C. Peters. Gravitational Radiation and the Motion of Two Point Masses. *Physical Review*, 136(4B):1224–1232, November 1964a. doi:10.1103/PhysRev.136.B1224.

Diptajyoti Mukherjee, Qirong Zhu, Go Ogiya, Carl L. Rodriguez, and Hy Trac. Evolution of massive black hole binaries in collisionally relaxed nuclear star clusters - Impact of mass segregation. MNRAS, 518(4):4801–4817, February 2023. doi:10.1093/mnras/stac3390.

Rio Yokota and Lorena A Barba. A tuned and scalable fast multipole method as a preeminent algorithm for exascale systems. *The International Journal of High Performance Computing Applications*, 26(4):337–346, 2012.

Walter Dehnen. A hierarchical o (n) force calculation algorithm. *Journal of Computational Physics*, 179(1):27–42, 2002.

Junichiro Makino, Piet Hut, Murat Kaplan, and Hasan Saygın. A time-symmetric block time-step algorithm for N-body simulations. New A, 12(2):124–133, November 2006a. doi:10.1016/j.newast.2006.06.003.

Federico I. Pelupessy, Jürgen Jänes, and Simon Portegies Zwart. N-body integrators with individual time steps from Hierarchical splitting. New A, 17(8):711–719, November 2012a. doi:10.1016/j.newast.2012.05.009.

Alessia Gualandris and David Merritt. Long-term Evolution of Massive Black Hole Binaries. IV. Mergers of Galaxies with Collisionally Relaxed Nuclei. ApJ, 744(1):74, January 2012a. doi:10.1088/0004-637X/744/1/74.

Miguel Preto, Ingo Berentzen, Peter Berczik, and Rainer Spurzem. Fast Coalescence of Massive Black Hole Binaries from Mergers of Galactic Nuclei: Implications for Low-frequency Gravitational-wave Astrophysics. ApJ, 732(2):L26, May 2011. doi:10.1088/2041-8205/732/2/L26.

Go Ogiya, Oliver Hahn, Chiara M. F. Mingarelli, and Marta Volonteri. Accelerated orbital decay of supermassive black hole binaries in merging nuclear star clusters. MNRAS, 493 (3):3676–3689, April 2020a. doi:10.1093/mnras/staa444.

Yihao Zhou, Diptajyoti Mukherjee, Nianyi Chen, Tiziana Di Matteo, Peter H. Johansson, Antti Rantala, Christian Partmann, Ugo NiccolDi Carlo, Simeon Bird, and Yueying Ni. MAGICS II. Seed black holes stripped of their surrounding stars do not sink. *arXiv e-prints*, art. arXiv:2409.19914, September 2024. doi:10.48550/arXiv.2409.19914.

Nianyi Chen, Diptajyoti Mukherjee, Tiziana Di Matteo, Yueying Ni, Simeon Bird, and Rupert Croft. MAGICS I. The First Few Orbits Encode the Fate of Seed Massive Black Hole Pairs. *The Open Journal of Astrophysics*, 7:28, April 2024. doi:10.33232/001c.116179.

Dieu D. Nguyen, Anil C. Seth, Mark den Brok, Nadine Neumayer, Michele Cappellari, Aaron J. Barth, Nelson Caldwell, Benjamin F. Williams, and Breanna Binder. Improved Dynamical Constraints on the Mass of the Central Black Hole in NGC 404. ApJ, 836(2): 237, February 2017. doi:10.3847/1538-4357/aa5cb4.

Dieu D. Nguyen, Anil C. Seth, Nadine Neumayer, Sebastian Kamann, Karina T. Voggel, Michele Cappellari, Arianna Picotti, Phuong M. Nguyen, Torsten Böker, Victor Debattista, Nelson Caldwell, Richard McDermid, Nathan Bastian, Christopher C. Ahn, and Renuka Pechetti. Nearby Early-type Galactic Nuclei at High Resolution: Dynamical Black Hole and Nuclear Star Cluster Mass Measurements. ApJ, 858(2):118, May 2018. doi:10.3847/1538-4357/aabe28.

Iskren Y. Georgiev, Torsten Böker, Nathan Leigh, Nora Lützgendorf, and Nadine Neumayer. Masses and scaling relations for nuclear star clusters, and their co-existence with central black holes. MNRAS, 457(2):2122–2138, April 2016. doi:10.1093/mnras/stw093.

David Merritt, Seppo Mikkola, and Andras Szell. Long-Term Evolution of Massive Black Hole Binaries. III. Binary Evolution in Collisional Nuclei. ApJ, 671(1):53–72, December 2007a. doi:10.1086/522691.

Fazeel Mahmood Khan and Kelly Holley-Bockelmann. Extremely efficient mergers of intermediate-mass black hole binaries in nucleated dwarf galaxies. MNRAS, 508(1):1174–1188, November 2021. doi:10.1093/mnras/stab2646.

Nadine Neumayer, Anil Seth, and Torsten Böker. Nuclear star clusters. A&A Rev., 28(1): 4, July 2020a. doi:10.1007/s00159-020-00125-0.

Bradley J. Kavanagh, David A. Nichols, Gianfranco Bertone, and Daniele Gaggero. Detecting dark matter around black holes with gravitational waves: Effects of dark-matter dynamics on the gravitational waveform. Phys. Rev. D, 102(8):083006, October 2020. doi:10.1103/PhysRevD.102.083006.

P. C. Peters. Gravitational Radiation and the Motion of Two Point Masses. *Physical Review*, 136(4B):1224–1232, November 1964b. doi:10.1103/PhysRev.136.B1224.

Hubert Nguyen. *Gpu gems 3*. Addison-Wesley Professional, 2007.

Lyman Spitzer. *Dynamical evolution of globular clusters*. 1987.

Douglas Heggie and Piet Hut. *The Gravitational Million-Body Problem: A Multidisciplinary Approach to Star Cluster Dynamics*. 2003.

Sverre J Aarseth. *Gravitational N-body simulations: tools and algorithms*. Cambridge University Press, 2003.

James Binney and Scott Tremaine. *Galactic dynamics*. Princeton university press, 2011.

Walter Dehnen and Justin I Read. N-body simulations of gravitational dynamics. *The European Physical Journal Plus*, 126(5):1–28, 2011.

Sverre J Aarseth. From nbody1 to nbody6: The growth of an industry. *Publications of the Astronomical Society of the Pacific*, 111(765):1333, 1999.

A Ahmad and L Cohen. A numerical integration scheme for the n-body gravitational problem. *Journal of Computational Physics*, 12(3):389–402, 1973.

Josh Barnes and Piet Hut. A hierarchical o (n log n) force-calculation algorithm. *nature*, 324(6096):446–449, 1986.

Long Wang, Masaki Iwasawa, Keigo Nitadori, and Junichiro Makino. petar: a high-performance n-body code for modelling massive collisional stellar systems. *Monthly Notices of the Royal Astronomical Society*, 497(1):536–555, 2020.

Leslie Greengard and Vladimir Rokhlin. A fast algorithm for particle simulations. *Journal of computational physics*, 73(2):325–348, 1987.

Hongwei Cheng, Leslie Greengard, and Vladimir Rokhlin. A fast adaptive multipole algorithm in three dimensions. *Journal of computational physics*, 155(2):468–498, 1999.

Walter Dehnen. A very fast and momentum-conserving tree code. *The Astrophysical Journal Letters*, 536(1):L39, 2000.

Walter Dehnen. A fast multipole method for stellar dynamics. *Computational Astrophysics and Cosmology*, 1(1):1, 2014.

Piet Hut, Jun Makino, and Steve McMillan. Building a Better Leapfrog. ApJ, 443:L93, April 1995. doi:10.1086/187844.

Walter Dehnen. Towards time symmetric n-body integration. *Monthly Notices of the Royal Astronomical Society*, 472(1):1226–1238, 2017.

Siu A. Chin and C. R. Chen. Forward Symplectic Integrators for Solving Gravitational Few-Body Problems. *Celestial Mechanics and Dynamical Astronomy*, 91(3-4):301–322, March 2005. doi:10.1007/s10569-004-4622-z.

Antti Rantala, Thorsten Naab, and Volker Springel. FROST: a momentum-conserving CUDA implementation of a hierarchical fourth-order forward symplectic integrator. MNRAS, January 2021a. doi:10.1093/mnras/stab057.

Haruo Yoshida. Construction of higher order symplectic integrators. *Physics Letters A*, 150 (5-7):262–268, November 1990. doi:10.1016/0375-9601(90)90092-3.

I. P. Omelyan. Extrapolated gradientlike algorithms for molecular dynamics and celestial mechanics simulations. Phys. Rev. E, 74(3):036703, September 2006. doi:10.1103/PhysRevE.74.036703.

Will M. Farr and Edmund Bertschinger. Variational Integrators for the Gravitational N-Body Problem. ApJ, 663(2):1420–1433, July 2007. doi:10.1086/518641.

Long Wang, Keigo Nitadori, and Junichiro Makino. A slow-down time-transformed symplectic integrator for solving the few-body problem. MNRAS, 493(3):3398–3411, April 2020. doi:10.1093/mnras/staa480.

P. Kustaanheimo, A. SCHINZEL, H. DAVENPORT, and E. STIEFEL. Perturbation theory of kepler motion based on spinor regularization. *Journal für die reine und angewandte Mathematik*, 1965(218):204–219, 1965. doi:doi:10.1515/crll.1965.218.204. URL `https://doi.org/10.1515/crll.1965.218.204`.

Seppo Mikkola and Kiyotaka Tanikawa. Explicit Symplectic Algorithms For Time-Transformed Hamiltonians. *Celestial Mechanics and Dynamical Astronomy*, 74(4):287–295, August 1999a. doi:10.1023/A:1008368322547.

Miguel Preto and Scott Tremaine. A Class of Symplectic Integrators with Adaptive Time Step for Separable Hamiltonian Systems. AJ, 118(5):2532–2541, November 1999. doi:10.1086/301102.

William B. Gragg. On Extrapolation Algorithms for Ordinary Initial Value Problems. *SIAM Journal on Numerical Analysis*, 2(3):384–403, January 1965. doi:10.1137/0702030.

W. H. Press. *Techniques and Tricks for N-Body Computation*, volume 267, page 184. 1986. doi:10.1007/BFb0116411.

Seppo Mikkola and Sverre J. Aarseth. A chain regularization method for the few-body problem. *Celestial Mechanics and Dynamical Astronomy*, 47(4):375–390, January 1990.

Yi-Han Wang, Nathan W. C. Leigh, Bin Liu, and Rosalba Perna. SpaceHub: A high-performance gravity integration toolkit for few-body problems in astrophysics. MNRAS, 505(1):1053–1070, July 2021a. doi:10.1093/mnras/stab1189.

M. Hénon. Monte Carlo Models of Star Clusters (Part of the Proceedings of the IAU Colloquium No. 10, held in Cambridge, England, August 12-15, 1970.). Ap&SS, 13(2): 284–299, October 1971a. doi:10.1007/BF00649159.

Lyman Spitzer, Jr. and Michael H. Hart. Random Gravitational Encounters and the Evolution of Spherical Systems. II. Models. ApJ, 166:483, June 1971. doi:10.1086/150977.

Arkadiusz Hypki and Mirek Giersz. MOCCA code for star cluster simulations - I. Blue stragglers, first results. MNRAS, 429(2):1221–1243, February 2013. doi:10.1093/mnras/sts415.

Kriten J Joshi, Frederic A Rasio, and Simon Portegies Zwart. Monte carlo simulations of globular cluster evolution. i. method and test calculations. *The Astrophysical Journal*, 540(2):969, 2000.

B Pattabiraman, S Umbreit, Liao Wk, A Choudhary, V Kalogera, G Memik, and FA Rasio. A parallel monte carlo code for simulating collisional n n-body systems. *Astrophys J Suppl*, 204:15, 2013.

Eugene Vasiliev. A new Monte Carlo method for dynamical evolution of non-spherical stellar systems. MNRAS, 446(3):3150–3161, January 2015. doi:10.1093/mnras/stu2360.

Pau Amaro-Seoane, Heather Audley, Stanislav Babak, John Baker, Enrico Barausse, Peter Bender, Emanuele Berti, Pierre Binetruy, Michael Born, Daniele Bortoluzzi, et al. Laser interferometer space antenna. *arXiv preprint arXiv:1702.00786*, 2017.

Jun Luo, Li-Sheng Chen, Hui-Zong Duan, Yun-Gui Gong, Shoucun Hu, Jianghui Ji, Qi Liu, Jianwei Mei, Vadim Milyukov, Mikhail Sazhin, et al. Tianqin: a space-borne gravitational wave detector. *Classical and Quantum Gravity*, 33(3):035010, 2016.

Tiziana Di Matteo, Daniel Angles-Alcazar, and Francesco Shankar. Massive black holes in galactic nuclei: Theory and Simulations. *arXiv e-prints*, art. arXiv:2304.11541, April 2023. doi:10.48550/arXiv.2304.11541.

M. C. Begelman, R. D. Blandford, and M. J. Rees. Massive black hole binaries in active galactic nuclei. Nature, 287(5780):307–309, September 1980. doi:10.1038/287307a0.

S. Chandrasekhar. Dynamical Friction. I. General Considerations: the Coefficient of Dynamical Friction. ApJ, 97:255, March 1943. doi:10.1086/144517.

Linhao Ma, Philip F. Hopkins, Xiangcheng Ma, Daniel Anglés-Alcázar, Claude-André Faucher-Giguère, and Luke Zoltan Kelley. Seeds don't sink: even massive black hole 'seeds' cannot migrate to galaxy centres efficiently. MNRAS, 508(2):1973–1985, December 2021. doi:10.1093/mnras/stab2713.

David Merritt. *Dynamics and evolution of galactic nuclei*, volume 23. Princeton University Press, 2013.

Gerald D. Quinlan. The dynamical evolution of massive black hole binaries I. Hardening in a fixed stellar background. New A, 1(1):35–56, July 1996. doi:10.1016/S1384-1076(96)00003-6.

Alberto Sesana, Francesco Haardt, and Piero Madau. Interaction of Massive Black Hole Binaries with Their Stellar Environment. I. Ejection of Hypervelocity Stars. ApJ, 651 (1):392–400, November 2006. doi:10.1086/507596.

Miloš Milosavljević and David Merritt. The Final Parsec Problem. In Joan M. Centrella, editor, *The Astrophysics of Gravitational Wave Sources*, volume 686 of *American Institute of Physics Conference Series*, pages 201–210, October 2003. doi:10.1063/1.1629432.

