

# Hanging on the cliff: Extreme mass ratio inspiral formation with local two-body relaxation and post-Newtonian dynamics

Davide Mancieri<sup>1,2,3,\*</sup>, Luca Broggi<sup>2,3</sup>, Matteo Bonetti<sup>2,3,4</sup>, and Alberto Sesana<sup>2,3,4</sup>

<sup>1</sup> Dipartimento di Fisica, Università degli Studi di Trento, Via Sommarive 14, 38123 Povo, Italy

<sup>2</sup> Dipartimento di Fisica “G. Occhialini”, Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy

<sup>3</sup> INFN – Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy

<sup>4</sup> INAF – Osservatorio Astronomico di Brera, Via Brera 20, 20121 Milano, Italy

Received 19 September 2024 / Accepted 20 January 2025

## ABSTRACT

Extreme mass ratio inspirals (EMRIs) are anticipated to be primary gravitational wave sources for the Laser Interferometer Space Antenna (LISA). They form in dense nuclear clusters when a compact object is captured by the central massive black holes (MBHs) as a consequence of the frequent two-body interactions occurring between orbiting objects. The physics of this process is complex and requires detailed statistical modelling of a multi-body relativistic system. We present a novel Monte Carlo approach to evolving the post-Newtonian (PN) equations of motion of a compact object orbiting an MBH. The approach accounts for the effects of two-body relaxation locally on the fly, without leveraging on the common approximation of orbit-averaging. We applied our method to study the function  $S(a_0)$ , describing the fraction of EMRI to total captures (including EMRIs and direct plunges, DPs) as a function of the initial semi-major axis  $a_0$  for compact objects orbiting central MBHs with  $M_\bullet \in [10^4 M_\odot, 4 \times 10^6 M_\odot]$ . The past two decades have consolidated a picture in which  $S(a_0) \rightarrow 0$  at large initial semi-major axes, with a sharp transition from EMRIs to DPs occurring around a critical scale  $a_c$ . A recent study challenges this notion for low-mass MBHs, finding EMRIs forming at  $a \gg a_c$ , which were called ‘cliffhangers’. Our simulations confirm the existence of cliffhanger EMRIs, which we find to be more common than previously inferred. Cliffhangers start to appear for  $M_\bullet \lesssim 3 \times 10^5 M_\odot$  and can account for up to 55% of the overall EMRIs forming at those masses. We find  $S(a_0) \gg 0$  for  $a \gg a_c$ , reaching values as high as 0.6 for  $M_\bullet = 10^4 M_\odot$ , much higher than previously found. We test how these results are influenced by different assumptions on the dynamics used to evolve the system and treatment of two-body relaxation. We find that the PN description of the system greatly enhances the number of EMRIs by shifting  $a_c$  to larger values at all MBH masses. Conversely, the local treatment of relaxation has a mass-dependent impact, significantly boosting the number of cliffhangers at low MBH masses compared to an orbit-averaged treatment. These findings highlight the shortcomings of standard approximations used in the EMRI literature and the importance of carefully modelling the (relativistic) dynamics of these systems. The emerging picture is more complex than previously thought, and should be considered in future estimates of rates and properties of EMRIs detectable by LISA.

**Key words.** black hole physics – gravitational waves – galaxies: nuclei – methods: numerical

## 1. Introduction

Most galaxies host a massive black hole (MBH) at their core, with masses in the range of  $10^5$ – $10^{10} M_\odot$  (Kormendy & Richstone 1995; Magorrian et al. 1998; Kormendy & Ho 2013). Often these MBHs are found at the centre of nuclear clusters (Carollo et al. 1997; Böker et al. 2002; Côté et al. 2006; Schödel et al. 2007, 2009; Turner et al. 2012): highly dense clusters of stars and compact objects (COs) observed in the nuclei of many galaxies. In the central region of nuclear clusters, densities can reach up to  $10^7 M_\odot \text{pc}^{-3}$  (Neumayer et al. 2020), facilitating the scattering of COs onto tight and eccentric orbits around the central MBH, thus forming binaries with extreme mass ratios (Binney & Tremaine 2008; Merritt 2013a).

General relativity (GR) predicts that these binary systems emit gravitational waves (GWs), mostly during each CO pericentre passage (Misner et al. 1973). Over time the CO orbit shrinks and circularises (Peters 1964) as a consequence of the loss of energy and angular momentum of the binary system, which

are carried away by GWs (Maggiore 2007). Accordingly, the CO slowly inspirals towards the MBH, completing thousands of orbital cycles, eventually merging with it (Amaro-Seoane 2018). These GW sources are known as extreme mass ratio inspirals (EMRIs) due to the large mass imbalance between the pair of objects that make up the binary system: the MBH and the orbiting CO (Amaro-Seoane et al. 2007; Amaro Seoane 2022). Alternatively, if the merger occurs after few orbits or in a head-on collision, the event is referred to as a direct plunge (DP).

Extreme mass ratio inspirals that form around MBHs with masses of the order of  $10^4$ – $10^7 M_\odot$  are expected to emit GWs in the millihertz frequency range (Barack & Cutler 2004; Amaro-Seoane 2018). The numerous bursts of GWs emitted during an EMRI are anticipated to build up a significant signal-to-noise ratio (Barack & Cutler 2004; Babak et al. 2017), enabling the detection of such events by the upcoming Laser Interferometer Space Antenna (LISA, Amaro-Seoane et al. 2017, 2023; Colpi et al. 2024). Detecting EMRIs will enable us to test GR (Ryan 1997; Barack & Cutler 2007; Ghosh & Chakravarti 2024), estimate cosmological parameters (MacLeod & Hogan 2008; Laghi et al. 2021), and investigate dormant MBHs (Gair et al. 2010; Gallo & Sesana 2019).

\* Corresponding author; [d.mancieri@campus.unimib.it](mailto:d.mancieri@campus.unimib.it)

Recent works (e.g. Metzger et al. 2022; Franchini et al. 2023; Linial & Metzger 2023) have also proposed that EMRIs might be responsible for the novel, currently unexplained, X-ray variability phenomenon of quasi-periodic eruptions (Miniutti et al. 2019; Giustini et al. 2020; Arcodia et al. 2021, 2024), and recent observations agree with such an interpretation (Nicholl et al. 2024). This would offer a way to detect and study EMRIs within the electromagnetic spectrum, enabling multi-messenger observations that would boost their potential as astrophysical and cosmological probes (e.g. Toubiana et al. 2021).

Extreme mass ratio inspirals can form in many different ways, including (i) the tidal separation of a stellar-mass black hole (BH) binary system by the central MBH followed by the capture of one of the binary components (Hills 1988; Miller et al. 2005); (ii) the formation or capture and subsequent migration of BHs within accretion discs surrounding MBHs (Levin 2007; Pan & Yang 2021); (iii) the capture of a BH remnant by an MBH as a consequence of supernova kicks (Bortolas et al. 2017; Bortolas & Mapelli 2019); and (iv) the scatter of BHs onto tight orbits due to the gravitational influence of a secondary MBH in the aftermath of galaxy mergers (Chen et al. 2011; Bode & Wegg 2014; Naoz et al. 2022; Mazzolari et al. 2022).

The standard and ubiquitous formation channel for EMRI formation, which we explore in this work, is the scatter of a CO onto a sufficiently tight and eccentric orbit as a consequence of two-body relaxation (Alexander 2017; Amaro-Seoane 2018). The estimated EMRI formation rate through this channel is in the range of  $10^{-8}$ – $10^{-6}$  EMRIs per year for a central MBH of mass  $M_{\bullet} = 10^6 M_{\odot}$  (Hills & Bender 1995; Gair et al. 2004; Rubbo et al. 2006). Considering also the uncertainty in the number density of MBHs inside the observable range of LISA and the fraction of these that are hosted in nuclear clusters, the predicted number of EMRIs that will be detected by LISA during its lifetime spans three orders of magnitude, from a few events per year to a few thousand (Babak et al. 2017).

Simulating the formation of EMRIs in this standard channel poses considerable challenges. The most accurate description of the whole nuclear cluster would involve a direct N-body computation of the orbits of all stars and COs within it. However, the need to follow these objects for thousands of orbits, and the high mass ratio between them and the central MBH pose an insurmountable challenge for N-body approaches, which show problems even in the simplest Newtonian regime (Amaro-Seoane 2018). Alternative methods, such as solving the steady-state Fokker-Planck equation for the distribution function of the system (e.g. Amaro-Seoane & Preto 2011; Bar-Or & Alexander 2016; Vasiliev 2017; Broggi et al. 2022; Kaur et al. 2024) or employing Monte Carlo codes (e.g. Hénon 1971; Giersz 1998; Joshi et al. 2000; Freitag & Benz 2001; Pattabiraman et al. 2013; Fragione & Sari 2018; Zhang & Amaro Seoane 2024), offer faster approximated solutions, but carry inherent limitations and assumptions.

From the aforementioned body of work a picture has emerged in which there exists a critical initial semi-major axis  $a_c$  for the CO orbit around the MBH determining its eventual fate (Hopman & Alexander 2005; Bar-Or & Alexander 2016; Raveh & Perets 2021). In short, COs starting their inspiral from orbits with  $a_0 < a_c$  would only result in the formation of EMRIs, while only DPs would form when  $a_0 > a_c$ , with a sharp transition between the two regimes around  $a_c$ . Notably, this finding has been derived from detailed simulations of nuclear clusters hosting an MBH with  $M_{\bullet} \geq 10^6 M_{\odot}$  and then extrapolated at lower masses. Recently, however, this dichotomy has been questioned

in Qunbar & Stone (2024) (hereafter QS24) for the case of low-mass MBHs and intermediate-mass BHs, opening to the possibility of EMRIs forming even from initially wide orbits, dubbed cliffhanger EMRIs. This is an intriguing claim that might bear interesting consequences for estimating LISA EMRI detection rates.

As mentioned above, the techniques and codes employed to verify both the existence of  $a_c$  and the appearance of cliffhanger EMRIs rely on a number of assumptions and approximations. A notable approximation is that of being allowed to average the effects on the orbit of two-body relaxation over an entire orbital period (e.g. Cohn & Kulsrud 1978; Hopman & Alexander 2005, QS24). However, this simplification might not accurately describe the system, as two-body encounters strongly depend on the instantaneous velocity of the orbiting CO (Binney & Tremaine 2008; Merritt 2013a), which significantly changes during a highly eccentric orbit. Another is to use Newtonian dynamics to describe the system. GR effects are usually included only by adding extra orbit-averaged loss terms due to GW emission and by adjusting the CO capture condition by a suitable factor.

In this work we present a novel post-Newtonian (PN) Monte Carlo code that evolves the PN dynamics of a CO orbiting an MBH. The code treats two-body relaxation locally, by applying multiple small kicks to the CO orbit within a single orbit, and can account for an arbitrary number of stellar components contributing to relaxation and to the potential in which the CO is evolving. As a first application, here we apply this code to study the fraction of EMRIs relative to DPs that form around MBHs of different masses as a function of the initial semi-major axis of the orbit  $a_0$ . This allows us to verify the existence of  $a_c$  and its dependence on the system parameters, and the emergence of the novel population of cliffhanger EMRIs found in QS24.

The study presented in QS24 has some limitations which our code can improve upon. It employs a Monte Carlo approach to evolve in time the Newtonian energy and angular momentum of an orbit subjected to two-body relaxation, which is orbit-averaged over more than a full orbit at times. The loss of energy and angular momentum due to GW radiation is accounted for through equations detailed in Peters (1964), which are orbit-averaged as well. Moreover, only two-body encounters between a stellar-mass BH of mass  $m_{\bullet} = 10 M_{\odot}$  and a single-mass population of  $m_{\star} = 1 M_{\odot}$  stars surrounding the central MBH are considered. The stars are assumed to follow a simple isotropic power-law numeric density, and their global gravitational potential is not considered. In comparison, in this work we track the orbits of a stellar-mass BH in the total potential well of an MBH and of two populations of objects surrounding it: stars with masses  $m_{\star}$  and stellar-mass BHs with masses  $m_{\bullet}$ . We describe both populations with Dehnen numeric density profiles (Dehnen 1993; Tremaine et al. 1994). We also solve the equations of motion up to the 2.5PN term<sup>1</sup>, thus naturally accounting for GW radiation, and we locally correct the shape of the orbit for two-body relaxation multiple times within each orbital period.

The paper is organised as follows. In Sect. 2 we review the aspects of the theory relevant to our implementation and predict some behaviours of cliffhanger EMRIs with analytical means. In Sect. 3 we delve into the specifics of our methods, to then present the results of our simulations in Sect. 4. Finally, in Sect. 5 we discuss our findings and draw our conclusions.

<sup>1</sup> In what follows, we call ‘nPN term’ the term of the Hamiltonian defined in Eq. (41), which is suppressed by a factor  $(1/c)^{2n}$  with respect to the leading term.

## 2. Theory of relaxation-driven EMRI formation

In this section we first introduce some notions on the treatment of two-body interactions in the loss cone framework. Then, we give operative definitions for EMRIs and DPs. Finally, we briefly review the literature on the study of the EMRI-to-DP ratio and draw some first theoretical results on cliffhanger EMRIs.

### 2.1. Loss cone and diffusion theory

The main actor in a nuclear cluster is the central MBH, as it strongly affects the motion of stars and COs within its influence radius, which we define in this work as (Peebles 1972)

$$R_{\text{inf}} = \frac{GM_{\bullet}}{\sigma^2}. \quad (1)$$

Here  $G$  is the gravitational constant,  $M_{\bullet}$  is the mass of the MBH, and  $\sigma$  is the velocity dispersion of stars in the bulge of the host galaxy. The velocity dispersion can be related to the mass of the central MBH through the  $M$ - $\sigma$  relation (Ferrarese & Merritt 2000; Gebhardt et al. 2000), that we employ using the best fit from Gültekin et al. (2009):

$$\sigma = 70 \left( \frac{M_{\bullet}}{1.53 \times 10^6 M_{\odot}} \right)^{1/4.24} \text{ km/s}. \quad (2)$$

The nuclear star cluster is then composed by many other celestial bodies, which can be grouped into populations based on their nature and mass (e.g. Sun-like stars, stellar-mass BHs, neutron stars, and so on). We start by considering an arbitrary number of populations, later specialising to a system with only two (which can be identified generically as solar-mass stars and stellar-mass BHs).

We label each population with an  $\alpha$  index<sup>2</sup> and model their number distribution with Dehnen profiles (Dehnen 1993; Tremaine et al. 1994):

$$n_{\alpha}(r) = \frac{3 - \gamma_{\alpha}}{4\pi} \frac{M_{\text{tot},\alpha}}{m_{\alpha}} \frac{r_{a,\alpha}}{r^{\gamma_{\alpha}} (r + r_{a,\alpha})^{4-\gamma_{\alpha}}}. \quad (3)$$

Here  $m_{\alpha}$  is the individual mass of the objects in the  $\alpha$  population, while  $M_{\text{tot},\alpha}$  is their total mass. The exponent  $\gamma_{\alpha}$ , comprised in the range  $[0, 3)$ , is the slope of the profile for  $r \rightarrow 0$  and  $r_{a,\alpha}$  is the scale radius at which the distribution smoothly switches to a slope of  $-4$  for  $r \rightarrow \infty$ . Finally,  $r$  measures the distance from the centre of the distribution. We consider each profile in a steady-state configuration, with their centres fixed at the origin of our coordinate system, just like the MBH.

The gravitational potential produced by population  $\alpha$  is (Dehnen 1993; Tremaine et al. 1994)

$$\phi_{\alpha}(r) = \begin{cases} -\frac{1}{2 - \gamma_{\alpha}} \left( 1 - \left( \frac{r}{r + r_{a,\alpha}} \right)^{2-\gamma_{\alpha}} \right) \frac{GM_{\text{tot},\alpha}}{r_{a,\alpha}} & \text{if } \gamma_{\alpha} \neq 2 \\ \ln \left( \frac{r}{r + r_{a,\alpha}} \right) \frac{GM_{\text{tot},\alpha}}{r_{a,\alpha}} & \text{if } \gamma_{\alpha} = 2 \end{cases}, \quad (4)$$

so the total potential at a given position is given by

$$\psi(r) = - \left( -\frac{GM_{\bullet}}{r} + \sum_{\alpha} \phi_{\alpha}(r) \right), \quad (5)$$

<sup>2</sup> To give a concrete example, we use  $\alpha = \star$  to refer to a population of solar-mass stars, and  $\alpha = \bullet$  to refer to a population of stellar-mass BHs.

where we flipped the overall sign adopting the convention that the potential experienced by any object in the distribution is positive definite.

The central MBH affects the distribution of surrounding objects also by removing bodies that pass too close to it at orbital pericentre. For example, stars can be tidally disrupted and subsequently accreted or directly swallowed, whereas COs can be gravitationally captured resulting in EMRIs and DPs. This happens if the velocity vector  $\mathbf{v}$  of a bound object points towards the central MBH within a small solid angle, inside the so-called loss cone: a cone-shape region in velocity space containing objects which will be lost inside the MBH at the pericentre of their current orbit (see Alexander 2017; Amaro-Seoane 2018). In nuclear clusters, objects can be brought closer to or directly inside the loss cone as a consequence of two-body encounters within each other.

Let us now consider a particular stellar-mass BH of mass  $m_{\bullet} = 10 M_{\odot}$  and velocity magnitude  $v$  having a sequence of two-body encounters with the background bodies; we decompose its velocity change as the sum of  $\Delta v_{\parallel}$ , parallel to the initial line of motion, and  $\Delta v_{\perp}$ , in the plane orthogonal to the original direction of motion. At any point, given the position  $r$  and velocity  $v$  of the stellar-mass BH, its orbital binding energy per unit mass can be computed as

$$\mathcal{E}(r, v) = - \left( \frac{H(r, v)}{m_{\bullet}} + \sum_{\alpha} \phi_{\alpha}(r) \right), \quad (6)$$

where  $H$  is the Hamiltonian of the binary comprised of the MBH and the orbiting stellar-mass BH, and we define the energy so that  $\mathcal{E} < 0$  for bound orbits. We report the expression of  $H$  we used in this work in Eq. (41).

Given the random nature of the scattering process, different encounters will result in different values for  $\Delta v_{\parallel}$  and  $\Delta v_{\perp}$ . It is for this reason that the problem is better approached statistically, by looking at the diffusion coefficients (i.e. the expectation values per unit of time) of  $\Delta v_{\parallel}$ ,  $\Delta v_{\perp}$ , and related quantities. We use the symbol  $D[X]$  for the diffusion coefficient of the generic quantity  $X$ , and we explicitly write those relevant to our study in Appendix A. Crucially, these are local diffusion coefficients, meaning they depend on the exact position and velocity of the orbiting stellar-mass BH. The usual approach in literature is instead that of employing orbit-averaged diffusion coefficients  $D_0[X]$  to account for two-body encounters, which follow from their local versions as (Binney & Tremaine 2008; Merritt 2013a):

$$D_0[X] = \frac{2}{P} \int_{r_-}^{r_+} \frac{dr}{v_r} D[X]. \quad (7)$$

Here  $r_+$  and  $r_-$  respectively indicate the apocentre and pericentre of the orbit,  $P$  its radial period, and  $v_r$  the radial component of the velocity. If we indicate with  $a$  and  $e$  the semi-major axis and the eccentricity of an orbit, their Keplerian estimates are

$$r_+ = a(1 + e), \quad r_- = a(1 - e) \quad (8)$$

and

$$P = 2\pi \left( \frac{a^3}{G(M_{\bullet} + m_{\bullet})} \right)^{1/2}. \quad (9)$$

We note that the orbit is an ellipse with semi-major axis  $a$  and eccentricity  $e$  only in the Keplerian approximation, that holds

where the stellar potential is negligible ( $r_+ \lesssim R_{\text{inf}}$ ), and the orbit is non-relativistic ( $r_- \gg R_g$ ). Here

$$R_g = \frac{GM_\bullet}{c^2} \quad (10)$$

is the gravitational radius of the MBH, where  $c$  is the speed of light in vacuum. For orbits where the Keplerian approximation does not hold, our definitions of  $a$  and  $e$  do not have a simple, direct interpretation (see Sect. 2.2 below).

In Appendix A we show that a key ingredient to compute the local diffusion coefficients is  $f_\alpha(\mathcal{E})$ : the ergodic distribution function of objects in population  $\alpha$ . In particular,  $\mathcal{E}$  is the orbital binding energy per unit mass of the generic  $\alpha$  particle. The ergodic distribution function corresponding to a spherical density profile can be computed using Eddington's formula (Eddington 1916; Merritt 2013a, see also our Appendix A), which simplifies to the expression

$$f_\alpha(\mathcal{E}) = \frac{1}{\sqrt{8}\pi^2} \int_0^\mathcal{E} \frac{d\psi}{\sqrt{\mathcal{E}-\psi}} \frac{d^2 n_\alpha}{d\psi^2} \quad (11)$$

given our assumptions. For a Dehnen density profile,  $n_\alpha \propto \psi^4$  at infinite distance, as  $n_\alpha \propto r^{-4}$  and  $\psi \propto r^{-1}$ . Thus,

$$\left. \frac{dn_\alpha}{d\psi} \right|_{\psi=0} = 4\psi^3 \Big|_{\psi=0} = 0. \quad (12)$$

In general, the spherically symmetric distribution function  $f$  can depend also on the quantity

$$\mathcal{R}(\mathcal{E}, J) = \frac{J^2}{J_c^2(\mathcal{E})}, \quad (13)$$

where  $J$  is the magnitude of the specific angular momentum of the orbiting body, and  $J_c$  is its value for a circular orbit of energy  $\mathcal{E}$ . Cohn & Kulsrud (1978) showed that

$$f(\mathcal{E}, \mathcal{R}) = f_N(\mathcal{E}) \frac{\ln \mathcal{R} - \ln \mathcal{R}_q}{\mathcal{R}_q - 1 - \ln \mathcal{R}_q} \quad (14)$$

is a steady-state expression for  $f$  near the loss cone. Here  $f_N$  is a normalisation factor, while  $\mathcal{R}_q$  is defined as

$$\mathcal{R}_q(\mathcal{E}) = \mathcal{R}_{\text{lc}}(\mathcal{E}) \exp\left(-\sqrt[4]{q^2 + q^4}\right). \quad (15)$$

In this definition,  $\mathcal{R}_{\text{lc}}$  denotes the value of  $\mathcal{R}$  at the edge of the loss cone, while  $q$  is the loss cone diffusivity, defined as the ratio between the expected relative change in  $\mathcal{R}^2$  due to two-body relaxation at the edge of the loss cone:

$$q(\mathcal{E}) = \left. \frac{\langle \Delta \mathcal{R}^2 \rangle}{\mathcal{R}^2} \right|_{\mathcal{R}=\mathcal{R}_{\text{lc}}}. \quad (16)$$

The  $q$  parameter distinguishes between the full ( $q \gg 1$ ) and the empty ( $q \ll 1$ ) loss cone regimes. In the full loss cone regime, two-body relaxation is particularly efficient in scattering angular momentum, with variations as large as the angular momentum itself. Therefore, objects instantaneously located on an orbit penetrating the loss cone are likely to be scattered away from it before reaching the pericentre, and escape the capture by the MBH. On the other hand, in the empty loss cone regime, once a body enters the loss cone it has very little chance of escaping it, as two-body relaxation induces very small kicks in angular momentum over a single orbit (Cohn & Kulsrud 1978; Merritt 2013a).

## 2.2. Gravitational wave emission and loss cone captures

The dynamics of a stellar-mass BHs orbiting an MBH in a nuclear cluster is defined by two fundamental timescales:

1.  $t_{\text{rlx}}$ , the average time needed for two-body encounters to significantly change the orbital parameters;
2.  $t_{\text{GW}}$ , the average time needed for the emission of GWs to significantly change the orbital parameters.

When the stellar-mass BH and the MBH are far enough, the emission of GWs is negligible. In this configuration, the orbit evolution is stochastic and results only from the effect of two-body encounters. These encounters shuffle the position of the orbit on the  $(1-e, a)$  plane and are particularly efficient in changing  $1-e$ . Observing that  $\mathcal{R} = 1-e^2$  for Keplerian orbits, we can then define the average time needed for two-body relaxation to significantly change  $1-e$  as

$$t_{\text{rlx}} = \frac{1-e}{D_o[\Delta(1-e)]} \simeq \frac{1-e^2}{D_o[\Delta(1-e^2)]} = \frac{\mathcal{R}}{D_o[\Delta\mathcal{R}]}, \quad (17)$$

where we use the fact that  $1-e^2$  goes like  $2(1-e)$  for  $e \rightarrow 1$ . The expression we used for  $D_o[\Delta\mathcal{R}]$  can be found in Appendix A and in Broggi et al. (2022).

During this random motion on the  $(1-e, a)$  plane, the orbit might end up in a situation where the stellar-mass BH passes so close to the MBH that the emission of GWs begins to contribute to the evolution of the orbit, pushing it to become tighter and less eccentric, as the system loses energy and angular momentum. We can measure the strength of this effect by looking at the  $t_{\text{GW}}$  timescale, which we define as

$$t_{\text{GW}} = (1-e) \left\langle \frac{d}{dt} (1-e) \right\rangle_{\text{GW}}^{-1} = -(1-e) \left\langle \frac{de}{dt} \right\rangle_{\text{GW}}^{-1}, \quad (18)$$

where angle brackets indicate orbit-averages. Peters (1964) showed the following:

$$\left\langle \frac{de}{dt} \right\rangle_{\text{GW}} = -\frac{304 G^3 M_\bullet m_\bullet (M_\bullet + m_\bullet)}{15 c^5 a^4} \frac{e}{(1-e^2)^{5/2}} \left(1 + \frac{121}{304} e^2\right). \quad (19)$$

Thus<sup>3</sup>,

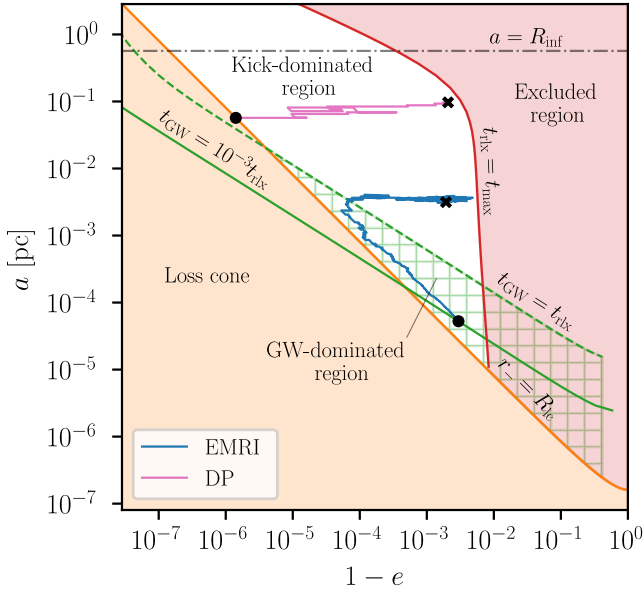
$$t_{\text{GW}} = \frac{15 c^5 a^4}{304 G^3 M_\bullet m_\bullet (M_\bullet + m_\bullet)} \frac{(1-e)(1-e^2)^{5/2}}{e} \left(1 + \frac{121}{304} e^2\right)^{-1} \\ \xrightarrow{e \rightarrow 1} \frac{12 \sqrt{2}}{85} \frac{c^5 a^4}{G^3 M_\bullet m_\bullet (M_\bullet + m_\bullet)} (1-e)^{7/2}. \quad (20)$$

At this point, we can define two regions in the  $(1-e, a)$  plane (see Fig. 1):

1. the kick-dominated region, where  $t_{\text{rlx}} < t_{\text{GW}}$  and the orbit evolution is stochastic and dominated by two-body encounters;
2. the GW-dominated region, where  $t_{\text{GW}} < t_{\text{rlx}}$  and the orbit evolution is quasi-deterministic and dominated by the emission of GWs.

Since strong velocity kicks might still push the orbit back into the kick-dominated region, or directly inside the loss cone

<sup>3</sup> We point out that the same expression of  $t_{\text{GW}}$  for  $e \rightarrow 1$  can be found by defining  $t_{\text{GW}} = |a| \langle da/dt \rangle_{\text{GW}}^{-1}$ .



**Fig. 1.** Two examples of orbit evolution in the  $(1 - e, a)$  plane. The blue line shows the formation of an EMRI, while the pink one results in a DP. Both are mainly randomised in eccentricity while inside the white kick-dominated region. When the blue line crosses into the GW-dominated region, which is filled with a green grid, it predominantly evolves via GW emission. DPs are identified by the crossing of the loss cone edge, which is represented by the solid orange curve. EMRIs instead must reach the portion of the solid green curve in between the loss cone and the excluded region (see Sect. 3.4). Here we set  $M_{\bullet} = 3 \times 10^5 M_{\odot}$ .

after few orbital periods<sup>4</sup>, we caution against identifying the EMRI formation region with the GW-dominated region. Following Hopman & Alexander (2005), a safest choice is that of labelling as EMRIs the orbits that reach the condition  $t_{\text{GW}} = 10^{-3} t_{\text{rlx}}$ . Here the orbital elements are so deep into the GW-dominated region that we can be confident that a merger will eventually take place after several thousands of periods. We instead define DPs as those events where the orbit enters the loss cone before reaching the  $t_{\text{GW}} = 10^{-3} t_{\text{rlx}}$  condition.

### 2.3. Loss cone definition in post-Newtonian gravity

From GR, we know that a test particle on a stable orbit of eccentricity  $e$  must always keep a distance from an MBH of at least  $(6 + 2e)/(1 + e)$  gravitational radii in order to avoid plunging into it (Cutler et al. 1994). Consistently, the innermost stable circular ( $e = 0$ ) orbit around an MBH has radius  $6R_g$ , while a body on a parabolic ( $e = 1$ ) trajectory can approach the MBH up to  $4R_g$ , still managing to escape it (Misner et al. 1973). Therefore,  $4R_g$  can be taken as the loss cone radius for very eccentric orbits in GR.

Relaxation in dense systems, however, is usually modelled with Newtonian gravity, which poses a problem for the loss cone definition. Crucially, there is no single method to translate Keplerian orbits of Newtonian gravity into geodesics of the Schwarzschild metric in GR (Servin & Kesden 2017). In particular, two different mappings can both recover the same

Newtonian ellipse for large pericentres. One mapping is that where both the general relativistic and Newtonian orbits share the same pericentre. This mapping has the property of conserving the circumference of circular orbits between the two theories (Servin & Kesden 2017), which is a property that does not appear to be much of use for our study, as we are interested in highly eccentric orbits. Instead, a more useful mapping is the one that preserves the angular momentum between the theories. In this case, we can identify plunging orbits around Schwarzschild MBHs with the same condition used in GR (Merritt et al. 2011; Sadeghian & Will 2011; Will 2012; Merritt 2013a):

$$J < \frac{4GM_{\bullet}}{c}. \quad (21)$$

From the Newtonian relation  $J^2 = GM_{\bullet}a(1 - e^2)$ , then follows the equivalent condition for the pericentre when  $e \rightarrow 1$ :

$$r_{-} < 8R_g. \quad (22)$$

To get our point across let us consider an example. Take a body moving in Newtonian gravity on a highly eccentric orbit, passing at slightly less than  $8R_g$  from the MBH at pericentre and evolve it back in time to the apocentre of its orbit. Given the high eccentricity we assume, the apocentre is large enough that GR simplifies to Newtonian gravity. If we now turn on GR and let the system evolve forward in time, the same body will reach a minimum distance from the MBH slightly lower than  $4R_g$ , plunging into its horizon. Thus, when simulating similar orbits in Newtonian dynamics, we have to consider as plunges all bodies that get to less than  $8R_g$  from the MBH.

However, for this work we neither solved the Einstein field equations to evolve our system nor relied on a Newtonian description of the dynamics. We evolved our systems using PN dynamics, which operates in an intermediate regime by definition. We are not aware of studies on the capture of eccentric orbit at the PN level, unlike in the case of circular orbits (e.g. Schäfer & Jaranowski 2024). Thus, we first test how the PN dynamics affects the position of the pericentre for a highly eccentric and deeply penetrating orbit within our code.

We took a test orbit initialised such that the stellar-mass BH reaches a distance from the MBH of  $8R_g$  at pericentre in Newtonian dynamics. Then, we ran the same initial conditions for three more times, adding a new PN term each time. We repeated this test for different combinations of masses and eccentricities, and found that only the latter have some small influence on the result. In particular, we tested for all combinations of  $e = \{0.9, 0.999, 0.99999\}$ ,  $m_{\bullet} = \{10, 100\} M_{\odot}$ , and  $M_{\bullet}/m_{\bullet} = \{10^3, 10^4, 10^5, 4 \times 10^5\}$ .

We find that, regardless of the choice of parameters, a highly eccentric orbit with  $r_{-} = 8R_g$  in Newtonian dynamics reaches a pericentre of  $\sim 6.45R_g$  at 1PN order and a pericentre of  $\sim 5.6R_g$  at 2PN and 2.5PN order. The results of this investigation are shown in Fig. 2.

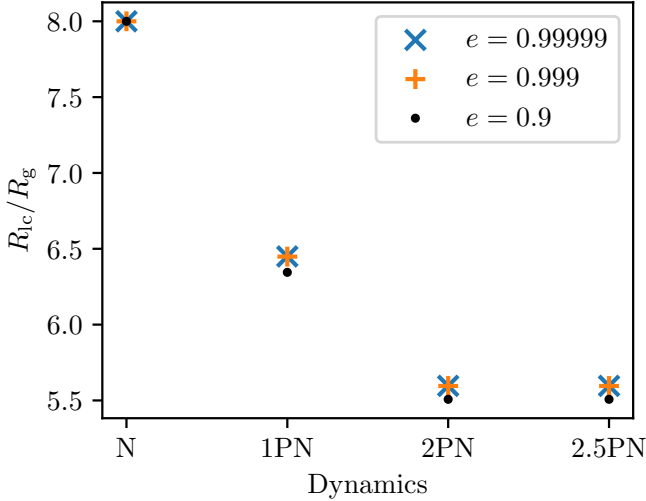
Thus, we define the loss cone radius as

$$R_{\text{lc}} = \zeta R_g, \quad (23)$$

with

$$\zeta = \begin{cases} 8 & \text{if Newtonian dynamics} \\ 6.45 & \text{if 1PN dynamics} \\ 5.6 & \text{if 2PN or 2.5PN dynamics} \end{cases}. \quad (24)$$

<sup>4</sup> The latter of these could be called EMRIs, as they spend some time in the GW-dominated region before merging. However, given the small number of cycles completed before the merger, the GW signal from these events will be too weak to be observed by LISA. We therefore classify them as DPs.