Peter Berczik, David Merritt, Rainer Spurzem, and Hans-Peter Bischof. Efficient Merger of Binary Supermassive Black Holes in Nonaxisymmetric Galaxies. ApJ, 642(1):L21–L24, May 2006. doi:10.1086/504426.

Fazeel Mahmood Khan, Kelly Holley-Bockelmann, Peter Berczik, and Andreas Just. Supermassive Black Hole Binary Evolution in Axisymmetric Galaxies: The Final Parsec Problem is Not a Problem. ApJ, 773(2):100, August 2013. doi:10.1088/0004-637X/773/2/100.

Eugene Vasiliev, Fabio Antonini, and David Merritt. The Final-parsec Problem in the Collisionless Limit. ApJ, 810(1):49, September 2015. doi:10.1088/0004-637X/810/1/49.

Eugene Vasiliev. A New Fokker-Planck Approach for the Relaxation-driven Evolution of Galactic Nuclei. ApJ, 848(1):10, October 2017. doi:10.3847/1538-4357/aa8cc8.

Alberto Sesana, Alessia Gualandris, and Massimo Dotti. Massive black hole binary eccentricity in rotating stellar systems. MNRAS, 415(1):L35–L39, July 2011. doi:10.1111/j.1745-3933.2011.01073.x.

Imran Nasim, Alessia Gualandris, Justin Read, Walter Dehnen, Maxime Delorme, and Fabio Antonini. Defeating stochasticity: coalescence time-scales of massive black holes in galaxy mergers. MNRAS, 497(1):739–746, September 2020. doi:10.1093/mnras/staa1896.

Alexander Rawlings, Matias Mannerkoski, Peter H. Johansson, and Thorsten Naab. Reviving stochasticity: uncertainty in SMBH binary eccentricity is unavoidable. MNRAS, 526 (2):2688–2695, December 2023. doi:10.1093/mnras/stad2891.

David Merritt. *Dynamics and Evolution of Galactic Nuclei*. 2013a.

M. Tremmel, F. Governato, M. Volonteri, and T. R. Quinn. Off the beaten path: a new approach to realistically model the orbital decay of supermassive black holes in galaxy formation simulations. MNRAS, 451(2):1868–1874, August 2015a. doi:10.1093/mnras/stv1060.

Nianyi Chen, Yueying Ni, Michael Tremmel, Tiziana Di Matteo, Simeon Bird, Colin DeGraf, and Yu Feng. Dynamical friction modelling of massive black holes in cosmological simulations and effects on merger rate predictions. MNRAS, 510(1):531–550, February 2022a. doi:10.1093/mnras/stab3411.

Linhao Ma, Philip F. Hopkins, Luke Zoltan Kelley, and Claude-André Faucher-Giguère. A new discrete dynamical friction estimator based on N-body simulations. MNRAS, 519 (4):5543–5553, March 2023. doi:10.1093/mnras/stad036.

Christian Partmann, Thorsten Naab, Natalia Lahén, Antti Rantala, Michaela Hirschmann, Jessica M. Hislop, Jonathan Petersson, and Peter H. Johansson. The importance of nuclear star clusters for massive black hole growth and nuclear star formation in simulated low-mass galaxies. MNRAS, 537(2):956–977, February 2025. doi:10.1093/mnras/staf002.

Joel Pfeffer, J. M. Diederik Kruijssen, Robert A. Crain, and Nate Bastian. The E-MOSAICS project: simulating the formation and co-evolution of galaxies and their star cluster populations. MNRAS, 475(4):4309–4346, April 2018. doi:10.1093/mnras/stx3124.

Carl L. Rodriguez, Zachary Hafen, Michael Y. Grudić, Astrid Lamberts, Kuldeep Sharma, Claude-André Faucher-Giguère, and Andrew Wetzel. Great balls of FIRE II: The evolution and destruction of star clusters across cosmic time in a Milky Way-mass galaxy. MNRAS, 521(1):124–147, May 2023. doi:10.1093/mnras/stad578.

H. H. Loose, E. Kruegel, and A. Tutukov. Bursts of star formation in the galactic centre. A&A, 105(2):342–350, January 1982.

Philip F. Hopkins and Eliot Quataert. How do massive black holes get their gas? MNRAS, 407(3):1529–1564, September 2010. doi:10.1111/j.1365-2966.2010.17064.x.

Philip F. Hopkins and Eliot Quataert. An explanation for the slopes of stellar cusps in galaxy spheroids. MNRAS, 411(1):L61–L65, February 2011. doi:10.1111/j.1745-3933.2010.00995.x.

Nicolas Guillard, Eric Emsellem, and Florent Renaud. New insights on the formation of nuclear star clusters. MNRAS, 461(4):3620–3629, October 2016. doi:10.1093/mnras/stw1570.

Gillen Brown, Oleg Y. Gnedin, and Hui Li. Nuclear Star Clusters in Cosmological Simulations. ApJ, 864(1):94, September 2018. doi:10.3847/1538-4357/aad595.

S. D. Tremaine, J. P. Ostriker, and L. Spitzer, Jr. The formation of the nuclei of galaxies. I. M31. ApJ, 196:407–411, March 1975. doi:10.1086/153422.

Markus Hartmann, Victor P. Debattista, Anil Seth, Michele Cappellari, and Thomas R. Quinn. Constraining the role of star cluster mergers in nuclear cluster formation: simulations confront integral-field data. MNRAS, 418(4):2697–2714, December 2011. doi:10.1111/j.1365-2966.2011.19659.x.

Fabio Antonini and Hagai B. Perets. Secular Evolution of Compact Binaries near Massive Black Holes: Gravitational Wave Sources and Other Exotica. ApJ, 757(1):27, September 2012. doi:10.1088/0004-637X/757/1/27.

Konstantinos Kritos, Emanuele Berti, and Joseph Silk. Massive black hole assembly in nuclear star clusters. Phys. Rev. D, 108(8):083012, October 2023. doi:10.1103/PhysRevD.108.083012.

Sandor Van Wassenhove, Pedro R. Capelo, Marta Volonteri, Massimo Dotti, Jillian M. Bellovary, Lucio Mayer, and Fabio Governato. Nuclear coups: dynamics of black holes in galaxy mergers. MNRAS, 439(1):474–487, March 2014. doi:10.1093/mnras/stu024.

Go Ogiya, Oliver Hahn, Chiara M. F. Mingarelli, and Marta Volonteri. Accelerated orbital decay of supermassive black hole binaries in merging nuclear star clusters. MNRAS, 493 (3):3676–3689, April 2020b. doi:10.1093/mnras/staa444.

Yanlong Shi, Kyle Kremer, and Philip F. Hopkins. From Seed to Supermassive Black Holes: Capture, Growth, Migration, and Pairing in Dense Proto-Bulge Environments. *arXiv e-prints*, art. arXiv:2405.17338, May 2024. doi:10.48550/arXiv.2405.17338.

Paolo Gondolo and Joseph Silk. Dark Matter Annihilation at the Galactic Center. Phys. Rev. Lett., 83(9):1719–1722, August 1999. doi:10.1103/PhysRevLett.83.1719.

Kazunari Eda, Yousuke Itoh, Sachiko Kuroyanagi, and Joseph Silk. New probe of dark-matter properties: Gravitational waves from an intermediate-mass black hole embedded in a dark-matter minispike. *Phys. Rev. Lett.*, 110:221101, May 2013. doi:10.1103/PhysRevLett.110.221101. URL https://link.aps.org/doi/10.1103/PhysRevLett.110.221101.

Kazunari Eda, Yousuke Itoh, Sachiko Kuroyanagi, and Joseph Silk. Gravitational waves as a probe of dark matter minispikes. Phys. Rev. D, 91(4):044045, February 2015. doi:10.1103/PhysRevD.91.044045.

Niklas Becker, Laura Sagunski, Lukas Prinz, and Saeed Rastgoo. Circularization versus eccentrification in intermediate mass ratio inspirals inside dark matter spikes. Phys. Rev. D, 105(6):063029, March 2022. doi:10.1103/PhysRevD.105.063029.

Niklas Becker and Laura Sagunski. Comparing accretion disks and dark matter spikes in intermediate mass ratio inspirals. Phys. Rev. D, 107(8):083003, April 2023. doi:10.1103/PhysRevD.107.083003.

Diptajyoti Mukherjee, Qirong Zhu, Hy Trac, and Carl L. Rodriguez. Fast Multipole Methods for N-body Simulations of Collisional Star Systems. ApJ, 916(1):9, July 2021a. doi:10.3847/1538-4357/ac03b2.

Diptajyoti Mukherjee, A. Miguel Holgado, Go Ogiya, and Hy Trac. Examining the effects of dark matter spikes on eccentric intermediate-mass ratio inspirals using N-body simulations. MNRAS, 533(2):2335–2355, September 2024. doi:10.1093/mnras/stae1989.

Diptajyoti Mukherjee, Yihao Zhou, Nianyi Chen, Ugo Niccolò Di Carlo, and Tiziana Di Matteo. MAGICS. III. Seeds Sink Swiftly: Nuclear Star Clusters Dramatically Accelerate Seed Black Hole Mergers. ApJ, 981(2):203, March 2025. doi:10.3847/1538-4357/adb1b0.

G. W. Clark. X-ray binaries in globular clusters. ApJ, 199:L143–L145, August 1975. doi:10.1086/181869.

Melvyn B. Davies and Brad M. S. Hansen. Neutron star retention and millisecond pulsar production in globular clusters. MNRAS, 301(1):15–24, November 1998. doi:10.1046/j.1365-8711.1998.01923.x.

N. Ivanova, C. O. Heinke, F. A. Rasio, K. Belczynski, and J. M. Fregeau. Formation and evolution of compact binaries in globular clusters - II. Binaries with neutron stars. MNRAS, 386(1):553–576, May 2008. doi:10.1111/j.1365-2966.2008.13064.x.

Charles J. Hailey, Kaya Mori, Franz E. Bauer, Michael E. Berkowitz, Jaesub Hong, and Benjamin J. Hord. A density cusp of quiescent X-ray binaries in the central parsec of the Galaxy. Nature, 556(7699):70–73, April 2018. doi:10.1038/nature25029.

S. Rappaport, A. Putney, and F. Verbunt. Evolution of Wide Binary Millisecond Pulsars in Globular Clusters. ApJ, 345:210, October 1989. doi:10.1086/167897.

Shrinivas R. Kulkarni, Ramesh Narayan, and Roger W. Romani. The Pulsar Content of Globular Clusters. ApJ, 356:174, June 1990. doi:10.1086/168828.

Steinn Sigurdsson and E. S. Phinney. Dynamics and Interactions of Binaries and Neutron Stars in Globular Clusters. ApJS, 99:609, August 1995. doi:10.1086/192199.

Claire S. Ye, Kyle Kremer, Sourav Chatterjee, Carl L. Rodriguez, and Frederic A. Rasio. Millisecond Pulsars and Black Holes in Globular Clusters. ApJ, 877(2):122, June 2019. doi:10.3847/1538-4357/ab1b21.

N. Ivanova, C. O. Heinke, F. A. Rasio, R. E. Taam, K. Belczynski, and J. Fregeau. Formation and evolution of compact binaries in globular clusters - I. Binaries with white dwarfs. MNRAS, 372(3):1043–1059, November 2006. doi:10.1111/j.1365-2966.2006.10876.x.

David Pooley and Piet Hut. Dynamical Formation of Close Binaries in Globular Clusters: Cataclysmic Variables. ApJ, 646(2):L143–L146, August 2006. doi:10.1086/507027.

Simon F. Portegies Zwart and Stephen L. W. McMillan. Black Hole Mergers in the Universe. ApJ, 528(1):L17–L20, January 2000. doi:10.1086/312422.

Carl L. Rodriguez, Meagan Morscher, Bharath Pattabiraman, Sourav Chatterjee, Carl-Johan Haster, and Frederic A. Rasio. Binary Black Hole Mergers from Globular Clusters: Implications for Advanced LIGO. Phys. Rev. Lett., 115(5):051101, July 2015. doi:10.1103/PhysRevLett.115.051101.

R Abbott, TD Abbott, S Abraham, F Acernese, K Ackley, C Adams, RX Adhikari, VB Adya, C Affeldt, M Agathos, et al. Gw190814: Gravitational waves from the coalescence of a 23 solar mass black hole with a 2.6 solar mass compact object. *The Astrophysical Journal Letters*, 896(2):L44, 2020a.

BP Abbott, R Abbott, TD Abbott, S Abraham, F Acernese, K Ackley, C Adams, RX Adhikari, VB Adya, C Affeldt, et al. Gw190425: Observation of a compact binary coalescence with total mass 3.4 m. *The Astrophysical Journal Letters*, 892(1):L3, 2020b.

R Abbott, TD Abbott, S Abraham, F Acernese, K Ackley, C Adams, RX Adhikari, VB Adya, C Affeldt, M Agathos, et al. Gw190412: Observation of a binary-black-hole coalescence with asymmetric masses. *arXiv preprint arXiv:2004.08342*, 2020c.

Benjamin P Abbott, Richard Abbott, TD Abbott, F Acernese, K Ackley, C Adams, T Adams, P Addesso, Rana X Adhikari, VB Adya, et al. Gw170814: a three-detector observation of gravitational waves from a binary black hole coalescence. *Physical review letters*, 119(14):141101, 2017.

Marc Freitag, M. Atakan Gürkan, and Frederic A. Rasio. Runaway collisions in young star clusters - II. Numerical results. MNRAS, 368(1):141–161, May 2006. doi:10.1111/j.1365-2966.2006.10096.x.

M. Atakan Gürkan, John M. Fregeau, and Frederic A. Rasio. Massive Black Hole Binaries from Collisional Runaways. ApJ, 640(1):L39–L42, March 2006. doi:10.1086/503295.

Mirek Giersz, Nathan Leigh, Arkadiusz Hypki, Nora Lützgendorf, and Abbas Askar. MOCCA code for star cluster simulations - IV. A new scenario for intermediate mass black hole formation in globular clusters. MNRAS, 454(3):3150–3165, December 2015. doi:10.1093/mnras/stv2162.

Toshikazu Ebisuzaki, Junichiro Makino, Takeshi Go Tsuru, Yoko Funato, Simon Portegies Zwart, Piet Hut, Steve McMillan, Satoki Matsushita, Hironori Matsumoto, and Ryohei Kawabe. Missing link found? the "runaway" path to supermassive black holes. *The Astrophysical Journal Letters*, 562(1):L19, 2001.

M. H. Hénon. The Monte Carlo Method (Papers appear in the Proceedings of IAU Colloquium No. 10 Gravitational N-Body Problem (ed. by Myron Lecar), R. Reidel Publ. Co. , Dordrecht-Holland.). Ap&SS, 14(1):151–167, November 1971b. doi:10.1007/BF00649201.