**Fig. 2.** Loss cone radius per unit of  $R_g$  as a function of  $e$  and dynamics used to evolve the system. We also tested for different  $M_\bullet$  and  $m_\bullet$  values, finding no discernible difference. On the x-axis, ‘N’ stands for Newtonian.

#### 2.4. EMRI-to-DP ratio and cliffhanger EMRIs

Through Monte Carlo simulations, [Hopman & Alexander \(2005\)](#) were the first to explore how the probability of EMRI versus DP depends on the initial semi-major axis of the orbit  $a_0$ .

The code they presented works in  $(\mathcal{E}, J)$  space and follows an object on a relativistic orbit, with small perturbations due to GW emission and stochastic two-body scattering. It accounts for GW radiation using ([Peters 1964](#))

$$\left\langle \frac{d\mathcal{E}}{dt} \right\rangle_{\text{GW}} = \frac{32 G^4 M_\bullet^2 m_\bullet (M_\bullet + m_\bullet)}{5 c^5 a^5 (1 - e^2)^{7/2}} \left( 1 + \frac{73}{24} e^2 + \frac{37}{96} e^4 \right) \quad (25)$$

and

$$\left\langle \frac{dJ}{dt} \right\rangle_{\text{GW}} = -\frac{32 G^{7/2} M_\bullet m_\bullet (M_\bullet + m_\bullet)^{3/2}}{5 c^5 a^{7/2} (1 - e^2)^2} \left( 1 + \frac{7}{8} e^2 \right), \quad (26)$$

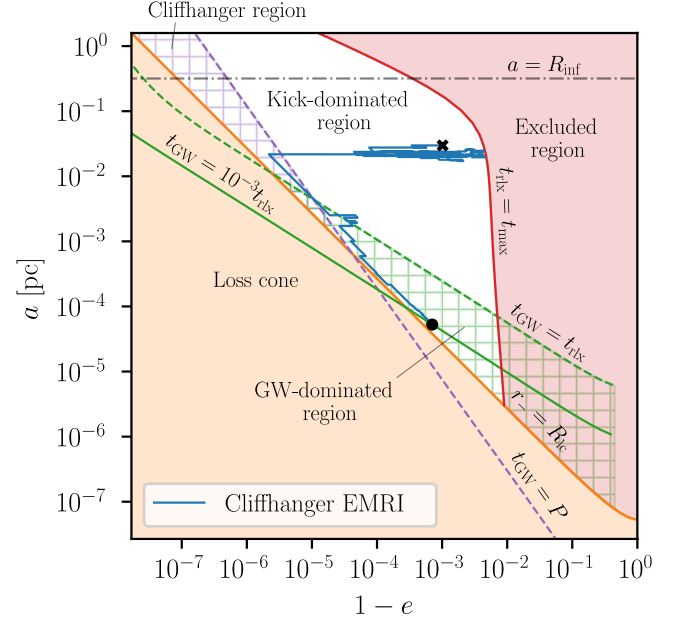
which are multiplied by  $P$  to compute an orbit-averaged estimate of the energy and angular momentum carried away by GWs during a period. Similarly, the code also updates the angular momentum once per orbit to account for two-body relaxation, by employing the orbit-averaged diffusion coefficients  $D_0[\Delta J]$  and  $D_0[(\Delta J)^2]$ . It instead neglects scattering in  $\mathcal{E}$ , as two-body relaxation is much more efficient in shuffling  $J$  rather than  $\mathcal{E}$  ([Binney & Tremaine 2008](#); [Merritt 2013a](#)).

After running many simulations for a central MBH of mass  $M_\bullet = 3 \times 10^6 M_\odot$  surrounded by stellar-mass BHs of mass  $m_\bullet = 10 M_\odot$ , [Hopman & Alexander \(2005\)](#) focused on the function

$$S(a_0) = \frac{N_{\text{EMRI}}(a_0)}{N_{\text{EMRI}}(a_0) + N_{\text{DP}}(a_0)}, \quad (27)$$

where  $N_{\text{EMRI}}$  and  $N_{\text{DP}}$  are respectively the counts of EMRIs and DPs formed given an initial  $a_0$ , and discovered the existence of a critical value  $a_c^{\text{HA}} \approx 5 \times 10^{-3} R_{\text{inf}}$  such that  $S(a_0) = 1$  for  $a_0 \ll a_c^{\text{HA}}$ , and  $S(a_0) = 0$  for  $a_0 \gg a_c^{\text{HA}}$ . [Raveh & Perets \(2021\)](#) fixed a sign error in the original implementation of [Hopman & Alexander \(2005\)](#), updating the result to  $a_c^{\text{RP}} \approx 2 \times 10^{-2} R_{\text{inf}}$ .

Recently, [QS24](#) studied again the problem, considering a more realistic set up in which the stellar-mass BH is scattered



**Fig. 3.** Example of a cliffhanger EMRI orbit in the  $(1 - e, a)$  plane. The cliffhanger region is filled with a purple grid, and it is delimited by the purple dashed line and the orange solid curve (see [Fig. 1](#) for the meaning of the other elements). The sharp tightening of the orbit at the passage inside the cliffhanger region occurs during a single pericentre passage. Here we set  $M_\bullet = 10^5 M_\odot$ .

by stars of mass  $m_\star = 1 M_\odot$  and scattering in energy is not neglected. Most notably, they extended the study for low-mass MBHs, finding that it is possible for EMRIs to form even from large  $a_0$ , thus breaking the classical notion of a somewhat sharp transition between  $a_0$  values from which only EMRIs form to those that result in DPs only.

[Qunbar & Stone \(2024\)](#) called EMRIs that form from initially wide orbits cliffhanger EMRIs. The name has to do with the suspenseful ending of the orbital evolution, leading to an EMRI in a region of the parameter space where only DPs were originally expected. In the classical picture, stellar-mass BHs on wide orbits generally end up plunging into the loss cone following a sequence of two-body encounters, which culminate in an almost radial orbit towards the MBH. However, a fraction of these orbits will not enter the capture sphere of the MBH, but graze its surface at a slightly larger distance. In this case, the burst of GW radiation emitted is so strong to easily both reduce  $a$  and increase  $1 - e$  by a factor of ten during a single passage (see [Fig. 3](#)). If the mass of the MBH is small and the semi-major axis large enough, following the large loss of energy and angular momentum, the stellar-mass BH will end up on a much tighter orbit, closer to or inside the GW-dominated region, with a high probability of later becoming an EMRI.

To better quantify how small the MBH mass should be and to better characterise the phenomenon, we here introduce the ‘cliffhanger region’: a region on the  $(1 - e, a)$  plane from which cliffhanger EMRIs come from. Given Eqs. (9) and (20), the semi-major axis  $a_{\text{cliff}}$  where  $t_{\text{GW}} = P$  as a function of the eccentricity is

$$a_{\text{cliff}} = \left( \frac{85\pi}{6\sqrt{2}} \right)^{2/5} \left( \frac{m_\bullet^2 (M_\bullet + m_\bullet)}{M_\bullet^3} \right)^{1/5} (1 - e)^{-7/5} R_g, \quad (28)$$

where we assume  $e \rightarrow 1$ . The loss cone boundary, from Eqs. (8) and (23), is instead identified by the equation  $r_- = R_{\text{lc}}$ ,

that is

$$a_{lc} = \frac{\zeta}{1-e} R_g. \quad (29)$$

The region in the  $(1-e, a)$  plane identified by the condition

$$a_{lc}(e) < a < a_{cliff}(e) \quad (30)$$

is what we refer to as the cliffhanger region (see Fig. 3). As the curves cross at eccentricity

$$\tilde{e} = 1 - \frac{85\pi}{6\sqrt{2}} \zeta^{-5/2} \left( \frac{m_\bullet^2 (M_\bullet + m_\bullet)}{M_\bullet^3} \right)^{1/2}, \quad (31)$$

the cliffhanger region is defined only for large eccentricities ( $e > \tilde{e}$ ).

Whenever a stellar-mass BH is scattered by two-body encounters onto an orbit which lays inside the cliffhanger region, the orbit will be drastically modified by GW radiation during a single pericentre passage, meaning the variation happens in less than an orbital period. The stellar-mass BH will not plunge onto the MBH after this strong emission. During this process the orbital parameters are related through (Peters 1964)

$$a(e) = c_0 \frac{e^{12/19}}{1-e^2} \left( 1 + \frac{121}{304} e^2 \right)^{870/2299}, \quad (32)$$

where  $c_0$  is determined by the initial conditions. For  $e \rightarrow 1$ , this implies that  $a(e) \propto (1-e)^{-1}$ . Therefore, GWs emission conserves the pericentre<sup>5</sup> for  $e \rightarrow 1$ , and in the  $(1-e, a)$  plane the orbit will evolve parallel to the loss cone edge.<sup>6</sup>

The scaling of the two curves with the mass of the central MBH is different: Eq. (29) implies that  $a_{lc} \propto M_\bullet$ , while Eq. (28) shows that  $a_{cliff} \propto M_\bullet^{3/5}$ . As we consider larger MBHs, the loss cone edge will cross the  $t_{GW} = P$  curve at larger eccentricities and semi-major axes, eventually to  $a > R_{inf}$  (see Fig. 4). As a result, it gets increasingly harder with larger values of  $M_\bullet$  for two-body relaxation to scatter orbits inside the cliffhanger region while avoiding DPs, as the corresponding range in  $e$  becomes extremely small. We can therefore predict the maximum MBH mass which allows the formation of cliffhanger EMRIs as in QS24. From Eqs. (9) and (25), follows that the change in energy in a period is

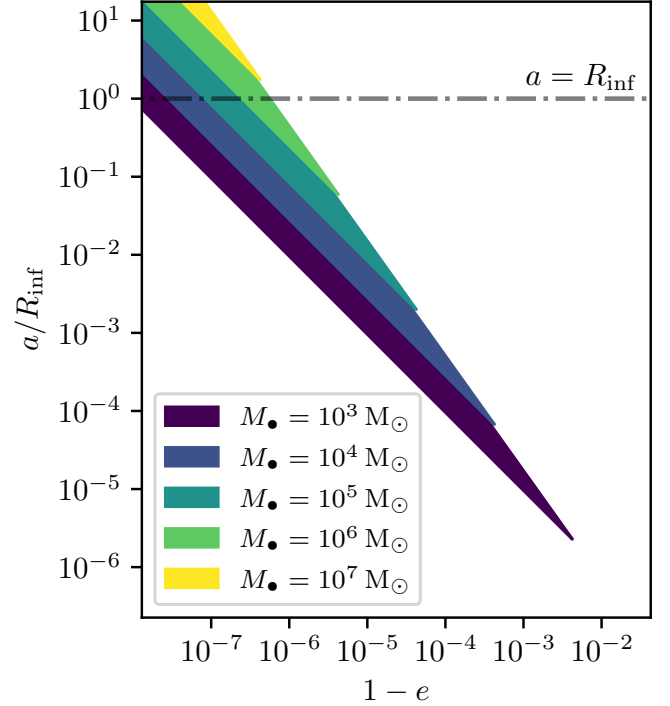
$$\begin{aligned} \Delta \mathcal{E}_{GW}(a, e) &= P \left\langle \frac{d\mathcal{E}}{dt} \right\rangle_{GW} \\ &\xrightarrow{e \rightarrow 1} \frac{85\pi}{12\sqrt{2}} \frac{G^{7/2} M_\bullet^2 m_\bullet (M_\bullet + m_\bullet)^{1/2}}{c^5} a^{-7/2} (1-e)^{-7/2}, \end{aligned} \quad (33)$$

which depends only on the pericentre in the  $e \rightarrow 1$  limit. Plugging Eq. (28) into Eq. (33) gives

$$\Delta \mathcal{E}_{GW}^{cliff}(a_{cliff}, e(a_{cliff})) = \frac{GM_\bullet}{2a_{cliff}} \quad (34)$$

<sup>5</sup> This also implies that  $J$  is approximately constant, as  $J \simeq \sqrt{2GM_\bullet r_-}$  for highly eccentric orbits.

<sup>6</sup> Hence the EMRI sort of ‘hangs on the cliff’ at the edge of the loss cone, re-proposing the pun at the origin of the title of the movie ‘Cliffhanger’ (1993).



**Fig. 4.** Cliffhanger region for different MBH masses in the  $(1-e, a)$  plane, where  $a$  is shown through its ratio with the influence radius. The darker regions are partly covered by lighter regions as all the regions extend up to the upper axis. All regions share a similar triangular shape, which is pushed towards the upper left corner of the plot as  $M_\bullet$  increases, effectively shrinking the available cliffhanger region for  $a < R_{inf}$ . For  $M_\bullet = 10^7 M_\odot$  the whole cliffhanger region is located outside the  $a = R_{inf}$  line.

which is the energy change in a period due to GWs emission on the  $t_{GW} = P$  curve. On this curve, GW radiation doubles  $\mathcal{E}$  during a single period,

$$\mathcal{E}_f^{cliff} = \mathcal{E}_i^{cliff} + \Delta \mathcal{E}_{GW}^{cliff}(a_{cliff,i}, e(a_{cliff,i})) = \frac{GM_\bullet}{a_{cliff,i}} = 2\mathcal{E}_i^{cliff}, \quad (35)$$

meaning that  $a$  is halved in the process.

Going deeper into the cliffhanger region allows larger energy jumps. At the limit of the loss cone edge,  $r_- = R_{lc}$  and Eq. (33) gives

$$\Delta \mathcal{E}_{GW}^{lc}(a_{lc}, e(a_{lc})) = \frac{85\pi}{12\sqrt{2}} \zeta^{-7/2} c^2 m_\bullet (M_\bullet + m_\bullet)^{1/2} M_\bullet^{-3/2}, \quad (36)$$

which does not depend on  $a_{lc}$  as we are setting the value of the pericentre. On the edge of the loss cone, the energy jump will be large, thus we can approximate

$$\mathcal{E}_f^{lc} = \mathcal{E}_i^{lc} + \Delta \mathcal{E}_{GW}^{lc} \simeq \Delta \mathcal{E}_{GW}^{lc}, \quad (37)$$

from which follows that all the cliffhanger jumps on the loss cone edge will end up with the same semi-major axis:

$$a_f^{lc} = \frac{6\sqrt{2}}{85\pi} \zeta^{7/2} \frac{G}{c^2} \frac{M_\bullet^{5/2}}{m_\bullet (M_\bullet + m_\bullet)^{1/2}} \simeq \frac{6\sqrt{2}}{85\pi} \zeta^{7/2} \frac{M_\bullet}{m_\bullet} R_g. \quad (38)$$

Finally, we can get a conservative estimate for the maximum MBH mass expected to produce cliffhanger EMRIs. The burst of GWs will end up forming an EMRI as long as the jump in  $a$  is inside or close to the EMRI formation region. Using the

classical estimate consistent with our results (see Sect. 4.1),  $a_f^{\text{lc}} \lesssim 10^{-2} R_{\text{inf}}$ , we obtain the following:

$$M_{\bullet} \lesssim \frac{17\pi}{120\sqrt{2}} \zeta^{-7/2} \frac{c^2}{\sigma^2} m_{\bullet}. \quad (39)$$

Given our choice of parameters for the  $M$ – $\sigma$  relation (Gültekin et al. 2009), we finally get

$$M_{\bullet} \lesssim 1.8 \times 10^7 \zeta^{-2.38} \left( \frac{m_{\bullet}}{10 M_{\odot}} \right)^{0.68} M_{\odot}. \quad (40)$$

It should be noted that this derivation takes into account only the properties of the MBH and the stellar-mass BHs, while the distributions of background bodies only marginally enters in the effective critical radius  $10^{-2} R_{\text{inf}}$ . In particular, it does not factor in how likely is for two-body relaxation to push the orbit inside the cliffhanger region, nor the fact that a cliffhanger EMRIs can be the result of more than one energy jump, as it is possible for the orbit to enter the cliffhanger region multiple times.

Plugging  $\zeta = 5.6$  and  $m_{\bullet} = 10 M_{\odot}$  into Eq. (40), we get  $M_{\bullet} \lesssim 3 \times 10^5 M_{\odot}$ , which is indeed roughly the MBH mass for which we start seeing cliffhanger EMRIs forming (see Sect. 4.1). For  $\zeta = 8$  and  $m_{\bullet} = 10 M_{\odot}$ , we find  $M_{\bullet} \lesssim 1.3 \times 10^5 M_{\odot}$ , which is remarkably close to the  $10^5 M_{\odot}$  value obtained by QS24, despite our different assumptions.

### 3. Numerical simulations of EMRI and DP formation

In this section we illustrate the features of our numerical implementation. We begin with an overview of the integration of the Hamiltonian evolution, then we explain how our treatment of local and orbit-averaged two-body relaxation works. Finally, we detail our procedure to generate the initial conditions for the runs we perform, together with the stopping criteria we use to identify the formation of EMRIs and DPs.

#### 3.1. Numerical overview and setup

In this work we employ an extended version of the code presented in Bonetti et al. (2016), a PN few-body integrator with adaptive step size, which we use to study the evolution of a binary system comprised of an MBH and a stellar-mass BH. The integrator leverages on a C++ implementation of the Bulirsch-Stoer algorithm (Bulirsch & Stoer 1966) coupled with Richardson extrapolation (Richardson 1911) to increase the accuracy of the numerical solution. The adaptive time step allows us to get smaller steps when the dynamics is faster, such as at the pericentre of the orbit.

The code integrates the equation of motion of two-body systems in the Hamiltonian form:

$$H = H_0 + \frac{1}{c^2} H_1 + \frac{1}{c^4} H_2 + \frac{1}{c^5} H_{2.5} + O\left(\frac{1}{c^6}\right). \quad (41)$$

Here,  $H_0$  is the non-relativistic Newtonian Hamiltonian, while  $H_1$ ,  $H_2$  and  $H_{2.5}$  are corrective terms introducing purely relativistic effects. Complete expressions for these terms, matching the form implemented numerically, can be found for example in Bonetti et al. (2016)<sup>7</sup>.  $H_{2.5}$  is the first non-conservative term

<sup>7</sup> Bonetti et al. (2016) focus on the three-body problem, and therefore report three-body PN Hamiltonian terms. We consistently use the same Hamiltonian, and reduce the evolution to a two-body problem by placing a third body with negligible mass at very large distance from the inner binary, effectively isolating its dynamics.

and accounts for GW radiation (Jaranowski & Schäfer 1997; Blanchet 2014). We only consider Schwarzschild BHs, thus we ignore the 0.5PN and 1.5PN terms of the expansion, as they account for spin and spin-orbit effects (Schäfer & Jaranowski 2024).

We use the Hamiltonian  $H$  to evolve the position  $\mathbf{x}$  and the momentum  $\mathbf{p}$  of the stellar-mass BH by integrating its Hamilton's equations in time  $t$ :

$$\begin{cases} \frac{d\mathbf{x}}{dt} = \sum_n \frac{1}{c^{2n}} \frac{\partial H_n}{\partial \mathbf{p}} \\ \frac{d\mathbf{p}}{dt} = - \sum_n \frac{1}{c^{2n}} \frac{\partial H_n}{\partial \mathbf{x}} \end{cases}. \quad (42)$$

At the same time we forcefully keep the MBH fixed and still at the origin of our Cartesian coordinate system. This choice is motivated by the fact that in a nuclear cluster the MBH will simultaneously interact with virtually all the objects in the stellar distribution, which will only cause a small Brownian motion-like displacement from the centre.

We ignore such a stochastic displacement of the MBH, even if it might be relevant as the number of stellar objects and the mass of the MBH gets lower (Merritt 2013a). Together with the evolution of the binary system, the code can account for the presence of additional external gravitational potentials, possibly representing a distribution of background bodies. We note again that only a single stellar-mass BH is evolved by directly solving Hamilton's equations.

In this work, we extend the code to include the effects of two-body relaxation on the orbit of the stellar-mass BH. These are due to the gravitational interaction of the stellar-mass BH with the rest of the stellar distribution. These stochastic interactions are fully characterised by the diffusion coefficients of the velocity component computed from their distribution function (Shapiro & Marchant 1978; Cohn & Kulsrud 1978; Merritt 2013a). We include the effects of two-body relaxation via a Monte Carlo approach by repeating the following procedure at the end of each time step of the integration:

1. we compute the mean and the standard deviation of the expected variation  $\langle \Delta \mathbf{v} \rangle$  of the velocity  $\mathbf{v}$  of the stellar-mass BH, considering the two-body encounters with the scattering objects which happened during the last time step;
2. we randomly draw a velocity change  $\Delta \mathbf{v}$  from the appropriate Gaussian distribution, which we build according to the previous step;
3. before moving to the next time step, we update the momentum of the stellar-mass BH to  $\mathbf{p}' = \mathbf{p} + m_{\bullet} \Delta \mathbf{v}$ , where  $m_{\bullet}$  is the mass of the stellar-mass BH<sup>8</sup>.

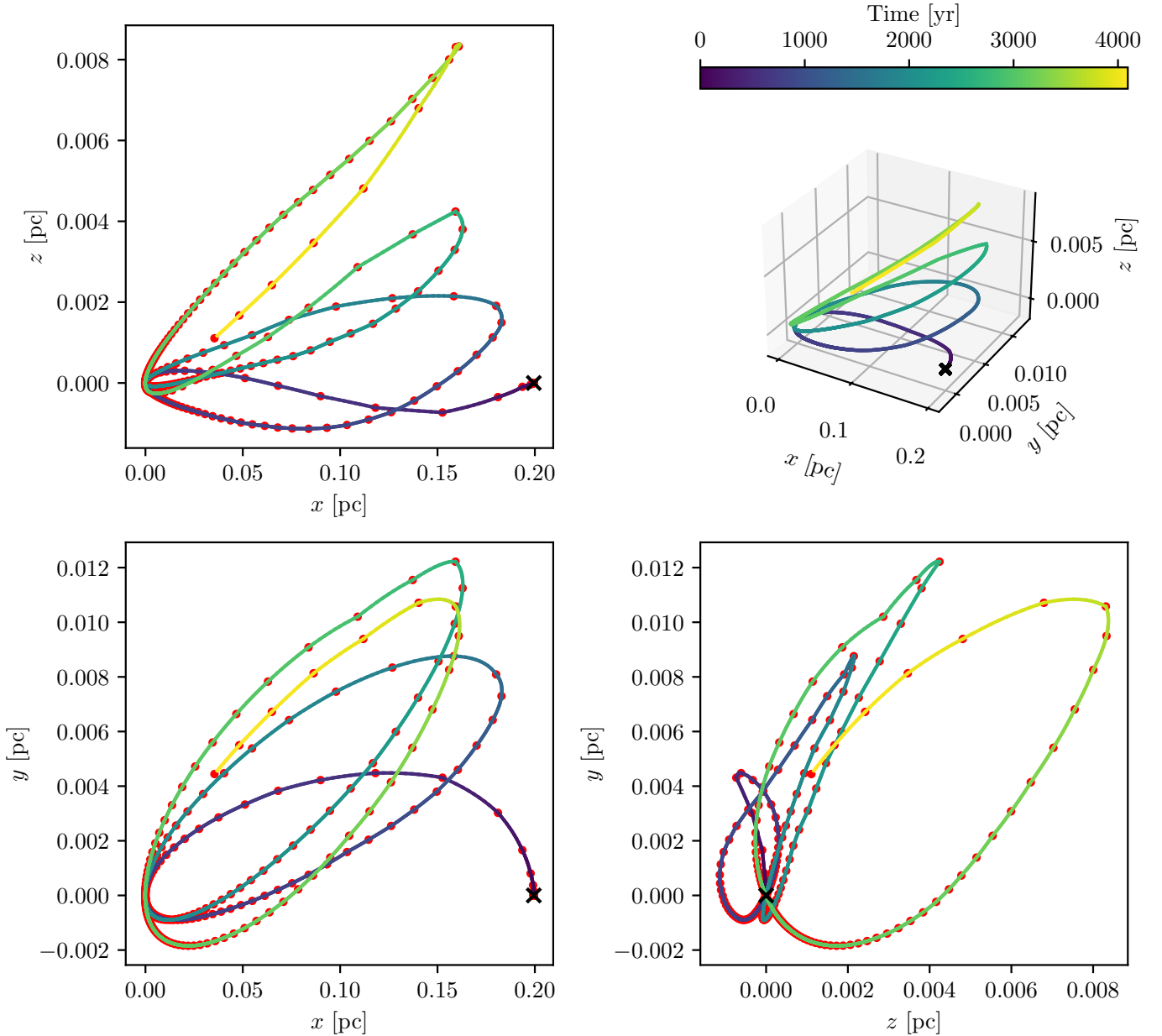
Following this procedure, we can statistically account for two-body encounters locally and multiple times within a single orbit, without leveraging on orbit-by-orbit approaches (QS24) or orbit-averages of the diffusion coefficients (Zhang & Amaro Seoane 2024).

For our simulations, we consider a stellar-mass BH of mass  $m_{\bullet} = 10 M_{\odot}$  and explore five possible masses for the MBH:  $M_{\bullet} = \{10^4, 10^5, 3 \times 10^5, 10^6, 4 \times 10^6\} M_{\odot}$ . We also group the background objects into two populations:

1. stars with  $m_{\star} = 1 M_{\odot}$ ,  $M_{\text{tot},\star} = 20 M_{\bullet}$ ,  $r_{a,\star} = 4 R_{\text{inf}}$  and  $\gamma_{\star} = 1.5$ ;
2. stellar-mass BHs with  $m_{\bullet} = 10 M_{\odot}$ ,  $M_{\text{tot},\bullet} = 0.2 M_{\bullet}$ ,  $r_{a,\bullet} = 4 R_{\text{inf}}$  and  $\gamma_{\bullet} = 1.7$ .

<sup>8</sup> Here we use the relation  $\mathbf{p} = m\mathbf{v}$ , which is not valid at PN order. Thus, while the evolution of the system is PN, the small velocity kicks we give to the stellar-mass BH at the end of each time step are Newtonian.



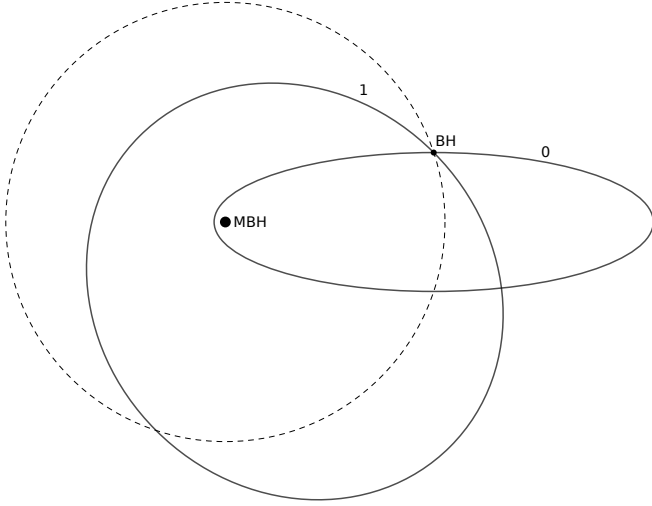


**Fig. 5.** Evolution of the position of a stellar-mass BH over few passages of the apocentre, using our local treatment of two-body relaxation. The top right panel shows the 3D trajectory of the stellar-mass BH, while the other three panels show its projections onto the three planes of the reference system. The colour of the curve shows the time elapsed since the beginning of the evolution. The red dots indicate the points where a velocity kick is given and they coincide with the time step of the main integrator. The points between the red dots are interpolated. The black cross shows the initial position of the stellar-mass BH, while the MBH is fixed at  $(0, 0, 0)$ . We note that the  $y$ - and  $z$ -axes have different scales for better readability. In this example  $M_{\bullet} = 4 \times 10^6 M_{\odot}$ ,  $a_0 = 0.1$  pc,  $e_0 = 0.999$ .

This choice corresponds to the two-components Bahcall-Wolf solution commonly used in literature (Bahcall & Wolf 1977,  $\gamma \approx 1.5$  for the lightest component,  $\gamma \approx 1.75$  for the heaviest component), where we slightly reduced the slope of the heaviest component to better represent their distribution in the inner regions of the stellar cluster, where they dominate relaxation (see Fig. 4 of Broggi et al. 2022, for the quasi-relaxed distribution of the stellar-mass BHs). The scale radius of the distribution  $r_a$  is set so that the stellar cluster in which our  $4 \times 10^6 M_{\odot}$  MBH is immersed is consistent with observational estimates of the density at the influence radius of SgrA\* (e.g. Schödel et al. 2007).

### 3.2. Local treatment of two-body relaxation

As explained in Appendix A, in order to compute local diffusion coefficients we have to first find  $f_{\alpha}(\mathcal{E})$  given  $n_{\alpha}$  and then integrate it over  $\mathcal{E}$ . This process involves nested integrals and derivatives, making it impractical to numerically perform such computations on the fly at each time step of our simulations. Thus, we precompute the auxiliary functions  $\mathcal{I}_{k,\alpha}$  and  $\mathcal{J}_{k,\alpha}$ , that are directly related to the diffusion coefficients, for some logarithmically spaced values of  $\mathcal{E}$  and  $\psi$ , and then, during the execution, we bilinearly interpolate these values of  $\mathcal{I}_{k,\alpha}$  and  $\mathcal{J}_{k,\alpha}$ . In this way, we get approximate local values for those coefficients



**Fig. 6.** Sketch depicting an example orbit change following our orbit-averaged procedure. The stellar-mass BH, initially on the orbit labelled 0, is kicked when  $r = a$ , and ends on the orbit labelled 1. Both orbits have the same semi-major axis  $a$ , but different eccentricities. The dashed circle represents a circular orbit with radius  $a$ . In this example there is no way to rotate either orbit around the focus occupied by the MBH such that the pericentre or apocentre of orbit 0 crosses orbit 1 at any point. Thus, neither pericentres nor apocentres are good spots to employ our orbit-averaged procedure. We note that the eccentricities of these example orbits are much lower compared to the ones investigated in this work. Moreover, the difference between subsequent orbits is much exaggerated here.

given the exact position and velocity of the stellar-mass BH at a particular time step.

The diffusion coefficients in velocity are then directly related to the parameters of the Gaussian distributions characterising  $\Delta v_{\parallel}$  and a  $\Delta v_{\perp}$  at each time step of the integration:

$$\Delta v_{\parallel} = \mu_{\parallel} + s_{\parallel}\sigma_{\parallel}, \quad \Delta v_{\perp} = \mu_{\perp} + s_{\perp}\sigma_{\perp}. \quad (43)$$

Here,  $s_{\parallel}$  and  $s_{\perp}$  are two uncorrelated random values, each extracted from a Normal distribution, while

$$\begin{aligned} \mu_{\parallel} &= D[\Delta v_{\parallel}] \Delta t, & \sigma_{\parallel}^2 &= D[(\Delta v_{\parallel})^2] \Delta t - (D[\Delta v_{\parallel}] \Delta t)^2, \\ \mu_{\perp} &= 0, & \sigma_{\perp}^2 &= D[(\Delta v_{\perp})^2] \Delta t. \end{aligned} \quad (44)$$

In these equations,  $\Delta t$  is the length of the particular time step we are considering. Thus, the values of  $\Delta v_{\parallel}$  and  $\Delta v_{\perp}$  we draw represent a statistical realisation of the change in velocity due to all of the encounters that have occurred during that time step.

Once  $\Delta v_{\perp}$  is drawn, we set its direction randomly, as it is distributed uniformly in the plane perpendicular to the vector  $\mathbf{v}$ . Finally, we vectorially sum together  $\Delta v_{\parallel}$  and  $\Delta v_{\perp}$  to get the total variation  $\Delta \mathbf{v}$  of the velocity of the stellar-mass BH, and update its momentum to  $\mathbf{p}' = \mathbf{p} + m_{\bullet} \Delta \mathbf{v}$  before the next step of the integration of Hamilton's equations begins. Figure 5 shows how this procedure affects a few consecutive apocentre passages. Deviations from regular elliptical orbits are evident, albeit small. Local velocity kicks are more frequent near pericentre due to the adaptive nature of the time step used by the main integrator.

### 3.3. Orbit-averaged treatment of two-body relaxation

The usual approach to implement two-body relaxation into the orbit evolution differs from what we have just described. Codes

usually work in the  $(\mathcal{E}, J)$  space and thus employ orbit-averaged diffusion coefficients of  $\mathcal{E}$  and  $J$ , which can be computed from those of velocity which we present in Appendix A (see Binney & Tremaine 2008; Merritt 2013a). In this work, we are interested in testing the reliability of the orbit-averaging approximation in opposition to our local approach. To make a meaningful comparison between the two methods, we present some simulations using the orbit-averaged approximation in Sect. 4.3.

Including orbit-averaging in our code is not straightforward, as we do not work with  $\mathcal{E}$  and  $J$ . Here we summarise our implementation:

1. at each time step of the integration, we draw  $\delta v_{\parallel}$  and  $\delta v_{\perp}$ <sup>9</sup> as explained in Sect. 3.2;
2. from  $\delta v_{\parallel}$  and  $\delta v_{\perp}$ , we compute the corresponding change in energy  $\delta \mathcal{E}$  and angular momentum  $\delta J$ ;
3. we move to the next time step without applying the kick and let the integration proceed, keeping track of the cumulative change in energy  $\Delta \mathcal{E} = \sum_i \delta \mathcal{E}_i$  and angular momentum  $\Delta J = \sum_i \delta J_i$ ;
4. once per orbit, from  $\Delta \mathcal{E}$  and  $\Delta J$  we compute a compatible kick  $\Delta \mathbf{v}$  and update  $\mathbf{p}$  to  $\mathbf{p}' = \mathbf{p} + m_{\bullet} \Delta \mathbf{v}$ ;
5. once the velocity of the stellar-mass BH is changed, we reset  $\Delta \mathcal{E}$  and  $\Delta J$  to zero, and repeat steps 1–4.

Following this procedure, we include the effects of stochastic kicks on the orbit once per period. The changes in  $\mathcal{E}$  and  $J$  are naturally orbit-integrated as the result of this procedure: the  $\delta \mathcal{E}_i$  and  $\delta J_i$  terms we sum together are indeed intrinsically weighted by the length of the time steps  $\delta t_i$ , which contribute through the  $\delta v_{\parallel,i}$  and  $\delta v_{\perp,i}$  terms.