M Giersz and R Spurzem. A stochastic monte carlo approach to model real star cluster evolution—ii. self-consistent models and primordial binaries. *Monthly Notices of the Royal Astronomical Society*, 317(3):581–606, 2000.

Carl L Rodriguez, Meagan Morscher, Long Wang, Sourav Chatterjee, Frederic A Rasio, and Rainer Spurzem. Million-body star cluster simulations: comparisons between monte carlo and direct n-body. *Monthly Notices of the Royal Astronomical Society*, 463(2):2109–2118, 2016.

Arkadiusz Hypki and Mirek Giersz. mocca code for star cluster simulations–v. initial globular cluster conditions influence on blue stragglers. *Monthly Notices of the Royal Astronomical Society*, 466(1):320–339, 2017.

Eugene Vasiliev. A new fokker–planck approach for the relaxation-driven evolution of galactic nuclei. *The Astrophysical Journal*, 848(1):10, 2017.

S Chandrasekhar. Principles of stellar dynamics, univ, 1942.

Yohai Meiron and Bence Kocsis. Resonant Relaxation in Globular Clusters. ApJ, 878(2): 138, June 2019. doi:10.3847/1538-4357/ab1b32.

Jun Yan Lau and James Binney. Relaxation of spherical stellar systems. MNRAS, 490(1): 478–490, November 2019. doi:10.1093/mnras/stz2567.

R. H. Miller. Irreversibility in Small Stellar Dynamical Systems. ApJ, 140:250, July 1964. doi:10.1086/147911.

Long Wang, Rainer Spurzem, Sverre Aarseth, Mirek Giersz, Abbas Askar, Peter Berczik, Thorsten Naab, Riko Schadow, and MBN Kouwenhoven. The dragon simulations: globular cluster evolution with a million stars. *Monthly Notices of the Royal Astronomical Society*, 458(2):1450–1465, 2016.

Stephen L. W. McMillan and Sverre J. Aarseth. An O(N N) Integration Scheme for Collisional Stellar Systems. ApJ, 414:200, September 1993. doi:10.1086/173068.

Masaki Iwasawa, Simon Portegies Zwart, and Junichiro Makino. Gpu-enabled particle-particle particle-tree scheme for simulating dense stellar cluster system. *Computational Astrophysics and Cosmology*, 2(1):6, 2015.

Qirong Zhu. A momentum-conserving n-body scheme with individual time steps. *New Astronomy*, 85:101481, 2021.

Jonathan P. Coles and Rebekka Bieri. An optimizing symbolic algebra approach for generating fast multipole method operators. *Computer Physics Communications*, 251:107081, June 2020. doi:10.1016/j.cpc.2019.107081.

Pierre Fortin and Maxime Touche. Dual tree traversal on integrated gpus for astrophysical n-body simulations. *The International Journal of High Performance Computing Applications*, 33(5):960–972, 2019. doi:10.1177/1094342019840806. URL `https://doi.org/10.1177/1094342019840806`.

Henry Crozier Plummer. On the problem of distribution in globular star clusters. *Monthly notices of the royal astronomical society*, 71:460–470, 1911.

Andreas HW Küpper, Thomas Maschberger, Pavel Kroupa, and Holger Baumgardt. Mass segregation and fractal substructure in young massive clusters–i. the mcluster code and method calibration. *Monthly Notices of the Royal Astronomical Society*, 417(3):2300–2317, 2011.

Long Wang, Rainer Spurzem, Sverre Aarseth, Keigo Nitadori, Peter Berczik, MBN Kouwenhoven, and Thorsten Naab. Nbody6++ gpu: ready for the gravitational million-body problem. *Monthly Notices of the Royal Astronomical Society*, 450(4):4070–4080, 2015.

DC Heggie and RD Mathieu. Standardised units and time scales. In *The use of supercomputers in stellar dynamics*, pages 233–235. Springer, 1986.

Mirek Giersz and Douglas C Heggie. Statistics of n-body simulations–i. equal masses before core collapse. *Monthly Notices of the Royal Astronomical Society*, 268(1):257–275, 1994.

Carl L Rodriguez, Bharath Pattabiraman, Sourav Chatterjee, Alok Choudhary, Wei-keng Liao, Meagan Morscher, and Frederic A Rasio. A new hybrid technique for modeling dense star clusters. *Computational Astrophysics and Cosmology*, 5(1):5, 2018.

Simon Portegies Zwart, Steve McMillan, Stefan Harfst, Derek Groen, Michiko Fujii, Breanndán Ó Nualláin, Evert Glebbeek, Douglas Heggie, James Lombardi, Piet Hut, et al. A multiphysics and multiscale software environment for modeling astrophysical systems. *New Astronomy*, 14(4):369–378, 2009.

F Inti Pelupessy, A van Elteren, N de Vries, SLW McMillan, N Drost, and SF Portegies Zwart. The astrophysical multipurpose software environment. *Astronomy & Astrophysics*, 557:A84, 2013.

S Portegies Zwart and Steve McMillan. *Astrophysical Recipes*. IOP, 2018.

S Casertano and P Hut. Core radius and density measurements in n-body experiments connections with theoretical and observational definitions. *The Astrophysical Journal*, 298:80–94, 1985.

Daniel J Eisenstein and Piet Hut. Hop: a new group-finding algorithm for n-body simulations. *The Astrophysical Journal*, 498(1):137, 1998.

Tobias Goerdt, Ben Moore, JI Read, Joachim Stadel, and Marcel Zemp. Does the fornax dwarf spheroidal have a central cusp or core? *Monthly Notices of the Royal Astronomical Society*, 368(3):1073–1077, 2006.

Shigeki Inoue. The test for suppressed dynamical friction in a constant density core of dwarf galaxies. *Monthly Notices of the Royal Astronomical Society*, 397(2):709–716, 2009.

Tobias Goerdt, Ben Moore, JI Read, and Joachim Stadel. Core creation in galaxies and halos via sinking massive objects. *The Astrophysical Journal*, 725(2):1707, 2010.

James A Petts, Justin I Read, and Alessia Gualandris. A semi-analytic dynamical friction model for cored galaxies. *Monthly Notices of the Royal Astronomical Society*, 463(1): 858–869, 2016.

JM Silva, JAS Lima, RE De Souza, Antonino Del Popolo, Morgan Le Delliou, and Xi-Guo Lee. Chandrasekhar's dynamical friction and non-extensive statistics. *Journal of Cosmology and Astroparticle Physics*, 2016(05):021, 2016.

Oreste Villa, Vidhya Gurumoorthi, Andrés Márquez, and Sriram Krishnamoorthy. Effects of floating-point non-associativity on numerical computations on massively multithreaded systems. In *Proceedings of Cray User Group Meeting (CUG)*, page 3, 2009.

Long Wang, Keigo Nitadori, and Junichiro Makino. Sdar: Slow-down algorithmic regularization code for solving few-body problems. *ascl*, pages ascl–2002, 2020.

Michael Y Grudić and Philip F Hopkins. A general-purpose time-step criterion for simulations with gravity. *Monthly Notices of the Royal Astronomical Society*, 495(4):4306–4313, 2020.

Cristián Maureira-Fredes and Pau Amaro-Seoane. GRAVIDY, a GPU modular, parallel direct-summation N-body integrator: dynamics with softening. MNRAS, 473(3):3113–3127, January 2018. doi:10.1093/mnras/stx2468.

Keigo Nitadori and Junichiro Makino. Sixth- and eighth-order Hermite integrator for N-body simulations. New A, 13(7):498–507, October 2008. doi:10.1016/j.newast.2008.01.010.

Takayuki R. Saitoh and Junichiro Makino. A Necessary Condition for Individual Time Steps in SPH Simulations. ApJ, 697(2):L99–L102, June 2009. doi:10.1088/0004-637X/697/2/L99.

U. Löckmann and H. Baumgardt. Tracing intermediate-mass black holes in the Galactic Centre. MNRAS, 384(1):323–330, February 2008. doi:10.1111/j.1365-2966.2007.12699.x.

John Kormendy and Karl Gebhardt. Supermassive black holes in galactic nuclei. In J. Craig Wheeler and Hugo Martel, editors, *20th Texas Symposium on relativistic astrophysics*, volume 586 of *American Institute of Physics Conference Series*, pages 363–381, October 2001. doi:10.1063/1.1419581.

John Kormendy and Luis C. Ho. Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies. ARA&A, 51(1):511–653, August 2013a. doi:10.1146/annurev-astro-082708-101811.

Vicente Rodriguez-Gomez, Annalisa Pillepich, Laura V. Sales, Shy Genel, Mark Vogelsberger, Qirong Zhu, Sarah Wellons, Dylan Nelson, Paul Torrey, Volker Springel, Chung-Pei Ma, and Lars Hernquist. The stellar mass assembly of galaxies in the Illustris simulation: growth by mergers and the spatial distribution of accreted stars. MNRAS, 458 (3):2371–2390, May 2016. doi:10.1093/mnras/stw456.

S. Komossa, V. Burwitz, G. Hasinger, P. Predehl, J. S. Kaastra, and Y. Ikebe. Discovery of a Binary Active Galactic Nucleus in the Ultraluminous Infrared Galaxy NGC 6240 Using Chandra. ApJ, 582(1):L15–L19, January 2003. doi:10.1086/346145.

S. Komossa. Observational evidence for binary black holes and active double nuclei. Mem. Soc. Astron. Italiana, 77:733, January 2006.

Tamara Bogdanović. Supermassive Black Hole Binaries: The Search Continues. In *Gravitational Wave Astrophysics*, volume 40 of *Astrophysics and Space Science Proceedings*, page 103, January 2015. doi:10.1007/978-3-319-10488-1_9.

Pau Amaro-Seoane, Heather Audley, Stanislav Babak, John Baker, Enrico Barausse, Peter Bender, Emanuele Berti, Pierre Binetruy, Michael Born, Daniele Bortoluzzi, Jordan Camp, Chiara Caprini, Vitor Cardoso, Monica Colpi, John Conklin, Neil Cornish, Curt Cutler, Karsten Danzmann, Rita Dolesi, Luigi Ferraioli, Valerio Ferroni, Ewan Fitzsimons, Jonathan Gair, Lluis Gesa Bote, Domenico Giardini, Ferran Gibert, Catia Grimani, Hubert Halloin, Gerhard Heinzel, Thomas Hertog, Martin Hewitson, Kelly

Holley-Bockelmann, Daniel Hollington, Mauro Hueller, Henri Inchauspe, Philippe Jetzer, Nikos Karnesis, Christian Killow, Antoine Klein, Bill Klipstein, Natalia Korsakova, Shane L Larson, Jeffrey Livas, Ivan Lloro, Nary Man, Davor Mance, Joseph Martino, Ignacio Mateos, Kirk McKenzie, Sean T McWilliams, Cole Miller, Guido Mueller, Germano Nardini, Gijs Nelemans, Miquel Nofrarias, Antoine Petiteau, Paolo Pivato, Eric Plagnol, Ed Porter, Jens Reiche, David Robertson, Norna Robertson, Elena Rossi, Giuliana Russano, Bernard Schutz, Alberto Sesana, David Shoemaker, Jacob Slutsky, Carlos F. Sopuerta, Tim Sumner, Nicola Tamanini, Ira Thorpe, Michael Troebs, Michele Vallisneri, Alberto Vecchio, Daniele Vetrugno, Stefano Vitale, Marta Volonteri, Gudrun Wanner, Harry Ward, Peter Wass, William Weber, John Ziemer, and Peter Zweifel. Laser Interferometer Space Antenna. *arXiv e-prints*, art. arXiv:1702.00786, February 2017.

Jun Luo, Li-Sheng Chen, Hui-Zong Duan, Yun-Gui Gong, Shoucun Hu, Jianghui Ji, Qi Liu, Jianwei Mei, Vadim Milyukov, Mikhail Sazhin, Cheng-Gang Shao, Viktor T. Toth, Hai-Bo Tu, Yamin Wang, Yan Wang, Hsien-Chi Yeh, Ming-Sheng Zhan, Yonghe Zhang, Vladimir Zharov, and Ze-Bing Zhou. TianQin: a space-borne gravitational wave detector. *Classical and Quantum Gravity*, 33(3):035010, February 2016. doi:10.1088/0264-9381/33/3/035010.

David Merritt. *Dynamics and Evolution of Galactic Nuclei*. 2013b.

Nadine Neumayer, Anil Seth, and Torsten Böker. Nuclear star clusters. A&A Rev., 28(1): 4, July 2020b. doi:10.1007/s00159-020-00125-0.

Rubén Sánchez-Janssen, Patrick Côté, Laura Ferrarese, Eric W. Peng, Joel Roediger, John P. Blakeslee, Eric Emsellem, Thomas H. Puzia, Chelsea Spengler, James Taylor, Karla A. Álamo-Martínez, Alessandro Boselli, Michele Cantiello, Jean-Charles Cuillandre, Pierre-Alain Duc, Patrick Durrell, Stephen Gwyn, Lauren A. MacArthur, Ariane Lançon, Sungsoon Lim, Chengze Liu, Simona Mei, Bryan Miller, Roberto Muñoz, J. Christopher Mihos, Sanjaya Paudel, Mathieu Powalka, and Elisa Toloba. The Next Generation Virgo Cluster Survey. XXIII. Fundamentals of Nuclear Star Clusters over Seven Decades in Galaxy Mass. ApJ, 878(1):18, June 2019. doi:10.3847/1538-4357/aaf4fd.

A. M. Ghez, S. Salim, N. N. Weinberg, J. R. Lu, T. Do, J. K. Dunn, K. Matthews, M. R. Morris, S. Yelda, E. E. Becklin, T. Kremenek, M. Milosavljevic, and J. Naiman. Measuring Distance and Properties of the Milky Way's Central Supermassive Black Hole with Stellar Orbits. ApJ, 689(2):1044–1062, December 2008. doi:10.1086/592738.

J. N. Bahcall and R. A. Wolf. Star distribution around a massive black hole in a globular cluster. ApJ, 209:214–232, October 1976. doi:10.1086/154711.

J. N. Bahcall and R. A. Wolf. The star distribution around a massive black hole in a globular cluster. II. Unequal star masses. ApJ, 216:883–907, September 1977. doi:10.1086/155534.

Tal Alexander and Clovis Hopman. Strong Mass Segregation Around a Massive Black Hole. ApJ, 697(2):1861–1869, June 2009. doi:10.1088/0004-637X/697/2/1861.

Walter Dehnen. Phase-space mixing and the merging of cusps. MNRAS, 360(3):892–900, July 2005. doi:10.1111/j.1365-2966.2005.09099.x.

Miguel Preto and Pau Amaro-Seoane. On Strong Mass Segregation Around a Massive Black Hole: Implications for Lower-Frequency Gravitational-Wave Astrophysics. ApJ, 708(1): L42–L46, January 2010. doi:10.1088/2041-8205/708/1/L42.

Qirong Zhu. A momentum-conserving N-body scheme with individual time steps. New A, 85:101481, May 2021a. doi:10.1016/j.newast.2020.101481.

Diptajyoti Mukherjee, Qirong Zhu, Hy Trac, and Carl L. Rodriguez. Fast Multipole Methods for N-body Simulations of Collisional Star Systems. ApJ, 916(1):9, July 2021b. doi:10.3847/1538-4357/ac03b2.

Seppo Mikkola and Kiyotaka Tanikawa. Algorithmic regularization of the few-body problem. MNRAS, 310(3):745–749, December 1999b. doi:10.1046/j.1365-8711.1999.02982.x.

Junichiro Makino and Sverre J. Aarseth. On a Hermite Integrator with Ahmad-Cohen Scheme for Gravitational Many-Body Problems. PASJ, 44:141–151, April 1992.

Federico I. Pelupessy, Jürgen Jänes, and Simon Portegies Zwart. N-body integrators with individual time steps from Hierarchical splitting. New A, 17(8):711–719, November 2012b. doi:10.1016/j.newast.2012.05.009.

Antti Rantala, Thorsten Naab, and Volker Springel. frost: a momentum-conserving CUDA implementation of a hierarchical fourth-order forward symplectic integrator. MNRAS, 502(4):5546–5562, April 2021b. doi:10.1093/mnras/stab057.

Siu A. Chin. Symplectic integrators from composite operator factorizations. *Physics Letters A*, 226(6):344–348, February 1997. doi:10.1016/S0375-9601(97)00003-0.

Siu A. Chin. Physics of symplectic integrators: Perihelion advances and symplectic corrector algorithms. Phys. Rev. E, 75(3):036701, March 2007. doi:10.1103/PhysRevE.75.036701.

Junichiro Makino, Piet Hut, Murat Kaplan, and Hasan Saygın. A time-symmetric block time-step algorithm for N-body simulations. New A, 12(2):124–133, November 2006b. doi:10.1016/j.newast.2006.06.003.

Yi-Han Wang, Nathan W. C. Leigh, Bin Liu, and Rosalba Perna. SpaceHub: A high-performance gravity integration toolkit for few-body problems in astrophysics. MNRAS, 505(1):1053–1070, July 2021b. doi:10.1093/mnras/stab1189.

O. Coulaud, P. Fortin, and J. Roman. High performance BLAS formulation of the multipole-to-local operator in the fast multipole method. *Journal of Computational Physics*, 227 (3):1836–1862, January 2008. doi:10.1016/j.jcp.2007.09.027.

Long Wang, Rainer Spurzem, Sverre Aarseth, Keigo Nitadori, Peter Berczik, M. B. N. Kouwenhoven, and Thorsten Naab. NBODY6++GPU: ready for the gravitational million-body problem. MNRAS, 450(4):4070–4080, July 2015. doi:10.1093/mnras/stv817.

Fabio Antonini. On the Distribution of Stellar Remnants around Massive Black Holes: Slow Mass Segregation, Star Cluster Inspirals, and Correlated Orbits. ApJ, 794(2):106, October 2014. doi:10.1088/0004-637X/794/2/106.

Pavel Kroupa. On the variation of the initial mass function. MNRAS, 322(2):231–246, April 2001. doi:10.1046/j.1365-8711.2001.04022.x.

H. Maness, F. Martins, S. Trippe, R. Genzel, J. R. Graham, C. Sheehy, M. Salaris, S. Gillessen, T. Alexander, T. Paumard, T. Ott, R. Abuter, and F. Eisenhauer. Evidence for a Long-standing Top-heavy Initial Mass Function in the Central Parsec of the Galaxy. ApJ, 669(2):1024–1041, November 2007. doi:10.1086/521669.

H. Bartko, F. Martins, S. Trippe, T. K. Fritz, R. Genzel, T. Ott, F. Eisenhauer, S. Gillessen, T. Paumard, T. Alexander, K. Dodds-Eden, O. Gerhard, Y. Levin, L. Mascetti, S. Nayakshin, H. B. Perets, G. Perrin, O. Pfuhl, M. J. Reid, D. Rouan, M. Zilka, and A. Sternberg. An Extremely Top-Heavy Initial Mass Function in the Galactic Center Stellar Disks. ApJ, 708(1):834–840, January 2010. doi:10.1088/0004-637X/708/1/834.

Gilles Chabrier. The Initial Mass Function: From Salpeter 1955 to 2005. In E. Corbelli, F. Palla, and H. Zinnecker, editors, *The Initial Mass Function 50 Years Later*, volume 327 of *Astrophysics and Space Science Library*, page 41, January 2005. doi:10.1007/978-1-4020-3407-7_5.

Eugene Vasiliev. AGAMA: action-based galaxy modelling architecture. MNRAS, 482(2): 1525–1544, January 2019a. doi:10.1093/mnras/sty2672.

W. Dehnen. A Family of Potential-Density Pairs for Spherical Galaxies and Bulges. MNRAS, 265:250, November 1993. doi:10.1093/mnras/265.1.250.

M. Baes, H. Dejonghe, and P. Buyle. The dynamical structure of isotropic spherical galaxies with a central black hole. A&A, 432(2):411–422, March 2005. doi:10.1051/0004-6361:20041907.

Clovis Hopman and Tal Alexander. The Effect of Mass Segregation on Gravitational Wave Sources near Massive Black Holes. ApJ, 645(2):L133–L136, July 2006. doi:10.1086/506273.

David Merritt and Andras Szell. Dynamical Cusp Regeneration. ApJ, 648(2):890–899, September 2006a. doi:10.1086/506010.

Su-Shu Huang. Modes of Mass Ejection by Binary Stars and the Effect on Their Orbital Periods. ApJ, 138:471, August 1963. doi:10.1086/147659.

Fazeel Mahmood Khan, Kelly Holley-Bockelmann, and Peter Berczik. Ultramassive Black Hole Coalescence. ApJ, 798(2):103, January 2015. doi:10.1088/0004-637X/798/2/103.

Alessia Gualandris, Fazeel Mahmood Khan, Elisa Bortolas, Matteo Bonetti, Alberto Sesana, Peter Berczik, and Kelly Holley-Bockelmann. Eccentricity evolution of massive black hole binaries from formation to coalescence. MNRAS, 511(4):4753–4765, April 2022. doi:10.1093/mnras/stac241.

Alberto Sesana, Francesco Haardt, and Piero Madau. Interaction of Massive Black Hole Binaries with Their Stellar Environment. III. Scattering of Bound Stars. ApJ, 686(1): 432–447, October 2008. doi:10.1086/590651.

Manuel Arca-Sedda and Alessia Gualandris. Gravitational wave sources from inspiralling globular clusters in the Galactic Centre and similar environments. MNRAS, 477(4): 4423–4442, July 2018. doi:10.1093/mnras/sty922.

Alberto Sesana. Self Consistent Model for the Evolution of Eccentric Massive Black Hole Binaries in Stellar Environments: Implications for Gravitational Wave Observations. ApJ, 719(1):851–864, August 2010. doi:10.1088/0004-637X/719/1/851.

Antti Rantala, Thorsten Naab, Francesco Paolo Rizzuto, Matias Mannerkoski, Christian Partmann, and Kristina Lautenschütz. BIFROST: simulating compact subsystems in star clusters using a hierarchical fourth-order forward symplectic integrator code. *arXiv e-prints*, art. arXiv:2210.02472, October 2022.

David Merritt. The Distribution of Stars and Stellar Remnants at the Galactic Center. ApJ, 718(2):739–761, August 2010. doi:10.1088/0004-637X/718/2/739.

Eugene Vasiliev, Fabio Antonini, and David Merritt. The Final-parsec Problem in Nonspherical Galaxies Revisited. ApJ, 785(2):163, April 2014. doi:10.1088/0004-637X/785/2/163.

Makoto Miyoshi, James Moran, James Herrnstein, Lincoln Greenhill, Naomasa Nakai, Philip Diamond, and Makoto Inoue. Evidence for a black hole from high rotation velocities in a sub-parsec region of NGC4258. Nature, 373(6510):127–129, January 1995. doi:10.1038/373127a0.

Scott Tremaine, Karl Gebhardt, Ralf Bender, Gary Bower, Alan Dressler, S. M. Faber, Alexei V. Filippenko, Richard Green, Carl Grillmair, Luis C. Ho, John Kormendy, Tod R. Lauer, John Magorrian, Jason Pinkney, and Douglas Richstone. The Slope of the Black Hole Mass versus Velocity Dispersion Correlation. ApJ, 574(2):740–753, August 2002. doi:10.1086/341002.

John Kormendy and Luis C. Ho. Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies. ARA&A, 51(1):511–653, August 2013b. doi:10.1146/annurev-astro-082708-101811.

Amy E. Reines, Jenny E. Greene, and Marla Geha. Dwarf Galaxies with Optical Signatures of Active Massive Black Holes. ApJ, 775(2):116, October 2013. doi:10.1088/0004-637X/775/2/116.

Jenny E. Greene, Jay Strader, and Luis C. Ho. Intermediate-Mass Black Holes. ARA&A, 58:257–312, August 2020. doi:10.1146/annurev-astro-032620-021835.

Rebecca L. Larson, Steven L. Finkelstein, Dale D. Kocevski, Taylor A. Hutchison, Jonathan R. Trump, Pablo Arrabal Haro, Volker Bromm, Nikko J. Cleri, Mark Dickinson, Seiji Fujimoto, Jeyhan S. Kartaltepe, Anton M. Koekemoer, Casey Papovich, Nor Pirzkal, Sandro Tacchella, Jorge A. Zavala, Micaela Bagley, Peter Behroozi, Jaclyn B. Champagne, Justin W. Cole, Intae Jung, Alexa M. Morales, Guang Yang, Haowen Zhang, Adi Zitrin, Ricardo O. Amorín, Denis Burgarella, Caitlin M. Casey, Óscar A. Chávez Ortiz, Isabella G. Cox, Katherine Chworowsky, Adriano Fontana, Eric Gawiser, Andrea Grazian, Norman A. Grogin, Santosh Harish, Nimish P. Hathi, Michaela Hirschmann, Benne W. Holwerda, Stéphanie Juneau, Gene C. K. Leung, Ray A. Lucas, Elizabeth J. McGrath, Pablo G. Pérez-González, Jane R. Rigby, Lise-Marie Seillé, Raymond C. Simons, Alexander de La Vega, Benjamin J. Weiner, Stephen M. Wilkins, L. Y. Aaron Yung, and Ceers Team. A CEERS Discovery of an Accreting Supermassive Black Hole 570 Myr after the Big Bang: Identifying a Progenitor of Massive z ¿ 6 Quasars. ApJ, 953(2):L29, August 2023. doi:10.3847/2041-8213/ace619.

Hannah Übler, Roberto Maiolino, Emma Curtis-Lake, Pablo G. Pérez-González, Mirko Curti, Michele Perna, Santiago Arribas, Stéphane Charlot, Madeline A. Marshall, Francesco D'Eugenio, Jan Scholtz, Andrew Bunker, Stefano Carniani, Pierre Ferruit, Peter Jakobsen, Hans-Walter Rix, Bruno Rodríguez Del Pino, Chris J. Willott, Torsten Boeker, Giovanni Cresci, Gareth C. Jones, Nimisha Kumari, and Tim Rawle. GA-NIFS: A massive black hole in a low-metallicity AGN at z $\sim$ 5.55 revealed by JWST/NIRSpec IFS. A&A, 677:A145, September 2023. doi:10.1051/0004-6361/202346137.

Dale D. Kocevski, Masafusa Onoue, Kohei Inayoshi, Jonathan R. Trump, Pablo Arrabal Haro, Andrea Grazian, Mark Dickinson, Steven L. Finkelstein, Jeyhan S. Kartaltepe, Michaela Hirschmann, James Aird, Benne W. Holwerda, Seiji Fujimoto, Stéphanie Juneau, Ricardo O. Amorín, Bren E. Backhaus, Micaela B. Bagley, Guillermo Barro, Eric F. Bell, Laura Bisigello, Antonello Calabrò, Nikko J. Cleri, M. C. Cooper, Xuheng Ding, Norman A. Grogin, Luis C. Ho, Taylor A. Hutchison, Akio K. Inoue, Linhua Jiang, Brenda Jones, Anton M. Koekemoer, Wenxiu Li, Zhengrong Li, Elizabeth J. McGrath, Juan Molina, Casey Papovich, Pablo G. Pérez-González, Nor Pirzkal, Stephen M. Wilkins, Guang Yang, and L. Y. Aaron Yung. Hidden Little Monsters: Spectroscopic Identification of Low-mass, Broad-line AGNs at z ¿ 5 with CEERS. ApJ, 954(1):L4, September 2023. doi:10.3847/2041-8213/ace5a0.

Yuichi Harikane, Yechi Zhang, Kimihiko Nakajima, Masami Ouchi, Yuki Isobe, Yoshiaki Ono, Shun Hatano, Yi Xu, and Hiroya Umeda. A JWST/NIRSpec First Census of Broad-line AGNs at z = 4-7: Detection of 10 Faint AGNs with M $_{BH}$ 10^6-10^8 M $_\odot$ and Their Host Galaxy Properties. ApJ, 959(1):39, December 2023. doi:10.3847/1538-4357/ad029e.

Andy D. Goulding, Jenny E. Greene, David J. Setton, Ivo Labbe, Rachel Bezanson, Tim B. Miller, Hakim Atek, Ákos Bogdán, Gabriel Brammer, Iryna Chemerynska, Sam E. Cutler, Pratika Dayal, Yoshinobu Fudamoto, Seiji Fujimoto, Lukas J. Furtak, Vasily Kokorev,

Gourav Khullar, Joel Leja, Danilo Marchesini, Priyamvada Natarajan, Erica Nelson, Pascal A. Oesch, Richard Pan, Casey Papovich, Sedona H. Price, Pieter van Dokkum, Bingjie Wang, John R. Weaver, Katherine E. Whitaker, and Adi Zitrin. UNCOVER: The Growth of the First Massive Black Holes from JWST/NIRSpec-Spectroscopic Redshift Confirmation of an X-Ray Luminous AGN at z = 10.1. ApJ, 955(1):L24, September 2023. doi:10.3847/2041-8213/acf7c5.