In Appendix B we show that, once  $\delta v_{\parallel}$  and  $\delta v_{\perp}$  are drawn, then

$$\delta \mathcal{E} = -v \delta v_{\parallel} - \frac{1}{2} (\delta v_{\parallel})^2 - \frac{1}{2} (\delta v_{\perp})^2, \quad (45)$$

$$\frac{\delta J}{J} = \frac{\delta v_{\parallel}}{v} + \delta v_{\perp} \frac{v_r}{v v_t} \sin \theta + \frac{1}{2} \left( \frac{\delta v_{\perp} \cos \theta}{v_t} \right)^2 + \mathcal{O}((\delta v)^3). \quad (46)$$

Here  $v_t = J/r$  is the tangential velocity,  $v_r = \sqrt{v^2 - v_t^2}$  is the radial velocity, and  $\theta$  is an angle uniformly drawn from the interval  $[0, 2\pi)$ . We compute  $J$  as

$$J = |\mathbf{J}| = \frac{\mathbf{x} \times \mathbf{p}}{m_{\bullet}}. \quad (47)$$

In Appendix C we show that a suitable orbit-integrated kick in velocity can be computed from  $\Delta \mathcal{E}$  and  $\Delta J$  as

$$\Delta v_r = \sqrt{v^2 - 2\Delta \mathcal{E} - \left( \frac{J + \Delta J}{r} \right)^2} - v_r, \quad (48)$$

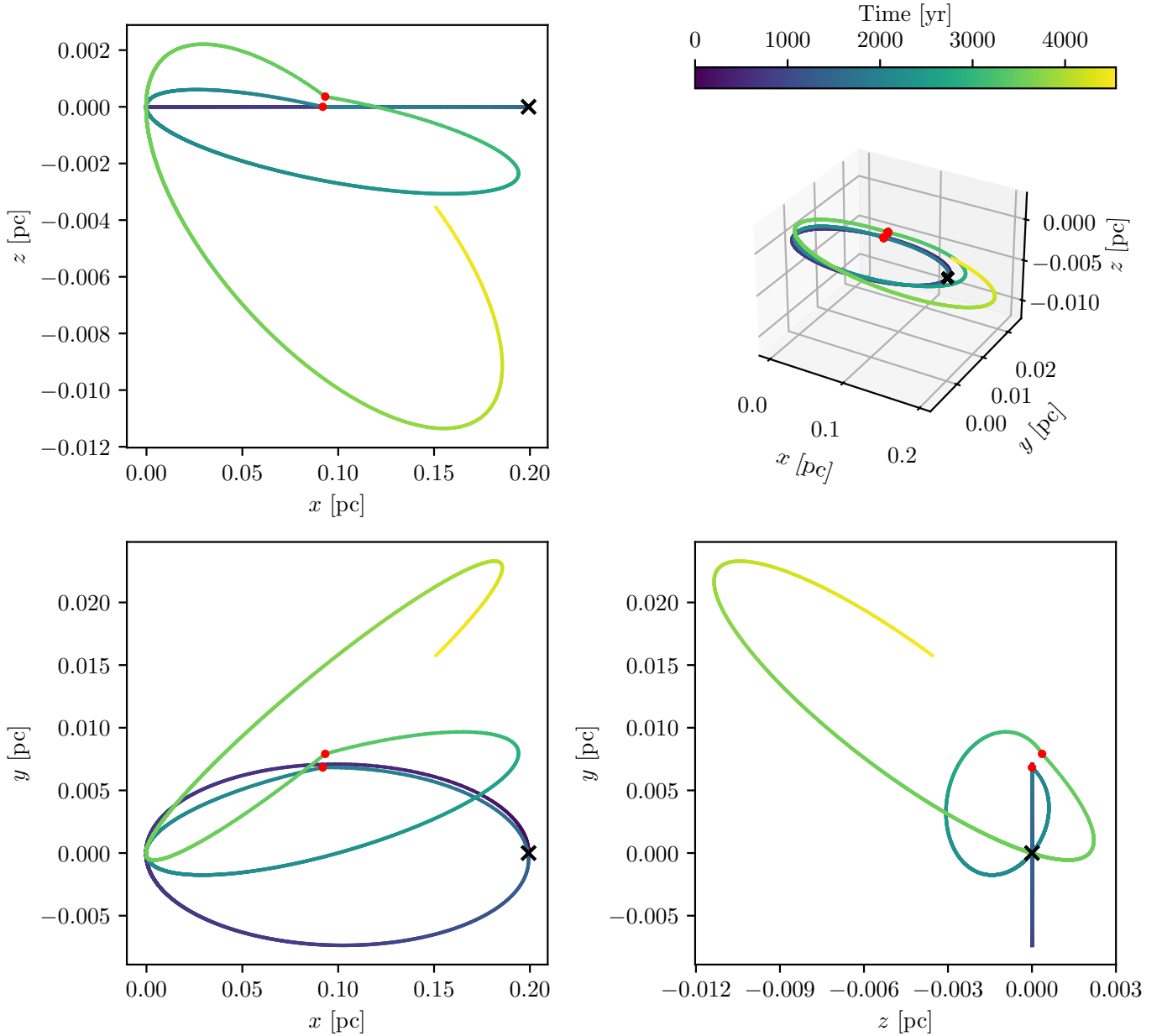
$$\Delta v_t = \frac{(J + \Delta J) \cos \beta}{r} - v_t, \quad (49)$$

$$\Delta v_3 = \frac{(J + \Delta J) \sin \beta}{r}. \quad (50)$$

Here the subscript 3 indicates the direction perpendicular to the orbital plane (thus,  $v_3 = 0$  before each kick), while  $\beta$  is an angle uniformly drawn from the interval  $[0, 2\pi)$ .

When updating the orbit, we change the velocity of the stellar-mass BH without displacing it. This is possible only if the old and the new orbit intersect at the location where  $\Delta \mathbf{v}$  is

<sup>9</sup> Here we introduce a new symbol:  $\delta \cdot \Delta \mathbf{v}$  is the kick imparted to the stellar-mass BH, while  $\delta v$  is the randomly drawn change in velocity over the intermediate steps along one orbit.



**Fig. 7.** Evolution of the position of a stellar-mass BH over a few orbits with the orbit-averaged treatment of two-body relaxation. The top right panel shows the 3D trajectory of the stellar-mass BH, while the other three panels show its projections onto the three planes of the reference system. The colour of the curve shows the time elapsed since the beginning of the evolution. The red dots indicate the points where a velocity kick is given, close to  $r = a$ . We let the body complete the first orbit without applying kicks to visualise its trajectory better. The black cross shows the initial position of the stellar-mass BH, while the MBH is fixed at  $(0, 0, 0)$ . We note that the  $y$ - and  $z$ -axes have different scales for better readability. In this example  $M_{\bullet} = 4 \times 10^6 M_{\odot}$ ,  $a_0 = 0.1$  pc,  $e_0 = 0.997$ .

applied, a fact that we ensure by applying the kick when  $r \simeq a$  (Fig. 6 illustrates the situation with a sketch). It is safe to assume that the orbits intersect there, as long as the energy of the particle (and thus  $a$ , since  $\mathcal{E} \simeq GM_{\bullet}/2a$ ) does not change drastically along a single period<sup>10</sup>. In case this assumption is not verified, the orbit-averaging approximation is not valid any more, as the time needed for two-body relaxation to significantly change  $\mathcal{E}$  is less than the orbital period (Binney & Tremaine 2008; Merritt 2013a).

<sup>10</sup> We note that the  $\Delta\mathcal{E}$  and  $\Delta J$  we are using here only come from the effects of two-body relaxation. GWs emission can strongly and quickly modify the energy of the orbit, but their effect is only included in the 2.5PN term of the Hamiltonian.

This motivates us to apply the kick  $\Delta\mathbf{v}$  to the stellar-mass BH once per orbit at the time step when  $r$  becomes smaller than  $a$ , coming from the apocentre. This choice reduces the number of cases where Eq. (48) has no real solutions to a negligible fraction, where orbit-averaging would fail because of the significant change received in one single orbit. In this case, the code simply halts, and we remove the run from the analysis.

The code computes  $a$  and  $e$  as

$$a = \frac{r_+ + r_-}{2}, \quad e = \frac{r_+ - r_-}{r_+ + r_-}, \quad (51)$$

each time the pericentre or apocentre of an orbit is reached, where these are identified as the distance from the centre when

the radial velocity flips sign. If  $a$  and  $e$  are computed at apocentre, then  $r_-$  is taken as the last pericentre passed. Instead, if they are computed at pericentre, then  $r_+$  is taken as the last apocentre passed. Figure 7 shows how this procedure affects a few consecutive orbits. In the figure, orbits are almost perfect ellipses, with small deviations due to PN effects.

### 3.4. Stopping criteria and initial conditions

The main aim of the code is that of identifying the formation of EMRIs and DPs. To this scope, we employ the definitions presented in Sect. 2.2, with some corrections needed given the computational cost of our hybrid few-body/Monte Carlo code.

As our aim was to gather a statistically significant number of EMRIs and DPs, we chose to limit the parameter space that we explored to a sub-portion close to the loss cone edge. Moreover, as we assume a steady-state stellar profile, we are interested in exploring orbital parameters for which the relaxation time  $t_{\text{rlx}}$  is not too large, as the initial conditions and the diffusion coefficients would be affected by the overall relaxation of the system. In particular, we consider as our timescale of investigation  $\sim 1\%$  of the time-to-peak of (classical) EMRI rates in the systems we are considering, which estimates the timescale of evolution of the distribution function of scatterers (Brogi et al. 2022)

$$t_{\text{max}} = 10 \left( \frac{M_\bullet}{4 \times 10^6 M_\odot} \right)^{1.29} \text{ Myr}, \quad (52)$$

and exclude from our investigations the region in the  $(1-e, a)$  phase space where  $t_{\text{rlx}} > t_{\text{max}}$ . The scaling with  $M_\bullet$  for  $t_{\text{max}}$  comes from the fact that (Merriitt 2013a)

$$t_{\text{rlx}} \propto M_\bullet^2 \sigma^{-2} \propto M_\bullet^{1.29}, \quad (53)$$

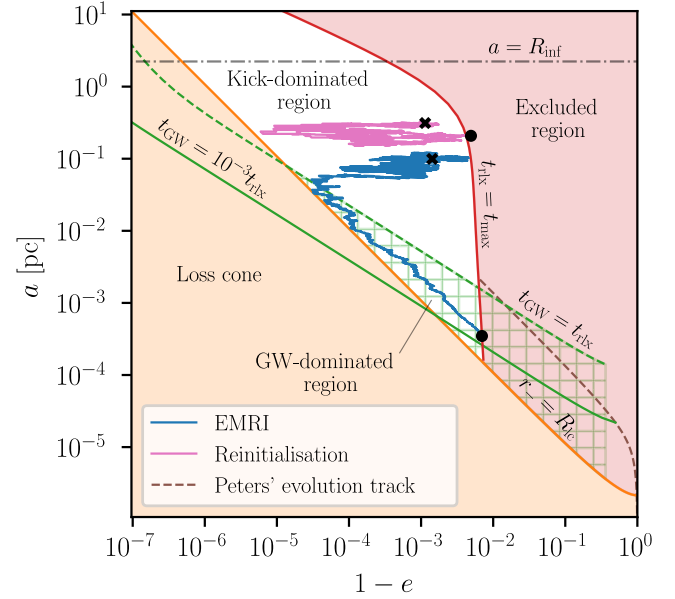
where we leverage on the  $M$ - $\sigma$  relation (Gültekin et al. 2009) to obtain the final scaling with the MBH mass. We note that  $t_{\text{max}}$  is not to be intended as a dynamical stopping condition (meaning the simulations do not halt at the time  $t = t_{\text{max}}$ ), rather the run is ended only when the orbital elements cross over the  $t_{\text{rlx}} = t_{\text{max}}$  curve on the  $(1-e, a)$  plane.

We initialise all our simulations inside the region of interest where  $t_{\text{rlx}} < t_{\text{max}}$  following a procedure described at the end of this section and we stop and discard any run where the orbit leaves the region. This choice does not affect the EMRI-to-DP ratio. In fact,  $t_{\text{max}}$  is so long with respect to the orbital period that any plunge or EMRI must enter this region before the capture. As long as the procedure is not biased towards EMRI or DP candidates, once a particle is discarded we are safe to reinitialise it as an uncorrelated particle.

The reinitialisation procedure becomes biased when the restriction to  $t_{\text{rlx}} < t_{\text{max}}$  excludes a significant part of the  $t_{\text{GW}} = 10^{-3} t_{\text{rlx}}$  curve. In this case, it can happen that some orbits well inside the GW emission dominated region but still far from the loss cone, can reach the  $t_{\text{rlx}} = t_{\text{max}}$  boundary before crossing the  $t_{\text{GW}} = 10^{-3} t_{\text{rlx}}$  separatrix, thus being erroneously stopped and reinitialised instead of being counted as EMRIs. A clear example of such an orbit is shown by the blue line in Fig. 8: the event will clearly produce an EMRI, but according to our reinitialisation criterion is effectively discarded and resimulated.

As DP candidates are not affected by this issue, the EMRI-to-DP ratio is influenced by these discarded EMRIs. This problem mostly affects the EMRI-to-DP ratio at small initial semi-major axes for large  $M_\bullet$ , as we show in Sect. 4.1.

For this reason, we correct our counts of EMRIs to also include orbits which cross into the excluded region while inside



**Fig. 8.** Two examples of orbit evolution in the  $(1-e, a)$  plane. The blue line shows the formation of an EMRI that is missed by our standard definition, and requires the correction described in the text to be rightfully counted. The pink line instead shows a regular reinitialisation. We also plot the evolution line of an orbit that passes through the minimum of the  $t_{\text{GW}} = 10^{-3} t_{\text{rlx}}$  curve following Eq. (32) (dashed brown line). It can be seen that all orbits that cross into the excluded region while being in the GW-dominated region fall below this evolution line, meaning they will reach the solid green curve, as long as the stochastic kicks due to two-body relaxation are small (see Fig. 1 for the meaning of the other elements). Here we set  $M_\bullet = 4 \times 10^6 M_\odot$ .

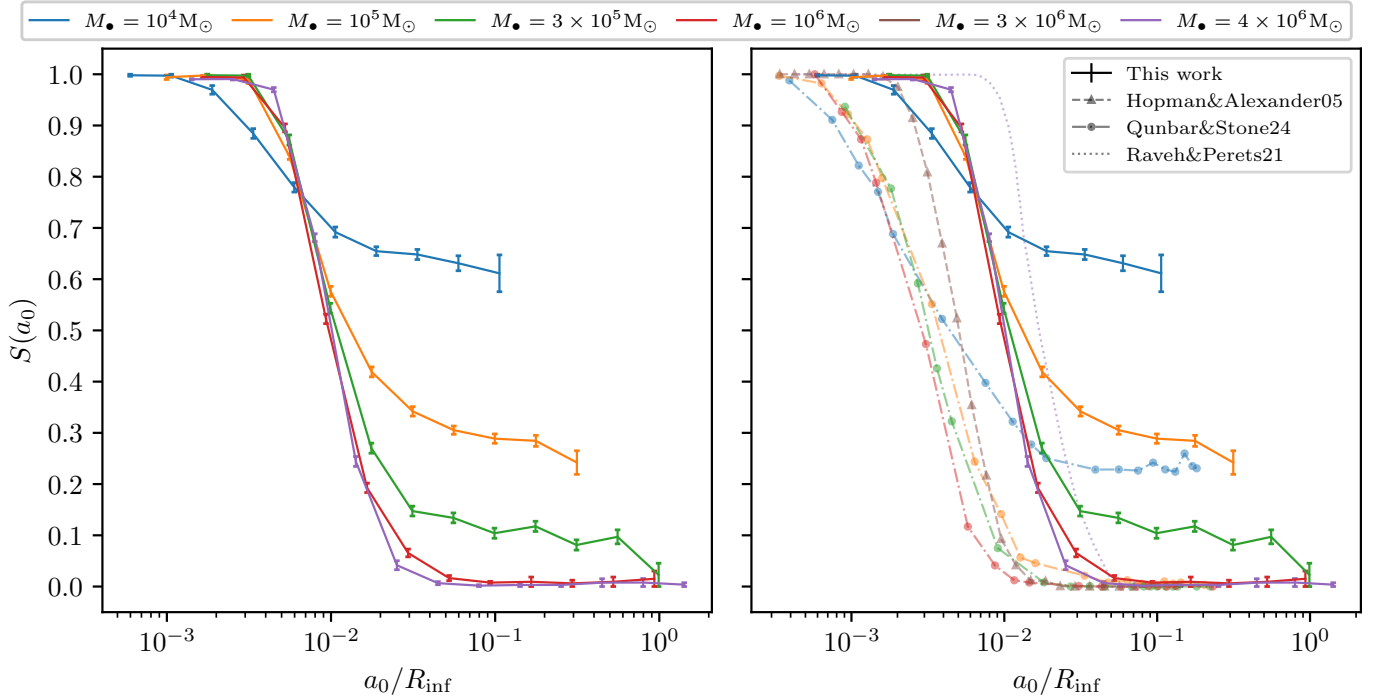
the GW-dominated region. Before accepting the orbit as an EMRI, we also verify that the final value of  $a$  is below the evolution line of an orbit which evolves as Eq. (32), where  $c_0$  is chosen so that the line crosses the minimum of the  $t_{\text{GW}} = 10^{-3} t_{\text{rlx}}$  curve (again, see Fig. 8). This means that, ignoring stochastic kicks due to two-body encounters, the orbit would eventually reach our condition for EMRI formation, while inside the excluded region. We also point out that near the  $t_{\text{rlx}} = t_{\text{max}}$  curve, the condition  $t_{\text{GW}} < t_{\text{rlx}}$  is likely sufficient to identify EMRIs anyway, as the stricter condition  $t_{\text{GW}} < 10^{-3} t_{\text{rlx}}$  is mainly introduced to avoid misclassifying DPs as EMRIs, and such a misclassification can only occur for  $1-e \lesssim 10^{-4}$ , where the  $t_{\text{GW}} = t_{\text{rlx}}$  curve closely approaches the loss cone.