Roberto Maiolino, Jan Scholtz, Emma Curtis-Lake, Stefano Carniani, William Baker, Anna de Graaff, Sandro Tacchella, Hannah Übler, Francesco D'Eugenio, Joris Witstok, Mirko Curti, Santiago Arribas, Andrew J. Bunker, Stéphane Charlot, Jacopo Chevallard, Daniel J. Eisenstein, Eiichi Egami, Zhiyuan Ji, Gareth C. Jones, Jianwei Lyu, Tim Rawle, Brant Robertson, Wiphu Rujopakarn, Michele Perna, Fengwu Sun, Giacomo Venturi, Christina C. Williams, and Chris Willott. JADES. The diverse population of infant Black Holes at 4¡z¡11: merging, tiny, poor, but mighty. *arXiv e-prints*, art. arXiv:2308.01230, August 2023. doi:10.48550/arXiv.2308.01230.

Roberto Maiolino, Jan Scholtz, Joris Witstok, Stefano Carniani, Francesco D'Eugenio, Anna de Graaff, Hannah Übler, Sandro Tacchella, Emma Curtis-Lake, Santiago Arribas, Andrew Bunker, Stéphane Charlot, Jacopo Chevallard, Mirko Curti, Tobias J. Looser, Michael V. Maseda, Timothy D. Rawle, Bruno Rodríguez del Pino, Chris J. Willott, Eiichi Egami, Daniel J. Eisenstein, Kevin N. Hainline, Brant Robertson, Christina C. Williams, Christopher N. A. Willmer, William M. Baker, Kristan Boyett, Christa DeCoursey, Andrew C. Fabian, Jakob M. Helton, Zhiyuan Ji, Gareth C. Jones, Nimisha Kumari, Nicolas Laporte, Erica J. Nelson, Michele Perna, Lester Sandles, Irene Shivaei, and Fengwu Sun. A small and vigorous black hole in the early Universe. Nature, 627 (8002):59–63, March 2024. doi:10.1038/s41586-024-07052-5.

Jorryt Matthee, Rohan P. Naidu, Gabriel Brammer, John Chisholm, Anna-Christina Eilers, Andy Goulding, Jenny Greene, Daichi Kashino, Ivo Labbe, Simon J. Lilly, Ruari Mackenzie, Pascal A. Oesch, Andrea Weibel, Stijn Wuyts, Mengyuan Xiao, Rongmon Bordoloi, Rychard Bouwens, Pieter van Dokkum, Garth Illingworth, Ivan Kramarenko, Michael V. Maseda, Charlotte Mason, Romain A. Meyer, Erica J. Nelson, Naveen A. Reddy, Irene Shivaei, Robert A. Simcoe, and Minghao Yue. Little Red Dots: An Abundant Population of Faint Active Galactic Nuclei at z $\sim$ 5 Revealed by the EIGER and FRESCO JWST Surveys. ApJ, 963(2):129, March 2024. doi:10.3847/1538-4357/ad2345.

Fabio Pacucci, Bao Nguyen, Stefano Carniani, Roberto Maiolino, and Xiaohui Fan. JWST CEERS and JADES Active Galaxies at z = 4-7 Violate the Local M -M $_\star$ Relation at ¿3σ: Implications for Low-mass Black Holes and Seeding Models. ApJ, 957(1):L3, November 2023. doi:10.3847/2041-8213/ad0158.

R. Barkana and A. Loeb. In the beginning: the first sources of light and the reionization of the universe. Phys. Rep., 349(2):125–238, July 2001. doi:10.1016/S0370-1573(01)00019-9.

Pratika Dayal, Elena M. Rossi, Banafsheh Shiralilou, Olmo Piana, Tirthankar Roy Choudhury, and Marta Volonteri. The hierarchical assembly of galaxies and black holes in the

first billion years: predictions for the era of gravitational wave astronomy. MNRAS, 486 (2):2336–2350, June 2019. doi:10.1093/mnras/stz897.

Fabio Pacucci and Abraham Loeb. Separating Accretion and Mergers in the Cosmic Growth of Black Holes with X-Ray and Gravitational-wave Observations. ApJ, 895(2):95, June 2020. doi:10.3847/1538-4357/ab886e.

Olmo Piana, Pratika Dayal, Marta Volonteri, and Tirthankar Roy Choudhury. The mass assembly of high-redshift black holes. MNRAS, 500(2):2146–2158, January 2021. doi:10.1093/mnras/staa3363.

Aklant K. Bhowmick, Laura Blecha, Paul Torrey, Luke Zoltan Kelley, Rainer Weinberger, Mark Vogelsberger, Lars Hernquist, Rachel S. Somerville, and Analis Eolyn Evans. Introducing the BRAHMA simulation suite: Signatures of low mass black hole seeding models in cosmological simulations. *arXiv e-prints*, art. arXiv:2402.03626, February 2024. doi:10.48550/arXiv.2402.03626.

Mitchell C. Begelman, Marta Volonteri, and Martin J. Rees. Formation of supermassive black holes by direct collapse in pre-galactic haloes. MNRAS, 370(1):289–298, July 2006. doi:10.1111/j.1365-2966.2006.10467.x.

L. Mayer, S. Kazantzidis, A. Escala, and S. Callegari. Direct formation of supermassive black holes via multi-scale gas inflows in galaxy mergers. Nature, 466(7310):1082–1084, August 2010. doi:10.1038/nature09294.

S. F. Portegies Zwart, J. Makino, S. L. W. McMillan, and P. Hut. Star cluster ecology. III. Runaway collisions in young compact star clusters. A&A, 348:117–126, August 1999. doi:10.48550/arXiv.astro-ph/9812006.

B. Devecchi and M. Volonteri. Formation of the First Nuclear Clusters and Massive Black Holes at High Redshift. ApJ, 694(1):302–313, March 2009. doi:10.1088/0004-637X/694/1/302.

A. Lupi, M. Colpi, B. Devecchi, G. Galanti, and M. Volonteri. Constraining the high-redshift formation of black hole seeds in nuclear star clusters with gas inflows. MNRAS, 442(4):3616–3626, August 2014. doi:10.1093/mnras/stu1120.

Arpan Das, Dominik R. G. Schleicher, Nathan W. C. Leigh, and Tjarda C. N. Boekholt. Formation of supermassive black hole seeds in nuclear star clusters via gas accretion and runaway collisions. MNRAS, 503(1):1051–1069, May 2021. doi:10.1093/mnras/stab402.

Abbas Askar, Vivienne F. Baldassare, and Mar Mezcua. Intermediate-Mass Black Holes in Star Clusters and Dwarf Galaxies. *arXiv e-prints*, art. arXiv:2311.12118, November 2023. doi:10.48550/arXiv.2311.12118.

Ugo N. Di Carlo, Nicola Giacobbo, Michela Mapelli, Mario Pasquato, Mario Spera, Long Wang, and Francesco Haardt. Merging black holes in young star clusters. MNRAS, 487 (2):2947–2960, August 2019. doi:10.1093/mnras/stz1453.

Ugo N. Di Carlo, Michela Mapelli, Mario Pasquato, Sara Rastello, Alessandro Ballone, Marco Dall'Amico, Nicola Giacobbo, Giuliano Iorio, Mario Spera, Stefano Torniamenti, and Francesco Haardt. Intermediate-mass black holes from stellar mergers in young star clusters. MNRAS, 507(4):5132–5143, November 2021. doi:10.1093/mnras/stab2390.

Fabio Pacucci, Priyamvada Natarajan, Marta Volonteri, Nico Cappelluti, and C. Megan Urry. Conditions for Optimal Growth of Black Hole Seeds. ApJ, 850(2):L42, December 2017. doi:10.3847/2041-8213/aa9aea.

Chiara M. F. Mingarelli, T. Joseph W. Lazio, Alberto Sesana, Jenny E. Greene, Justin A. Ellis, Chung-Pei Ma, Steve Croft, Sarah Burke-Spolaor, and Stephen R. Taylor. The local nanohertz gravitational-wave landscape from supermassive black hole binaries. *Nature Astronomy*, 1:886–892, November 2017. doi:10.1038/s41550-017-0299-6.

Luke Zoltan Kelley, Laura Blecha, Lars Hernquist, Alberto Sesana, and Stephen R. Taylor. Single sources in the low-frequency gravitational wave sky: properties and time to detection by pulsar timing arrays. MNRAS, 477(1):964–976, June 2018. doi:10.1093/mnras/sty689.

Gabriella Agazie, Akash Anumarlapudi, Anne M. Archibald, Zaven Arzoumanian, Paul T. Baker, Bence Bécsy, Laura Blecha, Adam Brazier, Paul R. Brook, Sarah Burke-Spolaor, Rand Burnette, Robin Case, Maria Charisi, Shami Chatterjee, Katerina Chatziioannou, Belinda D. Cheeseboro, Siyuan Chen, Tyler Cohen, James M. Cordes, Neil J. Cornish, Fronefield Crawford, H. Thankful Cromartie, Kathryn Crowter, Curt J. Cutler, Megan E. Decesar, Dallas Degan, Paul B. Demorest, Heling Deng, Timothy Dolch, Brendan Drachler, Justin A. Ellis, Elizabeth C. Ferrara, William Fiore, Emmanuel Fonseca, Gabriel E. Freedman, Nate Garver-Daniels, Peter A. Gentile, Kyle A. Gersbach, Joseph Glaser, Deborah C. Good, Kayhan Gültekin, Jeffrey S. Hazboun, Sophie Hourihane, Kristina Islo, Ross J. Jennings, Aaron D. Johnson, Megan L. Jones, Andrew R. Kaiser, David L. Kaplan, Luke Zoltan Kelley, Matthew Kerr, Joey S. Key, Tonia C. Klein, Nima Laal, Michael T. Lam, William G. Lamb, T. Joseph W. Lazio, Natalia Lewandowska, Tyson B. Littenberg, Tingting Liu, Andrea Lommen, Duncan R. Lorimer, Jing Luo, Ryan S. Lynch, Chung-Pei Ma, Dustin R. Madison, Margaret A. Mattson, Alexander McEwen, James W. McKee, Maura A. McLaughlin, Natasha McMann, Bradley W. Meyers, Patrick M. Meyers, Chiara M. F. Mingarelli, Andrea Mitridate, Priyamvada Natarajan, Cherry Ng, David J. Nice, Stella Koch Ocker, Ken D. Olum, Timothy T. Pennucci, Benetge B. P. Perera, Polina Petrov, Nihan S. Pol, Henri A. Radovan, Scott M. Ransom, Paul S. Ray, Joseph D. Romano, Shashwat C. Sardesai, Ann Schmiedekamp, Carl Schmiedekamp, Kai Schmitz, Levi Schult, Brent J. Shapiro-Albert, Xavier Siemens, Joseph Simon, Magdalena S. Siwek, Ingrid H. Stairs, Daniel R. Stinebring, Kevin Stovall, Jerry P. Sun, Abhimanyu Susobhanan, Joseph K. Swiggum, Jacob Taylor, Stephen R. Taylor, Jacob E. Turner, Caner Unal, Michele Vallisneri, Rutger van Haasteren, Sarah J. Vigeland, Haley M. Wahl, Qiaohong Wang, Caitlin A. Witt, Olivia Young, and Nanograv Collaboration. The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background. ApJ, 951(1):L8, July 2023a. doi:10.3847/2041-8213/acdac6.

Gabriella Agazie, Akash Anumarlapudi, Anne M. Archibald, Zaven Arzoumanian, Paul T. Baker, Bence Bécsy, Laura Blecha, Adam Brazier, Paul R. Brook, Sarah Burke-Spolaor, Robin Case, J. Andrew Casey-Clyde, Maria Charisi, Shami Chatterjee, Tyler Cohen, James M. Cordes, Neil J. Cornish, Fronefield Crawford, H. Thankful Cromartie, Kathryn Crowter, Megan E. Decesar, Paul B. Demorest, Matthew C. Digman, Timothy Dolch, Brendan Drachler, Elizabeth C. Ferrara, William Fiore, Emmanuel Fonseca, Gabriel E. Freedman, Nate Garver-Daniels, Peter A. Gentile, Joseph Glaser, Deborah C. Good, Kayhan Gültekin, Jeffrey S. Hazboun, Sophie Hourihane, Ross J. Jennings, Aaron D. Johnson, Megan L. Jones, Andrew R. Kaiser, David L. Kaplan, Luke Zoltan Kelley, Matthew Kerr, Joey S. Key, Nima Laal, Michael T. Lam, William G. Lamb, T. Joseph W. Lazio, Natalia Lewandowska, Tingting Liu, Duncan R. Lorimer, Jing Luo, Ryan S. Lynch, Chung-Pei Ma, Dustin R. Madison, Alexander McEwen, James W. McKee, Maura A. McLaughlin, Natasha McMann, Bradley W. Meyers, Patrick M. Meyers, Chiara M. F. Mingarelli, Andrea Mitridate, Cherry Ng, David J. Nice, Stella Koch Ocker, Ken D. Olum, Timothy T. Pennucci, Benetge B. P. Perera, Polina Petrov, Nihan S. Pol, Henri A. Radovan, Scott M. Ransom, Paul S. Ray, Joseph D. Romano, Shashwat C. Sardesai, Ann Schmiedekamp, Carl Schmiedekamp, Kai Schmitz, Brent J. Shapiro-Albert, Xavier Siemens, Joseph Simon, Magdalena S. Siwek, Ingrid H. Stairs, Daniel R. Stinebring, Kevin Stovall, Abhimanyu Susobhanan, Joseph K. Swiggum, Jacob Taylor, Stephen R. Taylor, Jacob E. Turner, Caner Unal, Michele Vallisneri, Rutger van Haasteren, Sarah J. Vigeland, Haley M. Wahl, Caitlin A. Witt, Olivia Young, and Nanograv Collaboration. The NANOGrav 15 yr Data Set: Bayesian Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries. ApJ, 951(2):L50, July 2023b. doi:10.3847/2041-8213/ace18a.