To summarise, the possible outcomes of a simulation are:

1. a DP, if the separation  $r$  between MBH and stellar-mass BH satisfies the condition  $r < R_{\text{lc}}$ ;
2. an EMRI, either if (i) the condition  $t_{\text{GW}} = 10^{-3} t_{\text{rlx}}$  is satisfied, or (ii) the curve  $t_{\text{rlx}} = t_{\text{max}}$  is reached with a final  $a$  such that the curve  $t_{\text{GW}} = 10^{-3} t_{\text{rlx}}$  is eventually reached at a later time assuming only GW emission;
3. a reinitialisation, in all other cases in which the condition  $t_{\text{rlx}} = t_{\text{max}}$  is reached.

Condition (ii) for EMRI formation is the correction introduced above, and its importance will be shown in Sect. 4.1. Examples of each of these outcomes are shown in Figs. 1, 3, and 8.

Finally, the code halts if Eq. (48) has no real solutions when looking for an orbit-averaged kick, or if  $\mathcal{E} < 0$  at some point during the simulation (meaning the stellar-mass BH has become unbound). In practice, the former situation excluded a negligible number of simulations, and the latter never occurred.



**Fig. 9.** Function  $S(a_0)$  for different  $M_*$  scenarios. The bars display  $1\sigma$  uncertainty. The panel on the left displays our results only, while that on the right compares them with those from other works. We note that the other works make different assumptions, both with respect to our work and to each other.

For each nuclear cluster configuration, we select a set of possible initial semi-major axes  $a_0$  uniformly distributed in logarithm. Then, we adopt the steps described below to randomly draw a different realistic value for the initial eccentricity  $e_0$  of each simulation. Given  $a_0$ , we compute

$$\mathcal{R}_{\max} = \mathcal{R}|_{t_{\text{rlx}}=t_{\max}} \quad (54)$$

and

$$\mathcal{R}_{\min} = \min\{\mathcal{R}_{\text{lc}}, \mathcal{R}|_{t_{\text{GW}}=10^{-3}t_{\text{rlx}}}\} \quad (55)$$

at the corresponding energy  $\mathcal{E}_0 = GM_*/2a_0$ . Then, we randomly draw an initial value for  $\mathcal{R}_0$  from the probability distribution function

$$p(\mathcal{R}) = \frac{\log(\mathcal{R}/\mathcal{R}_q)}{\mathcal{R}_{\max}(\ln(\mathcal{R}_{\max}/\mathcal{R}_q) - 1) - \mathcal{R}_{\min}(\ln(\mathcal{R}_{\min}/\mathcal{R}_q) - 1)} \quad (56)$$

to reproduce the Cohn-Kulsrud distribution (Cohn & Kulsrud 1978, also see Sect. 2.1), where the normalisation factor is such that  $\int_{\mathcal{R}_{\min}}^{\mathcal{R}_{\max}} d\mathcal{R} p(\mathcal{R}) = 1$ . Finally, from  $\mathcal{R}_0$ , we compute

$$e_0 = \sqrt{1 - \mathcal{R}_0} \quad (57)$$

and we initialise the orbit with semi-major axis  $a_0$  and eccentricity  $e_0$  with Newtonian initial conditions, placing the stellar-mass BH at the apocentre.

In our procedure, we initialise all particles at the apocentre of their initial orbits, rather than assigning a random location along the orbit. We select this point because PN corrections are least significant there and this choice does not bias the likelihood of a particle becoming a DP or an EMRI. Before reaching a disruptive pericentre, the particle will either quickly modify its orbital parameters through two-body scattering (when  $t_{\text{rlx}} \lesssim P$ ), effectively erasing any memory of the initial condition, or complete more than one period on the same orbit (when  $t_{\text{rlx}} \gtrsim P$ ), rendering the specific initial location along the orbit irrelevant.

## 4. Results

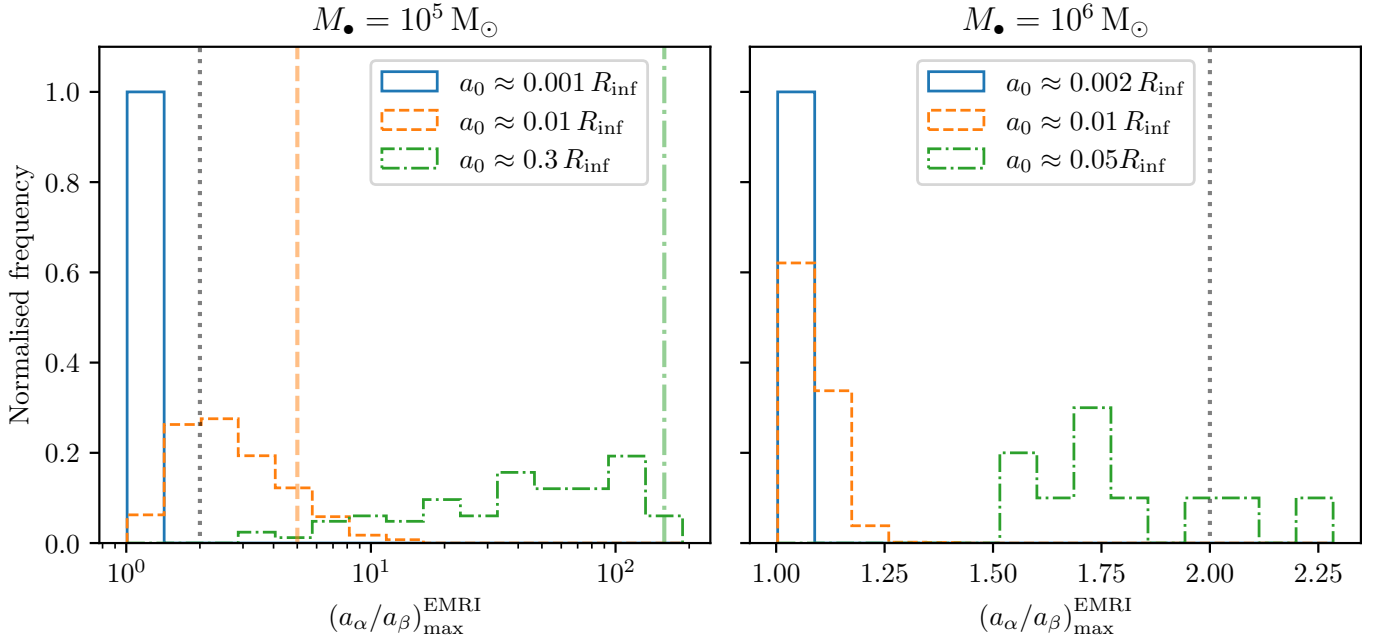
In this section we present the results we obtained running a large number of simulations with the code described in Sect. 3.

### 4.1. EMRI-to-DP ratio and cliffhanger EMRIs

For each simulated central MBH mass, we build the  $S(a_0)$  function given the counts of EMRIs and DPs. Given the success/failure nature of the problem (we can see EMRI formation as a success, and DP formation as a failure), we employ the binomial statistic to characterise the uncertainties on our findings. Specifically, we compute the corresponding binomial proportion confidence interval at  $1\sigma$  confidence level using the Wilson score interval (Wilson 1927), an improved version of the commonly used standard Wald interval. The Wilson score interval avoids issues when  $S(a_0)$  approaches 0 or 1, such as falsely implying certainty or error bars overshooting outside the  $[0, 1]$  range (Brown et al. 2001).

We report our main findings in the left panel of Fig. 9. For  $M_* = 10^6 M_\odot$  and  $M_* = 4 \times 10^6 M_\odot$ , we recover the expected dichotomy between EMRIs and DPs. In this case, EMRIs can only form from orbits with  $a_0 \lesssim a_c$ , while those with  $a_0 \gtrsim a_c$  exclusively result in DPs. We find that the transition between the two regimes occurs over about an order of magnitude in  $a_0$  at  $a_c \simeq 10^{-2} R_{\text{inf}}$ , which is consistent within a factor of two with previous works (Hopman & Alexander 2005; Raveh & Perets 2021, QS24). We believe that the quantitative difference is driven by differences in the nuclear cluster models, in the accuracy of the dynamical model (e.g. PN order) and in the treatment of two-body relaxation. We better explore the impact that some of these factors have on the shape of  $S(a_0)$  in Sects. 4.2 and 4.3.

Conversely, for  $M_* < 10^6 M_\odot$ , we observe a relevant number of EMRIs formed from initially wide orbits, with  $a_0 \gg a_c$ . These events are consistent with the cliffhanger EMRI population first



**Fig. 10.** Distribution of the maximum ratio between  $a$  evaluated at subsequent apocentres for EMRIs for selected values of  $M_\bullet$  and  $a_0$ . The black dotted vertical line shows  $(a_\alpha/a_\beta)_{\max}^{\text{EMRI}} = 2$ , which roughly delimits cliffhanger EMRIs to its right. In the left panel, the orange dashed and green dash-dotted vertical lines show  $a_0/a_\infty^c$  from Eq. (38), which is a rough prediction of the highest possible  $(a_\alpha/a_\beta)_{\max}^{\text{EMRI}}$  ratio. Each histogram is normalised to 1. We note the different  $x$ -axis scales and binning strategies in the two panels.

seen in Monte Carlo simulations by QS24 and, to the best of our knowledge, never found in previous work.

We observe cliffhanger EMRIs forming for  $M_\bullet \lesssim 3 \times 10^5 M_\odot$  and they appear to be more common as  $M_\bullet$  decreases, consistently with the argument based on the  $(1-e, a)$  phase space presented in Sect. 2.4.

Our results for  $S(a_0)$  are compared to those obtained by QS24 and other authors in the right panel of Fig. 9. First, we can notice the large scatter in  $a_c/R_{\text{inf}}$  that spans almost an order of magnitude, from  $\approx 3 \times 10^{-3}$  (QS24) to  $\approx 2 \times 10^{-2}$  (Raveh & Perets 2021). This is likely due to the differences in the parameters of the simulated systems as well as in the treatment of the dynamics and loss cone definition, as we discuss in Sect. 4.3. Focusing on cliffhanger EMRIs, QS24 observe their formation only for MBHs with  $M_\bullet \lesssim 5 \times 10^4 M_\odot$ , a threshold set almost an order of magnitude below our limiting value. Moreover, we see a much larger fraction of EMRIs over DPs at large  $a_0$ , with a maximum ratio of  $S(a_0) \approx 0.65$  at  $a_0 \gg 10^{-2} R_{\text{inf}}$  for  $M_\bullet = 10^4 M_\odot$ . As a further difference, our  $S(a_0)$  functions hint at a downward trend for  $a_0 \gg 10^{-2} R_{\text{inf}}$ , as opposed to tending to a constant at  $a_0 \rightarrow \infty$ . This could be due to the fact that QS24 stopped their investigation at  $a_{\max} \approx R_{\text{inf}}/3$ , as their implementation does not account for the stellar potential.

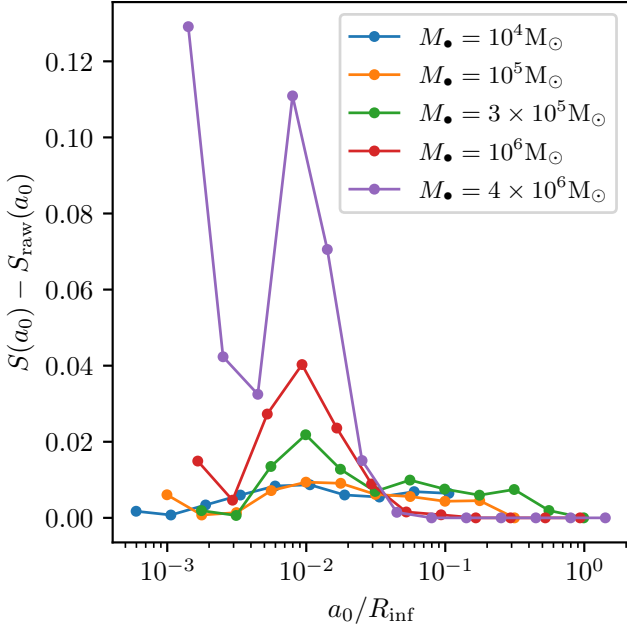
Consider now  $a_\alpha$  and  $a_\beta$  as the values of the semi-major axis of the orbit at two subsequent generic apocentres. For each run that resulted in the formation of an EMRI, we consider the largest ratio  $a_\alpha/a_\beta$  to set a criterion to distinguish cliffhanger and traditional EMRIs. Consistently with the reasoning that led us to Eq. (35), we determine the passage across the cliffhanger region if at some point in the orbital evolution the semi-major axis is cut in half or more during a single period, thus for all the runs where  $(a_\alpha/a_\beta)_{\max}^{\text{EMRI}} > 2$ . This threshold is somehow arbitrary, but offers a quantitative way to identify cliffhanger EMRIs. Figure 10 shows the result of this investigation for  $M_\bullet = 10^5 M_\odot$  and  $M_\bullet = 10^6 M_\odot$ . For both MBH masses, all EMRIs forming for

$a_0 \ll 10^{-2} R_{\text{inf}}$  are classical EMRIs. Both cases show an increase in the average value of  $(a_\alpha/a_\beta)_{\max}^{\text{EMRI}}$  for larger  $a_0$ , but only for  $M_\bullet = 10^5 M_\odot$  the threshold of 2 is crossed. Moreover, we verify that the prediction for the largest possible  $(a_\alpha/a_\beta)_{\max}^{\text{EMRI}}$  ratio (which occurs at the loss cone edge) given in Eq. (38) is verified for  $M_\bullet = 10^5 M_\odot$  for large  $a_0$ , with some deviations for smaller  $a_0$ . This is likely because the approximation given in Eq. (37) is less precise for smaller semi-major axes.

Finally, in Fig. 11, we show the impact on the function  $S(a_0)$  of the correction to the stopping criterion for EMRIs detailed in Sect. 3.4. Largest masses are more affected by the issue, in particular for low  $a_0$  and near the inflection point at  $a \approx 10^{-2} R_{\text{inf}}$ . The figure clearly shows that this correction is needed to correctly recover the limit of 1 of  $S(a_0)$  at  $a_0 \ll 10^{-2} R_{\text{inf}}$ , while it does not affect the curve at  $a_0 \gg R_{\text{inf}}$  region, where we observe cliffhanger EMRIs.

#### 4.2. Local versus orbit-averaged approach

One of the main objectives of our investigation is that of testing the reliability of the orbit-averaged approximation to describe the complex interactions in nuclear clusters. As explained in Sect. 2.1, this approach is based on the idea of orbit-averaging the local diffusion coefficients, which in reality strongly depend on the exact position  $r$  and velocity  $v$  of the orbiting body in consideration. In highly eccentric orbits, such as the one needed to form EMRIs, both  $r$  and  $v$  oscillate by orders of magnitude in a single period, and orbit-averaging diffusion coefficients might lead to a mismodelling of the very different features displayed by relaxation at  $r_-$  or  $r_+$ . Another common practice in the orbit-averaging approach that is likely to lead to inaccuracies is that of accounting for two-body relaxation only once per period to update the orbital elements. This effectively means that once the energy and angular momentum of the particle are updated (usually at apocentre) the orbit is frozen for a full period. This

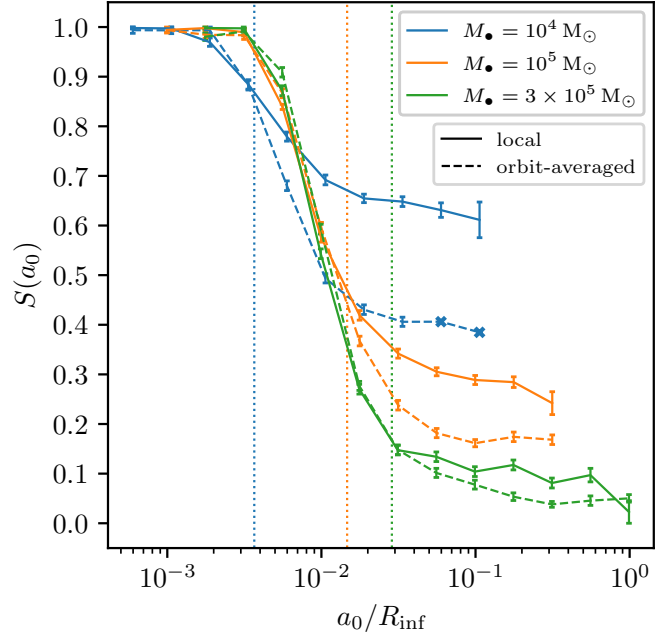


**Fig. 11.** Impact of the correction to the stopping criterion on the function  $S(a_0)$ . We show the difference between the function displayed in Fig. 9 and  $S_{\text{raw}}(a_0)$ , which is the one obtained without correcting the counts of EMRIs.

treatment is, however, at odd with the concept of full loss cone itself, as we come across the contradiction of being in the regime where escaping from the loss cone should be easy while being effectively locked onto an orbit, that thus cannot escape.