Pau Amaro-Seoane, Jeff Andrews, Manuel Arca Sedda, Abbas Askar, Quentin Baghi, Razvan Balasov, Imre Bartos, Simone S. Bavera, Jillian Bellovary, Christopher P. L. Berry, Emanuele Berti, Stefano Bianchi, Laura Blecha, Stéphane Blondin, Tamara Bogdanović, Samuel Boissier, Matteo Bonetti, Silvia Bonoli, Elisa Bortolas, Katelyn Breivik, Pedro R. Capelo, Laurentiu Caramete, Federico Cattorini, Maria Charisi, Sylvain Chaty, Xian Chen, Martyna Chruślińska, Alvin J. K. Chua, Ross Church, Monica Colpi, Daniel D'Orazio, Camilla Danielski, Melvyn B. Davies, Pratika Dayal, Alessandra De Rosa, Andrea Derdzinski, Kyriakos Destounis, Massimo Dotti, Ioana Dutan, Irina Dvorkin, Gaia Fabj, Thierry Foglizzo, Saavik Ford, Jean-Baptiste Fouvry, Alessia Franchini, Tassos Fragos, Chris Fryer, Massimo Gaspari, Davide Gerosa, Luca Graziani, Paul Groot, Melanie Habouzit, Daryl Haggard, Zoltan Haiman, Wen-Biao Han, Alina Istrate, Peter H. Johansson, Fazeel Mahmood Khan, Tomas Kimpson, Kostas Kokkotas, Albert Kong, Valeriya Korol, Kyle Kremer, Thomas Kupfer, Astrid Lamberts, Shane Larson, Mike Lau, Dongliang Liu, Nicole Lloyd-Ronning, Giuseppe Lodato, Alessandro Lupi, Chung-Pei Ma, Tomas Maccarone, Ilya Mandel, Alberto Mangiagli, Michela Mapelli, Stéphane Mathis, Lucio Mayer, Sean McGee, Barry McKernan, M. Coleman Miller, David F. Mota, Matthew Mumpower, Syeda S. Nasim, Gijs Nelemans, Scott Noble, Fabio Pacucci, Francesca Panessa, Vasileios Paschalidis, Hugo Pfister, Delphine Porquet, John Quenby, Angelo Ricarte, Friedrich K. Röpke, John Regan, Stephan Rosswog, Ashley Ruiter, Milton Ruiz, Jessie Runnoe, Raffaella Schneider, Jeremy Schnittman,

Amy Secunda, Alberto Sesana, Naoki Seto, Lijing Shao, Stuart Shapiro, Carlos Sopuerta, Nicholas C. Stone, Arthur Suvorov, Nicola Tamanini, Tomas Tamfal, Thomas Tauris, Karel Temmink, John Tomsick, Silvia Toonen, Alejandro Torres-Orjuela, Martina Toscani, Antonios Tsokaros, Caner Unal, Verónica Vázquez-Aceves, Rosa Valiante, Maurice van Putten, Jan van Roestel, Christian Vignali, Marta Volonteri, Kinwah Wu, Ziri Younsi, Shenghua Yu, Silvia Zane, Lorenz Zwick, Fabio Antonini, Vishal Baibhav, Enrico Barausse, Alexander Bonilla Rivera, Marica Branchesi, Graziella Branduardi-Raymont, Kevin Burdge, Srija Chakraborty, Jorge Cuadra, Kristen Dage, Benjamin Davis, Selma E. de Mink, Roberto Decarli, Daniela Doneva, Stephanie Escoffier, Poshak Gandhi, Francesco Haardt, Carlos O. Lousto, Samaya Nissanke, Jason Nordhaus, Richard O'Shaughnessy, Simon Portegies Zwart, Adam Pound, Fabian Schussler, Olga Sergijenko, Alessandro Spallicci, Daniele Vernieri, and Alejandro Vigna-Gómez. Astrophysics with the Laser Interferometer Space Antenna. *Living Reviews in Relativity*, 26(1):2, December 2023. doi:10.1007/s41114-022-00041-y.

Alice Damiano, Milena Valentini, Stefano Borgani, Luca Tornatore, Giuseppe Murante, Antonio Ragagnin, Cinthia Ragone-Figueroa, and Klaus Dolag. Dynamical friction and evolution of black holes in cosmological simulations: a new implementation in OpenGadget3. *arXiv e-prints*, art. arXiv:2403.12600, March 2024. doi:10.48550/arXiv.2403.12600.

Hugo Pfister, Marta Volonteri, Yohan Dubois, Massimo Dotti, and Monica Colpi. The erratic dynamical life of black hole seeds in high-redshift galaxies. MNRAS, 486(1): 101–111, June 2019. doi:10.1093/mnras/stz822.

Christian Partmann, Thorsten Naab, Antti Rantala, Anna Genina, Matias Mannerkoski, and Peter H. Johansson. The difficult path to coalescence: massive black hole dynamics in merging low mass dark matter haloes and galaxies. *arXiv e-prints*, art. arXiv:2310.08079, October 2023. doi:10.48550/arXiv.2310.08079.

Abbas Askar, Melvyn B. Davies, and Ross P. Church. Formation of supermassive black holes in galactic nuclei - I. Delivering seed intermediate-mass black holes in massive stellar clusters. MNRAS, 502(2):2682–2700, April 2021. doi:10.1093/mnras/stab113.

Simeon Bird, Yueying Ni, Tiziana Di Matteo, Rupert Croft, Yu Feng, and Nianyi Chen. The ASTRID simulation: galaxy formation and reionization. MNRAS, 512(3):3703–3716, May 2022. doi:10.1093/mnras/stac648.

Yueying Ni, Tiziana Di Matteo, Simeon Bird, Rupert Croft, Yu Feng, Nianyi Chen, Michael Tremmel, Colin DeGraf, and Yin Li. The ASTRID simulation: the evolution of supermassive black holes. MNRAS, 513(1):670–692, June 2022. doi:10.1093/mnras/stac351.

Nianyi Chen, Yueying Ni, A. Miguel Holgado, Tiziana Di Matteo, Michael Tremmel, Colin DeGraf, Simeon Bird, Rupert Croft, and Yu Feng. Massive black hole mergers with orbital information: predictions from the ASTRID simulation. MNRAS, 514(2):2220–2238, August 2022b. doi:10.1093/mnras/stac1432.

Seppo Mikkola and Kiyotaka Tanikawa. Algorithmic regularization of the few-body problem. MNRAS, 310(3):745–749, December 1999c. doi:10.1046/j.1365-8711.1999.02982.x.

Antti Rantala, Pauli Pihajoki, Peter H. Johansson, Thorsten Naab, Natalia Lahén, and Till Sawala. Post-Newtonian Dynamical Modeling of Supermassive Black Holes in Galactic-scale Simulations. ApJ, 840(1):53, May 2017. doi:10.3847/1538-4357/aa6d65.

Antti Rantala, Pauli Pihajoki, Matias Mannerkoski, Peter H. Johansson, and Thorsten Naab. MSTAR - a fast parallelized algorithmically regularized integrator with minimum spanning tree coordinates. MNRAS, 492(3):4131–4148, March 2020. doi:10.1093/mnras/staa084.

Qirong Zhu. A momentum-conserving N-body scheme with individual time steps. New A, 85:101481, May 2021b. doi:10.1016/j.newast.2020.101481.

Giuseppe Lodato and Priyamvada Natarajan. The mass function of high-redshift seed black holes. MNRAS, 377(1):L64–L68, May 2007. doi:10.1111/j.1745-3933.2007.00304.x.

Christian Partmann, Thorsten Naab, Natalia Lahén, Antti Rantala, Michaela Hirschmann, Jessica M. Hislop, Jonathan Petersson, and Peter H. Johansson. The importance of nuclear star clusters for massive black hole growth and nuclear star formation in simulated low-mass galaxies. *arXiv e-prints*, art. arXiv:2409.18096, September 2024. doi:10.48550/arXiv.2409.18096.

T. Di Matteo, V. Springel, and L. Hernquist. Energy input from quasars regulates the growth and activity of black holes and their host galaxies. Nature, 433:604–607, February 2005. doi:10.1038/nature03335.

N. I. Shakura and R. A. Sunyaev. Black holes in binary systems. Observational appearance. A&A, 24:337–355, 1973.

M. Tremmel, F. Governato, M. Volonteri, and T. R. Quinn. Off the beaten path: a new approach to realistically model the orbital decay of supermassive black holes in galaxy formation simulations. MNRAS, 451(2):1868–1874, August 2015b. doi:10.1093/mnras/stv1060.

Nianyi Chen, Yueying Ni, Michael Tremmel, Tiziana Di Matteo, Simeon Bird, Colin DeGraf, and Yu Feng. Dynamical friction modelling of massive black holes in cosmological simulations and effects on merger rate predictions. MNRAS, 510(1):531–550, February 2022c. doi:10.1093/mnras/stab3411.

Matias Mannerkoski, Alexander Rawlings, Peter H. Johansson, Thorsten Naab, Antti Rantala, Volker Springel, Dimitrios Irodotou, and Shihong Liao. KETJU - resolving small-scale supermassive black hole dynamics in GADGET-4. MNRAS, 524(3):4062–4082, September 2023. doi:10.1093/mnras/stad2139.

Thierry Mora and Clifford M. Will. Post-Newtonian diagnostic of quasiequilibrium binary configurations of compact objects. Phys. Rev. D, 69(10):104021, May 2004. doi:10.1103/PhysRevD.69.104021.

Antti Rantala, Thorsten Naab, and Volker Springel. frost: a momentum-conserving CUDA implementation of a hierarchical fourth-order forward symplectic integrator. MNRAS, 502(4):5546–5562, April 2021c. doi:10.1093/mnras/stab057.

Anna Genina, Volker Springel, and Antti Rantala. A calibrated model for N-body dynamical friction acting on supermassive black holes. *arXiv e-prints*, art. arXiv:2405.08870, May 2024. doi:10.48550/arXiv.2405.08870.

Fazeel Mahmood Khan, Ingo Berentzen, Peter Berczik, Andreas Just, Lucio Mayer, Keigo Nitadori, and Simone Callegari. Formation and Hardening of Supermassive Black Hole Binaries in Minor Mergers of Disk Galaxies. ApJ, 756(1):30, September 2012. doi:10.1088/0004-637X/756/1/30.

Fazeel Mahmood Khan, Davide Fiacconi, Lucio Mayer, Peter Berczik, and Andreas Just. Swift Coalescence of Supermassive Black Holes in Cosmological Mergers of Massive Galaxies. ApJ, 828(2):73, September 2016. doi:10.3847/0004-637X/828/2/73.

Raphaël Errani and Julio F. Navarro. The asymptotic tidal remnants of cold dark matter subhaloes. MNRAS, 505(1):18–32, July 2021. doi:10.1093/mnras/stab1215.

Xiaolong Du, Andrew Benson, Zhichao Carton Zeng, Tommaso Treu, Annika H. G. Peter, Charlie Mace, Fangzhou Jiang, Shengqi Yang, Charles Gannon, Daniel Gilman, Anna. M. Nierenberg, and Ethan O. Nadler. Tidal evolution of cored and cuspy dark matter halos. Phys. Rev. D, 110(2):023019, July 2024. doi:10.1103/PhysRevD.110.023019.

Renuka Pechetti, Anil Seth, Nadine Neumayer, Iskren Georgiev, Nikolay Kacharov, and Mark den Brok. Luminosity Models and Density Profiles for Nuclear Star Clusters for a Nearby Volume-limited Sample of 29 Galaxies. ApJ, 900(1):32, September 2020. doi:10.3847/1538-4357/abaaa7.

David Merritt and Andras Szell. Dynamical Cusp Regeneration. ApJ, 648(2):890–899, September 2006b. doi:10.1086/506010.

Alberto Sesana and Fazeel Mahmood Khan. Scattering experiments meet N-body - I. A practical recipe for the evolution of massive black hole binaries in stellar environments. MNRAS, 454(1):L66–L70, November 2015. doi:10.1093/mnrasl/slv131.

Federica Fastidio, Alessia Gualandris, Alberto Sesana, Elisa Bortolas, and Walter Dehnen. Eccentricity evolution of PTA sources from cosmological initial conditions. *arXiv e-prints*, art. arXiv:2406.02710, June 2024. doi:10.48550/arXiv.2406.02710.

Eugene Vasiliev. AGAMA: action-based galaxy modelling architecture. MNRAS, 482(2): 1525–1544, January 2019b. doi:10.1093/mnras/sty2672.

Fazeel Mahmood Khan, Fiza Javed, Kelly Holley-Bockelmann, Lucio Mayer, Peter Berczik, and Andrea V. Macciò. The potential for long-lived intermediate mass black hole binaries in the lowest density dwarf galaxies. *arXiv e-prints*, art. arXiv:2408.14541, August 2024. doi:10.48550/arXiv.2408.14541.

Alessia Gualandris and David Merritt. Long-term Evolution of Massive Black Hole Binaries. IV. Mergers of Galaxies with Collisionally Relaxed Nuclei. ApJ, 744(1):74, January 2012b. doi:10.1088/0004-637X/744/1/74.

P. C. Peters. Gravitational Radiation and the Motion of Two Point Masses. *Physical Review*, 136(4B):1224–1232, November 1964c. doi:10.1103/PhysRev.136.B1224.

Alexander Rasskazov and David Merritt. Evolution of Binary Supermassive Black Holes in Rotating Nuclei. ApJ, 837(2):135, March 2017. doi:10.3847/1538-4357/aa6188.

M. A. Mirza, A. Tahir, F. M. Khan, H. Holley-Bockelmann, A. M. Baig, P. Berczik, and F. Chishtie. Galaxy rotation and supermassive black hole binary evolution. MNRAS, 470(1):940–947, September 2017. doi:10.1093/mnras/stx1248.

Vera C. Rubin and Jr. Ford, W. Kent. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. ApJ, 159:379, February 1970. doi:10.1086/150317.

V. C. Rubin, Jr. Ford, W. K., and N. Thonnard. Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa -¿ Sc. ApJ, 225:L107–L111, November 1978. doi:10.1086/182804.

V. C. Rubin, Jr. Ford, W. K., and N. Thonnard. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc). ApJ, 238:471–487, June 1980. doi:10.1086/158003.

Massimo Persic, Paolo Salucci, and Fulvio Stel. The universal rotation curve of spiral galaxies — I. The dark matter connection. MNRAS, 281(1):27–47, July 1996. doi:10.1093/mnras/278.1.27.

J. A. Tyson, F. Valdes, and R. A. Wenk. Detection of Systematic Gravitational Lens Galaxy Image Alignments: Mapping Dark Matter in Galaxy Clusters. ApJ, 349:L1, January 1990. doi:10.1086/185636.

F. Hammer. Thin and Giant Luminous Arcs: A Strong Test of the Lensing Cluster Mass Distribution. ApJ, 383:66, December 1991. doi:10.1086/170764.

B. Fort, O. Le Fevre, F. Hammer, and M. Cailloux. An Arc System with a Radial Gravitational Image in the Cluster MS 2137-23. ApJ, 399:L125, November 1992. doi:10.1086/186623.

O. Le Fevre, F. Hammer, M. C. Angonin, I. M. Gioia, and G. A. Luppino. Imaging of 16 Distant EMSS Clusters with Z ¿= 0.2 and L X,44 ¿=4: New Arcs and First Consequences. ApJ, 422:L5, February 1994. doi:10.1086/187198.

Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. The Structure of Cold Dark Matter Halos. ApJ, 462:563, May 1996. doi:10.1086/177173.

Gianfranco Bertone. *Particle Dark Matter : Observations, Models and Searches*. 2010.

Piero Ullio, Hongsheng Zhao, and Marc Kamionkowski. Dark-matter spike at the galactic center? Phys. Rev. D, 64(4):043504, August 2001. doi:10.1103/PhysRevD.64.043504.

David Merritt, Milos Milosavljević, Licia Verde, and Raul Jimenez. Dark Matter Spikes and Annihilation Radiation from the Galactic Center. Phys. Rev. Lett., 88(19):191301, May 2002. doi:10.1103/PhysRevLett.88.191301.

David Merritt. Evolution of the Dark Matter Distribution at the Galactic Center. Phys. Rev. Lett., 92(20):201304, May 2004. doi:10.1103/PhysRevLett.92.201304.

Gianfranco Bertone and David Merritt. Time-dependent models for dark matter at the galactic center. Phys. Rev. D, 72(10):103502, November 2005. doi:10.1103/PhysRevD.72.103502.

HongSheng Zhao and Joseph Silk. Dark minihalos with intermediate mass black holes. *Phys. Rev. Lett.*, 95:011301, Jun 2005. doi:10.1103/PhysRevLett.95.011301. URL https://link.aps.org/doi/10.1103/PhysRevLett.95.011301.

Gianfranco Bertone, Andrew R. Zentner, and Joseph Silk. New signature of dark matter annihilations: Gamma rays from intermediate-mass black holes. *Phys. Rev. D*, 72:103517, Nov 2005. doi:10.1103/PhysRevD.72.103517. URL https://link.aps.org/doi/10.1103/PhysRevD.72.103517.

Francesc Ferrer, Augusto Medeiros da Rosa, and Clifford M. Will. Dark matter spikes in the vicinity of Kerr black holes. Phys. Rev. D, 96(8):083014, October 2017. doi:10.1103/PhysRevD.96.083014.

Xiao-Jun Yue and Wen-Biao Han. Gravitational waves with dark matter minispikes: The combined effect. Phys. Rev. D, 97(6):064003, March 2018. doi:10.1103/PhysRevD.97.064003.

Xiao-Jun Yue, Wen-Biao Han, and Xian Chen. Dark Matter: An Efficient Catalyst for Intermediate-mass-ratio-inspiral Events. ApJ, 874(1):34, March 2019. doi:10.3847/1538-4357/ab06f6.

Ning Dai, Yungui Gong, Tong Jiang, and Dicong Liang. Intermediate mass-ratio inspirals with dark matter minispikes. Phys. Rev. D, 106(6):064003, September 2022. doi:10.1103/PhysRevD.106.064003.

Seiji Kawamura, Takashi Nakamura, Masaki Ando, Naoki Seto, Tomotada Akutsu, Ikkoh Funaki, Kunihito Ioka, Nobuyuki Kanda, Isao Kawano, Mitsuru Musha, Kazuhiro Nakazawa, Shuichi Sato, Takeshi Takashima, Takahiro Tanaka, Kimio Tsubono, Jun'ichi Yokoyama, Kazuhiro Agatsuma, Koh-Suke Aoyanagi, Koji Arai, Akito Araya, Naoki Aritomi, Hideki Asada, Yoichi Aso, Dan Chen, Takeshi Chiba, Toshikazu Ebisuzaki, Satoshi Eguchi, Yumiko Ejiri, Motohiro Enoki, Yoshiharu Eriguchi, Masa-Katsu Fujimoto, Ryuichi Fujita, Mitsuhiro Fukushima, Toshifumi Futamase, Rina Gondo, Tomohiro Harada, Tatsuaki Hashimoto, Kazuhiro Hayama, Wataru Hikida, Yoshiaki Himemoto, Hisashi Hirabayashi, Takashi Hiramatsu, Feng-Lei Hong, Hideyuki Horisawa, Mizuhiko

Hosokawa, Kiyotomo Ichiki, Takeshi Ikegami, Kaiki T. Inoue, Hideki Ishihara, Takehiko Ishikawa, Hideharu Ishizaki, Hiroyuki Ito, Yousuke Itoh, Kiwamu Izumi, Shinya Kanemura, Nobuki Kawashima, Fumiko Kawazoe, Naoko Kishimoto, Kenta Kiuchi, Shiho Kobayashi, Kazunori Kohri, Hiroyuki Koizumi, Yasufumi Kojima, Keiko Kokeyama, Wataru Kokuyama, Kei Kotake, Yoshihide Kozai, Hiroo Kunimori, Hitoshi Kuninaka, Kazuaki Kuroda, Sachiko Kuroyanagi, Kei-Ichi Maeda, Hideo Matsuhara, Nobuyuki Matsumoto, Yuta Michimura, Osamu Miyakawa, Umpei Miyamoto, Shinji Miyoki, Mutsuko Y. Morimoto, Toshiyuki Morisawa, Shigenori Moriwaki, Shinji Mukohyama, Shigeo Nagano, Kouji Nakamura, Hiroyuki Nakano, Kenichi Nakao, Shinichi Nakasuka, Yoshinori Nakayama, Erina Nishida, Atsushi Nishizawa, Yoshito Niwa, Taiga Noumi, Yoshiyuki Obuchi, Naoko Ohishi, Masashi Ohkawa, Kenshi Okada, Norio Okada, Koki Okutomi, Kenichi Oohara, Norichika Sago, Motoyuki Saijo, Ryo Saito, Masaaki Sakagami, Shin-Ichiro Sakai, Shihori Sakata, Misao Sasaki, Takashi Sato, Masaru Shibata, Kazunori Shibata, Ayumi Shimo-Oku, Hisaaki Shinkai, Ayaka Shoda, Kentaro Somiya, Hajime Sotani, Aru Suemasa, Naoshi Sugiyama, Yudai Suwa, Rieko Suzuki, Hideyuki Tagoshi, Fuminobu Takahashi, Kakeru Takahashi, Keitaro Takahashi, Ryutaro Takahashi, Ryuichi Takahashi, Hirotaka Takahashi, Takamori Akiteru, Tadashi Takano, Nobuyuki Tanaka, Keisuke Taniguchi, Atsushi Taruya, Hiroyuki Tashiro, Yasuo Torii, Morio Toyoshima, Shinji Tsujikawa, Akitoshi Ueda, Ken-Ichi Ueda, Takafumi Ushiba, Masayoshi Utashima, Yaka Wakabayashi, Kent Yagi, Kazuhiro Yamamoto, Toshitaka Yamazaki, Chul-Moon Yoo, Shijun Yoshida, and Taizoh Yoshino. Space gravitational-wave antennas DECIGO and B-DECIGO. *International Journal of Modern Physics D*, 28(12):1845001, January 2019. doi:10.1142/S0218271818450013.

Mahiro Abe, Philip Adamson, Marcel Borcean, Daniela Bortoletto, Kieran Bridges, Samuel P. Carman, Swapan Chattopadhyay, Jonathon Coleman, Noah M. Curfman, Kenneth DeRose, Tejas Deshpande, Savas Dimopoulos, Christopher J. Foot, Josef C. Frisch, Benjamin E. Garber, Steve Geer, Valerie Gibson, Jonah Glick, Peter W. Graham, Steve R. Hahn, Roni Harnik, Leonie Hawkins, Sam Hindley, Jason M. Hogan, Yijun Jiang (), Mark A. Kasevich, Ronald J. Kellett, Mandy Kiburg, Tim Kovachy, Joseph D. Lykken, John March-Russell, Jeremiah Mitchell, Martin Murphy, Megan Nantel, Lucy E. Nobrega, Robert K. Plunkett, Surjeet Rajendran, Jan Rudolph, Natasha Sachdeva, Murtaza Safdari, James K. Santucci, Ariel G. Schwartzman, Ian Shipsey, Hunter Swan, Linda R. Valerio, Arvydas Vasonis, Yiping Wang, and Thomas Wilkason. Matter-wave Atomic Gradiometer Interferometric Sensor (MAGIS-100). *Quantum Science and Technology*, 6 (4):044003, October 2021. doi:10.1088/2058-9565/abf719.

Caio F. B. Macedo, Paolo Pani, Vitor Cardoso, and Luís C. B. Crispino. Into the Lair: Gravitational-wave Signatures of Dark Matter. ApJ, 774(1):48, September 2013. doi:10.1088/0004-637X/774/1/48.

Lorenz Zwick, Andrea Derdzinski, Mudit Garg, Pedro R. Capelo, and Lucio Mayer. Dirty waveforms: multiband harmonic content of gas-embedded gravitational wave sources. MNRAS, 511(4):6143–6159, April 2022. doi:10.1093/mnras/stac299.

Adam Coogan, Gianfranco Bertone, Daniele Gaggero, Bradley J. Kavanagh, and David A.

Nichols. Measuring the dark matter environments of black hole binaries with gravitational waves. Phys. Rev. D, 105(4):043009, February 2022. doi:10.1103/PhysRevD.105.043009.

Daniel Baumann, Gianfranco Bertone, John Stout, and Giovanni Maria Tomaselli. Sharp Signals of Boson Clouds in Black Hole Binary Inspirals. Phys. Rev. Lett., 128(22):221102, June 2022. doi:10.1103/PhysRevLett.128.221102.

Lorenz Zwick, Pedro R. Capelo, and Lucio Mayer. Priorities in gravitational waveforms for future space-borne detectors: vacuum accuracy or environment? MNRAS, 521(3): 4645–4651, May 2023. doi:10.1093/mnras/stad707.

Philippa S. Cole, Gianfranco Bertone, Adam Coogan, Daniele Gaggero, Theophanes Karydas, Bradley J. Kavanagh, Thomas F. M. Spieksma, and Giovanni Maria Tomaselli. Distinguishing environmental effects on binary black hole gravitational waveforms. *Nature Astronomy*, 7:943–950, August 2023a. doi:10.1038/s41550-023-01990-2.

Philippa S. Cole, Adam Coogan, Bradley J. Kavanagh, and Gianfranco Bertone. Measuring dark matter spikes around primordial black holes with Einstein Telescope and Cosmic Explorer. Phys. Rev. D, 107(8):083006, April 2023b. doi:10.1103/PhysRevD.107.083006.

Otto A. Hannuksela, Kenny C. Y. Ng, and Tjonnie G. F. Li. Extreme dark matter tests with extreme mass ratio inspirals. Phys. Rev. D, 102(10):103022, November 2020. doi:10.1103/PhysRevD.102.103022.

Vitor Cardoso, Caio F. B. Macedo, and Rodrigo Vicente. Eccentricity evolution of compact binaries and applications to gravitational-wave physics. Phys. Rev. D, 103(2):023015, January 2021. doi:10.1103/PhysRevD.103.023015.

Fani Dosopoulou. Dynamical friction in dark matter spikes: corrections to Chandrasekhar's formula. *arXiv e-prints*, art. arXiv:2305.17281, May 2023. doi:10.48550/arXiv.2305.17281.

Volker Springel. The cosmological simulation code GADGET-2. MNRAS, 364(4):1105–1134, December 2005. doi:10.1111/j.1365-2966.2005.09655.x.

Qirong Zhu. A momentum-conserving N-body scheme with individual time steps. New A, 85:101481, May 2021c. doi:10.1016/j.newast.2020.101481.

Luc Blanchet. Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries. *Living Reviews in Relativity*, 17(1):2, December 2014. doi:10.12942/lrr-2014-2.

Clifford M. Will. Incorporating post-Newtonian effects in N-body dynamics. Phys. Rev. D, 89(4):044043, February 2014. doi:10.1103/PhysRevD.89.044043.

S. F. Portegies Zwart, T. C. N. Boekholt, E. H. Por, A. S. Hamers, and S. L. W. McMillan. Chaos in self-gravitating many-body systems. Lyapunov time dependence of N and the influence of general relativity. A&A, 659:A86, March 2022. doi:10.1051/0004-6361/202141789.

Seppo Mikkola and Sverre Aarseth. A Time-Transformed Leapfrog Scheme. *Celestial Mechanics and Dynamical Astronomy*, 84(4):343–354, December 2002. doi:10.1023/A:1021149313347.

Christian Hellström and Seppo Mikkola. Explicit algorithmic regularization in the few-body problem for velocity-dependent perturbations. *Celestial Mechanics and Dynamical Astronomy*, 106(2):143–156, February 2010. doi:10.1007/s10569-009-9248-8.

Hanno Rein and David S. Spiegel. IAS15: a fast, adaptive, high-order integrator for gravitational dynamics, accurate to machine precision over a billion orbits. MNRAS, 446(2): 1424–1437, January 2015. doi:10.1093/mnras/stu2164.

H. Rein and S. F. Liu. REBOUND: an open-source multi-purpose N-body code for collisional dynamics. A&A, 537:A128, January 2012. doi:10.1051/0004-6361/201118085.

Mathieu Boudaud, Thomas Lacroix, Martin Stref, Julien Lavalle, and Pierre Salati. In-depth analysis of the clustering of dark matter particles around primordial black holes. Part I. Density profiles. J. Cosmology Astropart. Phys., 2021(8):053, August 2021. doi:10.1088/1475-7516/2021/08/053.

David Merritt, Stefan Harfst, and Gianfranco Bertone. Collisionally regenerated dark matter structures in galactic nuclei. Phys. Rev. D, 75(4):043517, February 2007b. doi:10.1103/PhysRevD.75.043517.

Niklas Becker. IMRIpy: Intermediate Mass Ratio Inspirals simulator. Astrophysics Source Code Library, record ascl:2307.018, July 2023.

D. Lynden-Bell. Can spherical clusters rotate? MNRAS, 120:204, January 1960. doi:10.1093/mnras/120.3.204.

Kelly Holley-Bockelmann and Fazeel Mahmood Khan. Galaxy Rotation and Rapid Supermassive Binary Coalescence. ApJ, 810(2):139, September 2015. doi:10.1088/0004-637X/810/2/139.

Fazeel Mahmood Khan, Muhammad Awais Mirza, and Kelly Holley-Bockelmann. Inward bound: the incredible journey of massive black holes as they pair and merge - I. The effect of mass ratio in flattened rotating galactic nuclei. MNRAS, 492(1):256–267, February 2020. doi:10.1093/mnras/stz3360.