The relaxation timescale for a particle with energy  $\mathcal{E}$  and pericentre  $r_-$  estimates the average time that the particle passes on the orbit before being scattered away. When  $r_-$  is much smaller than the semi-major axis of the orbit, it estimates the time before  $r_-$  is changed by the order of itself, as energy remains almost constant (Cohn & Kulsrud 1978). In this small  $r_-$  regime the relaxation timescale for the orbit can be estimated as  $t_{\text{rlx}} \approx P r_- / (q r_{\text{lc}})$  (Merritt 2013a; Bortolas et al. 2023; Broggi et al. 2024), where  $P$  is the orbital period. Consider now an orbit with energy  $\mathcal{E}$  and  $r_- < r_{\text{lc}}$ : in a time  $P$  it will be reached by a number  $dN$  of stellar objects due to two-body relaxation. In the orbit-averaged approach, they will all be DPs produced in the subsequent orbital period, with a rate from that orbit of  $d\dot{N} = dN/P$ . In the local approach, if  $t_{\text{rlx}} < P$ , only the stellar objects that will take less than  $t_{\text{rlx}}$  to reach the pericentre will be DPs, and the corresponding rate will be  $d\dot{N} \approx dN t_{\text{rlx}} / P^2 = r_- / (q r_{\text{lc}}) dN / P$ . Therefore, the orbit-averaged approach will overestimate the rate of DPs for  $q \gtrsim 1$  due to the ‘frozen orbit’ approach, as they have  $r_- < r_{\text{lc}}$  by definition. The same reasoning applies to cliffhangers, that are produced at very small pericentres, too. However, being their pericentre larger than  $r_{\text{lc}}$  by definition, the overestimation of cliffhangers in the orbit-averaged assumption will be lower. Consequently, we expect that the orbit-averaged approach will underestimate  $S(a)$ .

In order to test this expectation we re-ran the three sets of simulations with  $M_{\bullet} = 10^4, 10^5, 3 \times 10^5 M_{\odot}$  with the exact same system parameters but employing the orbit-averaged procedure described in Sec. 3.3. Results of these test runs are compared to our default simulations in Fig. 12, and appear to confirm our expectation. The  $S(a_0)$  functions computed in the local vs orbit-averaged simulations nicely match up to the value of  $a_0/R_{\text{inf}}$



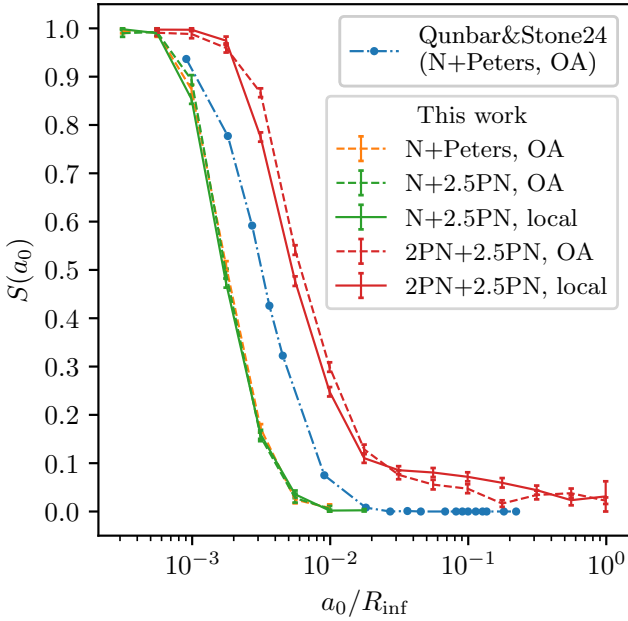
**Fig. 12.** Comparison of the  $S(a_0)$  function obtained by employing the local (solid lines) vs orbit-averaged (dashed lines) treatment of two-body relaxation. The vertical dotted lines mark, for each case, the value of  $a_0/R_{\text{inf}}$  at which  $q = 1$ , thus marking the transition from the empty (at smaller  $a_0$ ) to the full loss cone regime. The crosses indicate the points for which the orbit-averaged procedure explained in Sect. 3.3 failed many times.

for which  $q = 1$ , which marks the transition from empty to full loss cone (vertical dotted lines in the figure). At larger  $a_0$ , the orbit-averaged approach tends to produce less cliffhanger EMRIs and more plunges, thus resulting in a lower  $S(a_0)$ . We note that the trend becomes more severe by lowering the central MBH mass, consistent with the fact that the cliffhanger region becomes larger and larger. In practice, the smaller the central MBH, the bigger gets the cliffhanger region compared to the loss cone, and it becomes increasingly more difficult for the orbit-averaged approximation to capture the detailed dynamics of particle scattered within the cliffhanger region in real time along the orbit close to the pericentre, before getting to the loss cone, thus leading to an over-estimation of DPs.

#### 4.3. Detailed investigation of the differences with QS24

Our main results display some deviations with respect to other works in the literature. In particular the emergence of cliffhanger EMRIs and their abundance does not match findings by QS24. This is not surprising, given the many differences between our simulations and theirs. In particular, we simulate a two-population system, including the effect of solar-mass stars and stellar-mass BHs in the computation of the diffusion coefficient, we adopt a local treatment of two-body relaxation and we evolve the stellar-mass BH using PN equations of motion. Conversely, in QS24 scattering comes from a single population of solar-mass object, is treated as orbit-averaged and the dynamics of the stellar-mass BH is Newtonian.

In the previous section, we already highlighted the importance of the local treatment of relaxation, which leads to a much larger relative fraction of EMRIs in the full loss cone regime. This can certainly account for some of the differences with QS24. For example in the case of  $M_{\bullet} = 10^4 M_{\odot}$ , the



**Fig. 13.** Function  $S(a_0)$  given different assumptions on the dynamics and treatment of two-body relaxation. For the dynamics, we distinguish between Newtonian (N) or PN evolution of the system, and between employing the equations detailed in Peters (1964) or the 2.5PN term to treat loss of energy and angular momentum due to GW emission. Instead, two-body relaxation can be accounted for using the orbit-averaged approximation (indicated with OA), or locally. In all cases  $M_\bullet = 3 \times 10^5 M_\odot$  and the MBH is surrounded by a distribution of  $m_\star = 1 M_\odot$  stars only.

orbit-averaged tests shown in Fig. 12 result in  $S(a_0) \approx 0.4$  at large  $a_0$ , which is much closer to the value of  $\approx 0.25$  found in QS24 compared to the  $\gtrsim 0.6$  found in our default simulations employing a local treatment of relaxation.

In order to further test how the differences between the QS24 and our studies affect the results, we perform further targeted simulations, focusing on the  $M_\bullet = 3 \times 10^5 M_\odot$  case, as it is close to the MBH mass value at which the majority of LISA EMRI detections is expected (Babak et al. 2017). We try to reproduce the results by QS24<sup>11</sup> by running our simulations with a setup which is as close as possible to theirs. We use a single population of scattering objects of  $m_\star = 1 M_\odot$ , distributed according to a Denhen density profile characterised by  $M_{\text{tot},\star} = 1425 M_\bullet$ ,  $r_{a,\star} = 223.37 R_{\text{inf}}$  and  $\gamma_\star = 1.75$ . With these parameters, our density profile departs from the single  $r^{-1.75}$  power law simulated by QS24 by less than 1% within the MBH influence radius.

Qunbar & Stone (2024) use Newtonian dynamics, and take into account emission of GWs through the equations detailed in Peters (1964). We do the same by setting  $H_1 = H_2 = H_{2.5} = 0$  and adding the contributions

$$\Delta \mathcal{E}_{\text{GW}} = P \left\langle \frac{d\mathcal{E}}{dt} \right\rangle_{\text{GW}}, \quad \Delta J_{\text{GW}} = P \left\langle \frac{dJ}{dt} \right\rangle_{\text{GW}} \quad (58)$$

to the orbital sums of the stochastic kicks  $\Delta \mathcal{E} = \sum_i \delta \mathcal{E}_i$  and  $\Delta J = \sum_i \delta J_i$  once per orbit, before applying the velocity kick following the orbit-averaged procedure described in Sect. 3.3.

We report the results of these tests in Fig. 13. Despite our attempt to reproduce the system employed in QS24 as close as

<sup>11</sup> QS24 do not present results for a  $M_\bullet = 3 \times 10^5 M_\odot$  scenario; however, the authors kindly agreed to share the results with us and to let us show them in Fig. 13.

possible, we cannot replicate their results. Although the shape of the  $S(a_0)$  function is very similar, our  $a_c$  is approximately a factor of two smaller than what they found. It is likely that this discrepancy stems from the neglect of the stellar potential in QS24, or in differences in the specific implementation of orbit-averaging. Further investigation is still needed to get a definitive answer. Moreover, we observe that  $S(a_0)$  is not modified by using the 2.5PN term to account for GW radiation instead of the equations detailed in Peters (1964). As discussed in the previous section, for this central MBH mass, the local treatment of two-body relaxation does not have a strong impact on the  $S(a_0)$  function. We find that in Newtonian dynamics the local treatment results in no discernible deviation of  $S(a_0)$  compared to the orbit-averaged approximation. At second PN order instead, we observe some more significant deviations. In particular, as noted in the previous section, our local approach predicts only slightly more DPs for  $a_0 \approx 3 \times 10^{-3} R_{\text{inf}}$  and more EMRIs for  $a_0 \gtrsim 2 \times 10^{-2} R_{\text{inf}}$ . Finally, a large difference comes from the choice of dynamics. When going from Newtonian to PN dynamics,  $S(a_0)$  is shifted to larger semi-major axes and cliffhanger EMRIs become more likely. This is because PN orbits can get closer to the MBH without being accreted as DPs ( $R_{\text{lc}} = 5.6 R_g$  instead of  $8 R_g$ ), thus having the chance of experiencing much larger jumps in the  $(a, 1-e)$  plane along the loss cone line. This, in turn, means that cliffhanger EMRIs can form from larger  $a_0$ , consistent with Fig. 13.

These results show that  $a_c$  (i.e. the EMRI-DP separatrix) as well as the abundance of cliffhanger EMRIs depends on the treatment of the stellar-mass BH orbit dynamics. Since PN orbits are a better approximation to GR than Newtonian gravity, we speculate that our PN results are closer to reality than those obtained by employing Newtonian dynamics. However, a full geodesic orbital description in GR would be needed to provide an accurate characterisation of the  $S(a_0)$  function.

#### 4.4. Effect on EMRI rates

We now apply our results on  $S(a_0)$  to assess the impact of cliffhanger EMRIs on the rates of EMRI predicted by Fokker-Planck codes. Classically, the rate of EMRIs is computed from the differential rate in energy of loss cone captures

$$\mathcal{F}_{\text{lc}}(\mathcal{E}) \equiv \frac{d\dot{N}}{d\mathcal{E}} \quad (59)$$

that can be directly computed during the evolution (Cohn & Kulsrud 1978). The total rate of EMRIs and DPs per unit time can be computed as follows:

$$\dot{N}_{\text{tot}} = \int_0^\infty d\mathcal{E} \mathcal{F}_{\text{lc}}(\mathcal{E}). \quad (60)$$

In general,  $S(a_0)$  acts on the differential rate as a transfer function, giving the fraction of loss cone events happening at semi-major axis  $a_0$  that will result in EMRIs. The function  $1 - S(a_0)$  gives, symmetrically, the fraction of DPs. In order to apply the transfer function  $S(a_0)$ , one needs to map the orbit in a generic potential to a typical scale radius, interpreted as the semi-major axis when the stellar object enters the loss cone or the GW-dominated region, meaning it is considered to be captured. A reliable choice for objects captured at a given energy is the radius  $r_c$  of the circular orbit with the same energy, that reduces to the semi-major axis in the Keplerian regime:

$$S(\mathcal{E}) = S(r_c(\mathcal{E})). \quad (61)$$



Another possibility is  $a_0 = (r_- + r_+)/2$ , but it is affected by the stellar potential at smaller  $\mathcal{E}$ . The two mappings for typical loss cone orbits depend on the total potential evaluated at  $r_c = a_0$  and  $r_+ \simeq 2 a_0$  respectively. The rates of DPs and EMRIs can be computed as

$$\dot{N}_{\text{DP}} = \int_0^\infty d\mathcal{E} [1 - S(\mathcal{E})] \mathcal{F}_{\text{lc}}(\mathcal{E}), \quad (62)$$

$$\dot{N}_{\text{EMRI}} = \int_0^\infty d\mathcal{E} S(\mathcal{E}) \mathcal{F}_{\text{lc}}(\mathcal{E}). \quad (63)$$

When using the classical discriminant between EMRIs and DPs, one sets  $S(\mathcal{E}) = \Theta(\mathcal{E} - \mathcal{E}_c)$ , where  $\Theta(x)$  is the Heaviside step function, that is zero when its argument is negative and one otherwise, and  $\mathcal{E}_c$  is the energy such that  $r_c(\mathcal{E}_c) = a_c$ . Accordingly, the classical rate of the two classes of captures, which we denote with the upper index ‘cl’, is

$$\dot{N}_{\text{DP}}^{\text{cl}} = \int_0^{\mathcal{E}_c} d\mathcal{E} \mathcal{F}_{\text{lc}}(\mathcal{E}), \quad (64)$$

$$\dot{N}_{\text{EMRI}}^{\text{cl}} = \int_{\mathcal{E}_c}^\infty d\mathcal{E} \mathcal{F}_{\text{lc}}(\mathcal{E}). \quad (65)$$

#### 4.4.1. The case $M_\bullet = 3 \times 10^5 M_\odot$ : implications for the most relevant LISA systems

We first specialise to the case  $M_\bullet = 3 \times 10^5 M_\odot$ , which is where the bulk of LISA detections are expected, and use the code presented in Broggi et al. (2022) to simulate the relaxation of the nuclear cluster surrounding it. The initial Dehnen profile of stars and stellar-mass BHs has  $\gamma_\star = \gamma_\bullet = 1.5$ , while the other properties match those of Sect. 3.1. At later times, the system relaxes and two-body scattering in the inner region are dominated by stellar-mass BHs, whose distribution reaches  $\gamma_\bullet = 1.7$ , as in Sect. 3.1. The code accounts for a non-evolving stellar potential set by the initial conditions and includes the effects of two-body relaxation, but does not account for GWs emission. Since the GWs driven region of phase space encloses the loss cone for  $a < a_c$ , we interpret  $\mathcal{F}_{\text{lc}}$  for  $\mathcal{E} > \mathcal{E}_c$  as the differential rate entering the GW dominated region<sup>12</sup>. For this system  $R_{\text{inf}} = 0.57$  pc and  $\mathcal{E}_c = 63 \sigma_{\text{inf}}^2$ , where  $\sigma_{\text{inf}} = \sigma(R_{\text{inf}})$ . We build  $S(\mathcal{E})$  by interpolating the function shown in Fig. 9 linearly in the log-log space. We extrapolate as  $S(\mathcal{E}) = 0$  for  $\mathcal{E} \rightarrow 0$  ( $a_0 \rightarrow \infty$ ), and  $S(\mathcal{E}) = 1$  for  $\mathcal{E} \rightarrow \infty$  ( $a_0 \rightarrow 0$ ).

In Fig. 14, we show the differential rate  $\mathcal{F}_{\text{lc}}$  computed at  $t = 25$  Myr and  $t = 45$  Myr, both after the initial numerical transient (coming from the perfectly isotropic initial conditions), but before the peak of the classical EMRI rates, which happens at  $t = 75$  Myr. We also show the differential rate of EMRIs

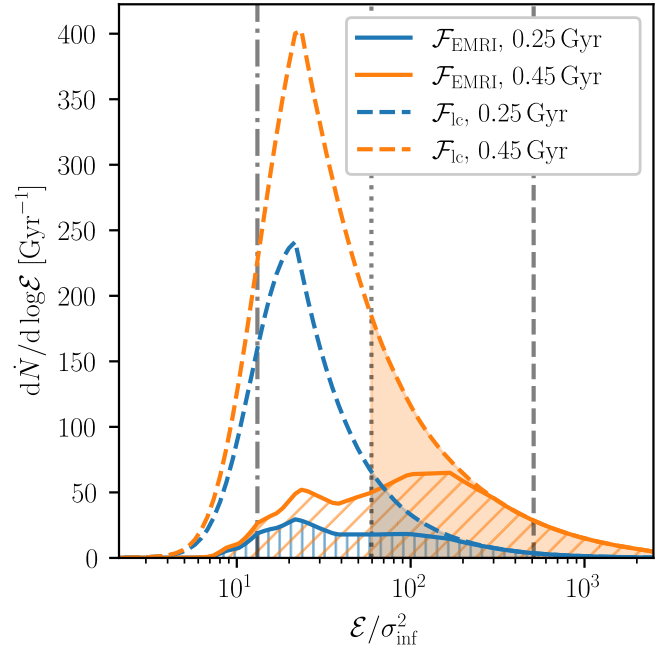
$$\mathcal{F}_{\text{EMRI}}(\mathcal{E}) \equiv \mathcal{F}_{\text{lc}}(\mathcal{E}) S(\mathcal{E}) \quad (66)$$

for a direct comparison with its classical estimate

$$\mathcal{F}_{\text{EMRI}}^{\text{cl}}(\mathcal{E}) \equiv \mathcal{F}_{\text{lc}}(\mathcal{E}) \Theta(\mathcal{E} - \mathcal{E}_c). \quad (67)$$

The instantaneous rate of EMRIs and DPs according to our and the classical prescriptions are reported in Table 1. At  $t = 25$  Myr, when EMRIs classically account for 15% of the total loss cone event rate, Eq. (62) gives a total rate of EMRIs larger by

<sup>12</sup> Some recent works refer to the line  $t_{\text{GW}} \simeq t_{\text{rlx}}$  as the loss cone for  $a < a_c$  (Bondani et al. 2024; Kaur et al. 2024; Rom et al. 2024). In this work, with ‘loss cone curve’ we always refer to the  $r_- = R_{\text{lc}}$  curve.