Ludovica Varisco, Elisa Bortolas, Massimo Dotti, and Alberto Sesana. Stellar hardening of massive black hole binaries: the impact of the host rotation. MNRAS, 508(1):1533–1542, November 2021. doi:10.1093/mnras/stab2649.

Long Wang, Peter Berczik, Rainer Spurzem, and M. B. N. Kouwenhoven. The Link between Ejected Stars, Hardening and Eccentricity Growth of Super Massive Black Holes in Galactic Nuclei. ApJ, 780(2):164, January 2014. doi:10.1088/0004-637X/780/2/164.

David Merritt, Alessia Gualandris, and Seppo Mikkola. Explaining the Orbits of the Galactic Center S-Stars. ApJ, 693(1):L35–L38, March 2009. doi:10.1088/0004-637X/693/1/L35.

Masaki Iwasawa, Sangyong An, Tatsushi Matsubayashi, Yoko Funato, and Junichiro Makino. Eccentric Evolution of Supermassive Black Hole Binaries. ApJ, 731(1):L9, April 2011. doi:10.1088/2041-8205/731/1/L9.

A. Tanikawa and M. Umemura. Successive Merger of Multiple Massive Black Holes in a Primordial Galaxy. ApJ, 728(2):L31, February 2011. doi:10.1088/2041-8205/728/2/L31.

Matteo Bonetti, Francesco Haardt, Alberto Sesana, and Enrico Barausse. Post-Newtonian evolution of massive black hole triplets in galactic nuclei - II. Survey of the parameter space. MNRAS, 477(3):3910–3926, July 2018. doi:10.1093/mnras/sty896.

Matias Mannerkoski, Peter H. Johansson, Antti Rantala, Thorsten Naab, and Shihong Liao. Resolving the Complex Evolution of a Supermassive Black Hole Triplet in a Cosmological Simulation. ApJ, 912(2):L20, May 2021. doi:10.3847/2041-8213/abf9a5.

H. Koehn, A. Just, P. Berczik, and M. Tremmel. Dynamics of supermassive black hole triples in the ROMULUS25 cosmological simulation. A&A, 678:A11, October 2023. doi:10.1051/0004-6361/202347093.

H. Bondi and F. Hoyle. On the mechanism of accretion by stars. MNRAS, 104:273, January 1944. doi:10.1093/mnras/104.5.273.

Richard Edgar. A review of Bondi-Hoyle-Lyttleton accretion. New A Rev., 48(10):843–859, September 2004. doi:10.1016/j.newar.2004.06.001.

Patryk Mach and Andrzej Odrzywołek. Accretion of dark matter onto a moving schwarzschild black hole: An exact solution. *Phys. Rev. Lett.*, 126:101104, Mar 2021. doi:10.1103/PhysRevLett.126.101104. URL `https://link.aps.org/doi/10.1103/PhysRevLett.126.101104`.

David A. Nichols, Benjamin A. Wade, and Alexander M. Grant. Secondary accretion of dark matter in intermediate mass-ratio inspirals: Dark-matter dynamics and gravitational-wave phase. *arXiv e-prints*, art. arXiv:2309.06498, September 2023. doi:10.48550/arXiv.2309.06498.

Brian D. Fields, Stuart L. Shapiro, and Jessie Shelton. Galactic Center Gamma-Ray Excess from Dark Matter Annihilation: Is There a Black Hole Spike? Phys. Rev. Lett., 113(15): 151302, October 2014. doi:10.1103/PhysRevLett.113.151302.

Stuart L. Shapiro and Jessie Shelton. Weak annihilation cusp inside the dark matter spike about a black hole. Phys. Rev. D, 93(12):123510, June 2016. doi:10.1103/PhysRevD.93.123510.

Thomas Lacroix. Dynamical constraints on a dark matter spike at the Galactic centre from stellar orbits. A&A, 619:A46, November 2018. doi:10.1051/0004-6361/201832652.

Eugene Vasiliev. Dark matter annihilation near a black hole: Plateau versus weak cusp. Phys. Rev. D, 76(10):103532, November 2007. doi:10.1103/PhysRevD.76.103532.

MAGIC Collaboration. Limits to dark matter annihilation cross-section from a combined analysis of MAGIC and Fermi-LAT observations of dwarf satellite galaxies. J. Cosmology Astropart. Phys., 2016(2):039–039, February 2016. doi:10.1088/1475-7516/2016/02/039.

Jessie Shelton, Stuart L. Shapiro, and Brian D. Fields. Black Hole Window into p - Wave Dark Matter Annihilation. Phys. Rev. Lett., 115(23):231302, December 2015. doi:10.1103/PhysRevLett.115.231302.

A. de Angelis, V. Tatischeff, I. A. Grenier, J. McEnery, M. Mallamaci, M. Tavani, U. Oberlack, L. Hanlon, R. Walter, A. Argan, P. von Ballmoos, A. Bulgarelli, A. Bykov, M. Hernanz, G. Kanbach, I. Kuvvetli, M. Pearce, A. Zdziarski, J. Conrad, G. Ghisellini, A. Harding, J. Isern, M. Leising, F. Longo, G. Madejski, M. Martinez, M. N. Mazziotta, J. M. Paredes, M. Pohl, R. Rando, M. Razzano, A. Aboudan, M. Ackermann, A. Addazi, M. Ajello, C. Albertus, J. M. Álvarez, G. Ambrosi, S. Antón, L. A. Antonelli, A. Babic, B. Baibussinov, M. Balbo, L. Baldini, S. Balman, C. Bambi, U. Barres de Almeida, J. A. Barrio, R. Bartels, D. Bastieri, W. Bednarek, D. Bernard, E. Bernardini, T. Bernasconi, B. Bertucci, A. Biland, E. Bissaldi, M. Boettcher, V. Bonvicini, V. Bosch-Ramon, E. Bottacini, V. Bozhilov, T. Bretz, M. Branchesi, V. Brdar, T. Bringmann, A. Brogna, C. Budtz Jørgensen, G. Busetto, S. Buson, M. Busso, A. Caccianiga, S. Camera, R. Campana, P. Caraveo, M. Cardillo, P. Carlson, S. Celestin, M. Cermeño, A. Chen, C. C. Cheung, E. Churazov, S. Ciprini, A. Coc, S. Colafrancesco, A. Coleiro, W. Collmar, P. Coppi, R. Curado da Silva, S. Cutini, F. D'Ammando, B. de Lotto, D. de Martino, A. De Rosa, M. Del Santo, L. Delgado, R. Diehl, S. Dietrich, A. D. Dolgov, A. Domínguez, D. Dominis Prester, I. Donnarumma, D. Dorner, M. Doro, M. Dutra, D. Elsaesser, M. Fabrizio, A. Fernández-Barral, V. Fioretti, L. Foffano, V. Formato, N. Fornengo, L. Foschini, A. Franceschini, A. Franckowiak, S. Funk, F. Fuschino, D. Gaggero, G. Galanti, F. Gargano, D. Gasparrini, R. Gehrz, P. Giammaria, N. Giglietto, P. Giommi, F. Giordano, M. Giroletti, G. Ghirlanda, N. Godinovic, C. Gouiffés, J. E. Grove, C. Hamadache, D. H. Hartmann, M. Hayashida, A. Hryczuk, P. Jean, T. Johnson, J. José, S. Kaufmann, B. Khelifi, J. Kiener, J. Knödlseder, M. Kole, J. Kopp, V. Kozhuharov, C. Labanti, S. Lalkovski, P. Laurent, O. Limousin, M. Linares, E. Lindfors, M. Lindner, J. Liu, S. Lombardi, F. Loparco, R. López-Coto, M. López Moya, B. Lott, P. Lubrano, D. Malyshev, N. Mankuzhiyil, K. Mannheim, M. J. Marchã, A. Marcianò, B. Marcote, M. Mariotti, M. Marisaldi, S. McBreen, S. Mereghetti, A. Merle, R. Mignani, G. Minervini, A. Moiseev, A. Morselli, F. Moura, K. Nakazawa, L. Nava, D. Nieto, M. Orienti, M. Orio, E. Orlando, P. Orleanski, S. Paiano, R. Paoletti, A. Papitto, M. Pasquato, B. Patricelli, M. Á. Pérez-García, M. Persic, G. Piano, A. Pichel, M. Pimenta, C. Pittori, T. Porter, J. Poutanen, E. Prandini, N. Prantzos, N. Produit, S. Profumo, F. S. Queiroz, S. Rainó, A. Raklev, M. Regis, I. Reichardt, Y. Rephaeli, J. Rico, W. Rodejohann, G. Rodriguez Fernandez, M. Roncadelli, L. Roso, A. Rovero, R. Ruffini, G. Sala, M. A. Sánchez-Conde, A. Santangelo, P. Saz Parkinson, T. Sbarrato, A. Shearer, R. Shellard, K. Short, T. Siegert, C. Siqueira, P. Spinelli, A. Stamerra, S. Starrfield, A. Strong, I. Strümke, F. Tavecchio, R. Taverna, T. Terzić, D. J. Thompson, O. Tibolla, D. F.

Torres, R. Turolla, A. Ulyanov, A. Ursi, A. Vacchi, J. van den Abeele, G. Vankova-Kirilovai, C. Venter, F. Verrecchia, P. Vincent, X. Wang, C. Weniger, X. Wu, G. Zaharijaš, L. Zampieri, S. Zane, S. Zimmer, A. Zoglauer, and E-Astrogam Collaboration. Science with e-ASTROGAM. A space mission for MeV-GeV gamma-ray astrophysics. *Journal of High Energy Astrophysics*, 19:1–106, August 2018. doi:10.1016/j.jheap.2018.07.001.

Jarrod R. Hurley, Onno R. Pols, and Christopher A. Tout. Comprehensive analytic formulae for stellar evolution as a function of mass and metallicity. MNRAS, 315(3):543–569, July 2000. doi:10.1046/j.1365-8711.2000.03426.x.

Stanislav Babak, Jonathan Gair, Alberto Sesana, Enrico Barausse, Carlos F. Sopuerta, Christopher P. L. Berry, Emanuele Berti, Pau Amaro-Seoane, Antoine Petiteau, and Antoine Klein. Science with the space-based interferometer LISA. V. Extreme mass-ratio inspirals. Phys. Rev. D, 95(10):103012, May 2017. doi:10.1103/PhysRevD.95.103012.

Niklas Becker. Dancing above the abyss: Environmental effects and dark matter signatures in inspirals into massive black holes. *arXiv e-prints*, art. arXiv:2404.02808, April 2024. doi:10.48550/arXiv.2404.02808.

Kayhan Gültekin, Douglas O. Richstone, Karl Gebhardt, Tod R. Lauer, Scott Tremaine, M. C. Aller, Ralf Bender, Alan Dressler, S. M. Faber, Alexei V. Filippenko, Richard Green, Luis C. Ho, John Kormendy, John Magorrian, Jason Pinkney, and Christos Siopis. The M-σ and M-L Relations in Galactic Bulges, and Determinations of Their Intrinsic Scatter. ApJ, 698(1):198–221, June 2009. doi:10.1088/0004-637X/698/1/198.

John Kormendy and Luis C. Ho. Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies. ARA&A, 51(1):511–653, August 2013c. doi:10.1146/annurev-astro-082708-101811.

David Merritt, Seppo Mikkola, and Andras Szell. Long-Term Evolution of Massive Black Hole Binaries. III. Binary Evolution in Collisional Nuclei. ApJ, 671(1):53–72, December 2007c. doi:10.1086/522691.

Verónica Vázquez-Aceves, Pau Amaro Seoane, Dana Kuvatova, Maxim Makukov, Chingis Omarov, and Denis Yurin. Intermediate-mass ratio inspirals in merging elliptical galaxies. MNRAS, 518(2):2113–2118, January 2023. doi:10.1093/mnras/stac3286.

Ning Dai, Yungui Gong, Yang Zhao, and Tong Jiang. Extreme mass ratio inspirals in galaxies with dark matter halos. *arXiv e-prints*, art. arXiv:2301.05088, January 2023. doi:10.48550/arXiv.2301.05088.

Becca Ewing, Surabhi Sachdev, Ssohrab Borhanian, and B. S. Sathyaprakash. Archival searches for stellar-mass binary black holes in LISA data. Phys. Rev. D, 103(2):023025, January 2021. doi:10.1103/PhysRevD.103.023025.

Bradley J. Kavanagh, Theophanes K. Karydas, Gianfranco Bertone, Pierfrancesco Di Cintio, and Mario Pasquato. Sharpening the dark matter signature in gravitational waveforms II: Numerical simulations with the NbodyIMRI code. *arXiv e-prints*, art. arXiv:2402.13762, February 2024. doi:10.48550/arXiv.2402.13762.

S. Portegies Zwart, S. L. W. McMillan, E. van Elteren, I. Pelupessy, and N. de Vries. Multi-physics simulations using a hierarchical interchangeable software interface. *Computer Physics Communications*, 184(3):456–468, March 2013. doi:10.1016/j.cpc.2012.09.024.

Simon Portegies Zwart and Steve McMillan. *Astrophysical Recipes; The art of AMUSE*. 2018. doi:10.1088/978-0-7503-1320-9.

J. Bédorf and S. Portegies Zwart. A pilgrimage to gravity on GPUs. *European Physical Journal Special Topics*, 210(1):201–216, August 2012. doi:10.1140/epjst/e2012-1647-6.

R. Capuzzo-Dolcetta, M. Spera, and D. Punzo. A fully parallel, high precision, N-body code running on hybrid computing platforms. *Journal of Computational Physics*, 236: 580–593, March 2013. doi:10.1016/j.jcp.2012.11.013.

Peter Berczik, Keigo Nitadori, Shiyan Zhong, Rainer Spurzem, Tsuyoshi Hamada, Xiaowei Wang, Ingo Berentzen, Alexander Veles, and Wei Ge. High performance massively parallel direct n-body simulations on large gpu clusters. In *International conference on high performance computing*, pages 8–18, 2011.

Fazeel Mahmood Khan, Andreas Just, and David Merritt. Efficient Merger of Binary Supermassive Black Holes in Merging Galaxies. ApJ, 732(2):89, May 2011. doi:10.1088/0004-637X/732/2/89.

Jeroen Bédorf, Evghenii Gaburov, and Simon Portegies Zwart. A sparse octree gravitational N-body code that runs entirely on the GPU processor. *Journal of Computational Physics*, 231(7):2825–2839, April 2012. doi:10.1016/j.jcp.2011.12.024.

P.R. Atkinson. *Enabling Task Parallelism for Many-core Architectures*. University of Bristol, 2021. URL `https://books.google.com/books?id=CQvRzgEACAAJ`.