**Fig. 14.** Differential rate of EMRIs  $\mathcal{F}_{\text{EMRI}}$  (solid lines) and of EMRIs and DPs  $\mathcal{F}_{\text{lc}}$  (dashed lines) as functions of energy. Quantities refer to a nuclear cluster with  $M_\bullet = 3 \times 10^5 M_\odot$  that has relaxed for 25 Myr (blue lines) and 45 Myr (orange lines). We report for reference vertical lines at the values of  $\mathcal{E}$  such that  $r_c(\mathcal{E})$  is  $0.1 R_{\text{inf}}$  (dash-dotted),  $0.01 R_{\text{inf}}$  (dotted) and  $0.001 R_{\text{inf}}$  (dashed). The colour-filled area of  $\mathcal{F}_{\text{lc}}$  gives the classical EMRI rate at  $r_c(\mathcal{E}) > 0.01 R_{\text{inf}}$ , to be compared with the hatched area of  $\mathcal{F}_{\text{EMRI}}$  over the full energy range. Apparent areas in the plot are proportional to the various integrals of the form  $\int d\mathcal{E} \mathcal{F}$ .

$\sim 35\%$ . The rate of DPs is only reduced by  $\sim 5\%$ , as they are overall more numerous. Therefore, with our new estimates EMRIs account for 20% of the total loss cone event rate. At  $t = 45$  Myr, EMRIs classically account for  $\sim 30\%$  of the total loss cone event rate, with our new estimate of the EMRI and DP rates comparable to the classical estimates.

Our model predicts a fraction of EMRIs produced in the classical region  $\mathcal{E} > \mathcal{E}_c$  of 45% at  $t = 25$  Myr, and 60% at  $t = 45$  Myr. Even with a small value  $S(a_0) \simeq 0.1$  at large semi-major axes, cliffhanger EMRIs are a relevant contribution to the overall count of EMRIs due to the shape of the loss cone flux.

Aside from the rates, the inclusion of  $S(a_0)$  has three direct consequences on the energy distribution of EMRIs and DPs:

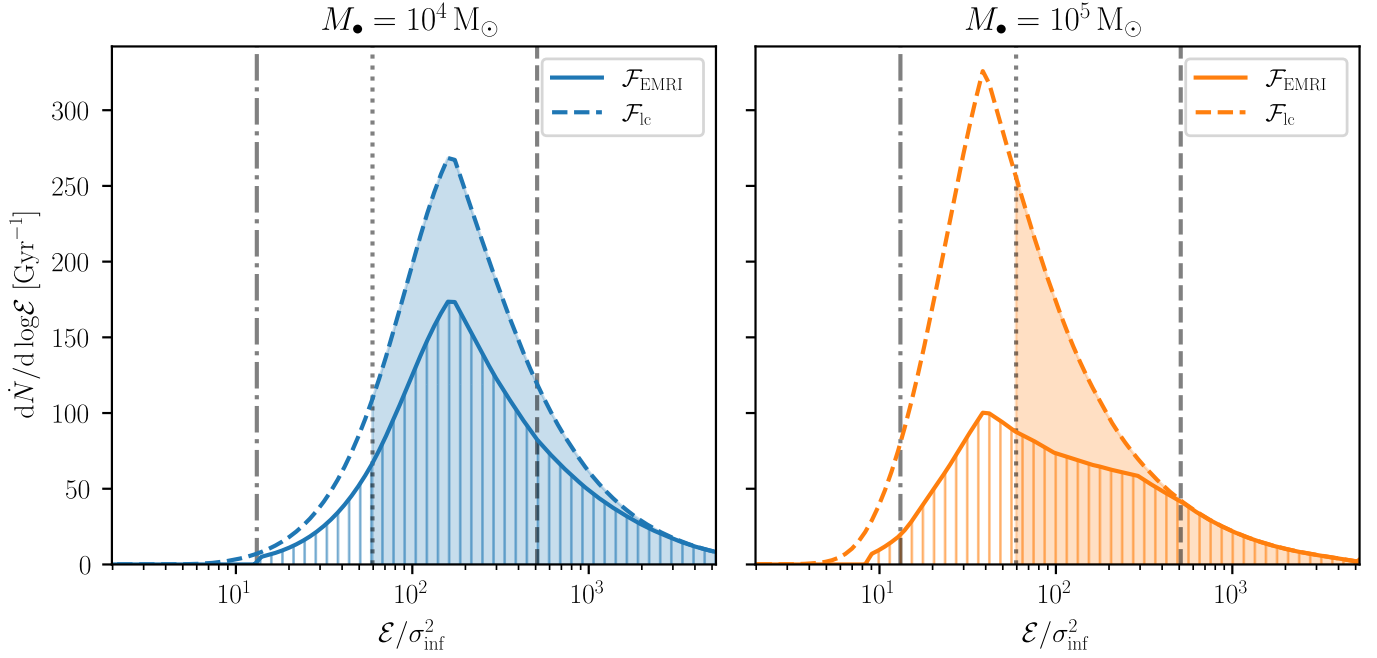
- about 10% of DPs are produced for  $a_0 < a_c$ , mainly between  $\mathcal{E}_c$  ( $r_c = 0.01 R_{\text{inf}}$ ) and  $5 \mathcal{E}_c$  ( $r_c = 0.002 R_{\text{inf}}$ );
- there is a significant deviation from the classical expectation that most EMRIs are produced with  $a_0 \simeq a_c$ , even restricting to  $a_0 \leq a_c$ ;
- the differential rate of EMRIs has a local maximum where the overall loss cone rate has a peak, in this case at  $r_c \simeq 0.07 R_{\text{inf}}$ , corresponding to cliffhangers.

The first two points depend mainly on the fact that  $S(a_0)$  takes approximately an order of magnitude in semi-major axis (or, equivalently, in energy) to smoothly decrease from 1 to its minimum value  $S_{\text{min}}$ , with  $S_{\text{min}} < S(a_c) < 1$  at any value of  $M_\bullet$ . This shows the limitations of approximating  $S(\mathcal{E})$  as a step function. The last point, on the other hand, is exclusively caused by the novelty that  $S(a_0) > 0$  even for  $a_0 \gg a_c$  introduced by QS24.

**Table 1.** Instantaneous EMRI+DP ( $\dot{N}$ ), EMRI ( $\dot{N}_{\text{EMRI}}$ ), and DP ( $\dot{N}_{\text{DP}}$ ) rates produced in the snapshots we investigated.

$M_{\bullet}$ [ $M_{\odot}$ ]	$t$ [Myr]	$\dot{N}$ [ $\text{yr}^{-1}$ ]	$\dot{N}_{\text{EMRI}}^{\text{cl}}$ [ $\text{yr}^{-1}$ ]	$\dot{N}_{\text{EMRI}}$ [ $\text{yr}^{-1}$ ]	$\dot{N}_{\text{DP}}^{\text{cl}}$ [ $\text{yr}^{-1}$ ]	$\dot{N}_{\text{DP}}$ [ $\text{yr}^{-1}$ ]
$10^4$	0.58	$5.9 \times 10^{-7}$	$5.3 \times 10^{-7}$	$4.0 \times 10^{-7}$	$6.5 \times 10^{-8}$	$1.9 \times 10^{-7}$
$10^5$	11	$6.6 \times 10^{-7}$	$3.1 \times 10^{-7}$	$2.9 \times 10^{-7}$	$3.7 \times 10^{-7}$	$3.8 \times 10^{-7}$
$3 \times 10^5$	25	$3.5 \times 10^{-7}$	$4.9 \times 10^{-8}$	$6.7 \times 10^{-8}$	$3.0 \times 10^{-7}$	$2.8 \times 10^{-7}$
$3 \times 10^5$	45	$7.4 \times 10^{-7}$	$2.0 \times 10^{-7}$	$2.1 \times 10^{-7}$	$5.4 \times 10^{-7}$	$5.3 \times 10^{-7}$

**Notes.** We report estimates according to our formulation Eq. (62) (no superscript) based on the transfer function  $S(a_0)$ , and the classical estimates Eq. (64) (superscript cl) based solely on the critical semi-major axis  $a_c = 0.01 R_{\text{inf}}$ .



**Fig. 15.** Differential rate of EMRIs  $\mathcal{F}_{\text{EMRI}}$  (solid lines) and of EMRIs and DPs  $\mathcal{F}_{\text{c}}$  (dashed lines) as functions of energy. The quantities refer to a nuclear cluster with  $M_{\bullet} = 10^4 M_{\odot}$  that has relaxed for 580 kyr (left panel) and one with  $M_{\bullet} = 10^5 M_{\odot}$  that has relaxed for 11 Myr (right panel). We report for reference vertical lines at the values of  $\mathcal{E}$  such that  $r_c(\mathcal{E})$  is  $0.1 R_{\text{inf}}$  (dash-dotted),  $0.01 R_{\text{inf}}$  (dotted) and  $0.001 R_{\text{inf}}$  (dashed). The colour-filled area of  $\mathcal{F}_{\text{c}}$  gives the classical EMRI rate at  $r_c(\mathcal{E}) > 0.01 R_{\text{inf}}$ , to be compared with the hatched area of  $\mathcal{F}_{\text{EMRI}}$  over the full energy range. Apparent areas in the plot are proportional to the various integrals of the form  $\int d\mathcal{E} \mathcal{F}$ .

#### 4.4.2. Effects on EMRIs from lower central MBH masses

Having explored in detail the most relevant LISA case, it is also interesting to have a look at the EMRI rate implications for lower MBH masses, where cliffhanger EMRIs have been shown to dominate the  $S(a_0)$  function. Results are shown in Fig. 15 and quantified in Table 1 for the cases  $M_{\bullet} = 10^4, 10^5 M_{\odot}$ . Rates are computed at  $t/t_p = 0.6$ , where  $t_p$  is the peaking time of the classical EMRI rate for the system under scrutiny (Broggi et al. 2022).

Despite cliffhangers becoming increasingly more common at lower MBH masses, their effect on the overall EMRI rates gets progressively less important. This is mainly because the peak energy of the loss cone flux,  $\mathcal{E}_{\text{lc,p}}$ , becomes comparable (and even bigger) than  $\mathcal{E}_{\text{c}}$  at low MBH masses. In the  $M_{\bullet} = 10^5 M_{\odot}$  case,  $\mathcal{E}_{\text{c}} \gtrsim \mathcal{E}_{\text{lc,p}}$ , therefore adding a significant population of cliffhanger EMRIs at lower energies has a relevant impact on the overall EMRI population. Despite the overall EMRI rate being substantially unaffected (Table 1) we find that  $\sim 37\%$  of those EMRIs are cliffhanger and that the peak of the EMRI flux occurs at  $a > a_c$  ( $\mathcal{E} < \mathcal{E}_{\text{c}}$ ) which might result in extremely eccentric systems. Conversely, the overall EMRI rate decreases in the  $M_{\bullet} = 10^4 M_{\odot}$  case, with cliffhangers accounting only to  $\sim 10\%$

of the overall EMRI population. Notably, for such a low MBH mass,  $\mathcal{E}_{\text{c}} < \mathcal{E}_{\text{lc,p}}$  meaning that most EMRIs form well inside  $a_c$  potentially resulting in lower eccentric systems.

This analysis shows that while the impact of the detailed treatment of nuclear dynamics and the appearance of cliffhangers on the EMRI rate might be only moderate, the energy distribution of EMRIs (and hence their detailed properties) can be highly affected. In particular, a simplistic treatment of EMRIs as systems mostly forming at  $a = a_c$  fails in capturing the rich and MBH mass dependent features of this class of GW sources.

## 5. Discussion and conclusions

In this work we developed a formalism to locally account for the effects of two-body relaxation on the formation of EMRIs and DPs around an MBH at the core of a nuclear cluster, for which we considered a two-population model. We examined a binary system comprised of an MBH and a stellar-mass BH orbiting around it, evolving the system by using PN dynamics up to the 2.5PN term. We ran a number of simulations and studied the EMRI-to-DP ratio as a function of the initial semi-major axis  $a_0$  of the orbit through the examination of the function  $S(a_0)$ ,

for five different central MBH masses. Our main findings can be summarised as follows:

1. For low MBH masses the classical picture, for which there is a critical value  $a_c$  such that EMRIs only form at  $a_0 \ll a_c$  while no EMRIs occur at  $a_0 \gg a_c$ , breaks down. Instead, we find a significant population of EMRIs forming for large values of  $a_0$ . Such a population is consistent with the novel phenomenon of cliffhanger EMRIs, first identified in QS24.
2. Cliffhanger EMRIs are more common for low-mass MBHs because there is a larger region in the  $(1 - e, a)$  phase space which allows for a drastic reduction of the orbit in a single pericentre passage due to GW radiation. Following this analytical argument, we predict that the maximum MBH mass expected to produce cliffhanger EMRIs, given our assumption, is roughly  $3 \times 10^5 M_\odot$ , which was verified in our numerical investigation.
3. Employing PN dynamics to evolve the system leads to more EMRIs, strongly modifying the shape of the function  $S(a_0)$  by shifting  $a_c$  to larger values and boosting the tail of cliffhanger EMRIs produced at  $a_0 \gg a_c$ .
4. The choice of locally accounting for two-body relaxation does significantly influence the EMRI-to-DP ratio for  $M_\bullet \lesssim 10^5 M_\odot$ , leading to a much larger fraction of cliffhanger EMRIs compared to that found in the orbit-averaged approximation.
5. Cliffhanger EMRIs account for a significant fraction of the total EMRI rate for the MBH masses where the loss cone capture differential rate peaks at  $a_p > a_c$  (up to 55% in the snapshots we tested). This is due to a large fraction of EMRIs being produced near this peak.

These results might have a significant impact on the estimates of EMRI detection rates by LISA. Currently, these estimates are still highly uncertain, ranging from a few to a few thousand EMRIs detected per year (Babak et al. 2017), and do not account for the novel population of cliffhanger EMRIs. Furthermore, we note that many cliffhanger EMRIs get to the GW emission region while grazing the loss cone edge, which might imply a significant sub-population of extremely eccentric EMRIs in the LISA band. We defer the investigation of observable cliffhanger EMRI properties by LISA to future work.

Interestingly, we observed that our results are quite sensitive to the choice of the loss cone radius  $R_{lc}$ . As explained in Sect. 4.3, part of the reason for the differences between the  $S(a_0)$  obtained with Newtonian and PN dynamics likely stems from switching from  $R_{lc} = 8R_g$  to  $R_{lc} = 5.6R_g$ , as plunges are more difficult to form and the cliffhanger region grows. For this reason, we believe that the closest approach to a BH on an eccentric orbit should be better explored at PN order, similarly to how it has already been done for circular orbits (e.g. Schäfer & Jaranowski 2024). Compared to Schwarzschild MBHs, spinning MBHs present innermost stable orbits which are closer or further away, respectively, depending on whether the orbit is prograde or retrograde. This again influences the loss cone radius, and consequently the EMRI-to-DP ratio. For this reason, we plan on updating our code to introduce spin effects.

Despite improving on realism by not employing orbit-averages in accounting for two-body relaxation, our approach is still limited on some fronts. Our analysis currently addresses spherical systems with a central Schwarzschild MBH. However, these results might not hold for axisymmetric or triaxial systems due to aspherical two-body relaxation and orbits naturally penetrating the loss cone (Vasiliev & Merritt 2013; Kaur & Stone 2024), or for spinning MBHs (Amaro-Seoane et al. 2013). Furthermore, the two-component Bahcall-Wolf profile we assumed

might underestimate the number of stellar-mass BHs for  $r < R_{inf}$ , as previous works (Hopman & Alexander 2006; Broggi et al. 2022; Rom et al. 2024) have found more segregated profiles. As a result, in our simulations stars dominate two-body relaxation over the entire  $r$  range, while for a completely relaxed system stellar-mass BHs should dominate for  $r \lesssim 0.1R_{inf}$ . While this could affect the shape of  $S(a_0)$ , and in particular the value of  $a_c$ , we note that for the EMRI rate computations of Sect. 4.4 the density profiles were consistently relaxed using the code presented in Broggi et al. (2022), thus resulting in a more segregated distribution of stellar-mass BHs.

Finally, even restricting the discussion to the class of systems we are investigating, we explored a limited region of the  $(1 - e, a)$  plane. Since this choice was not biased (at least for low MBH masses) towards EMRIs or DPs, this did not influence our results on the EMRI-to-DP ratio, but it required the introduction of corrections on the counts of EMRIs for high MBH masses and limits the maximum  $a_0$  that we can explore. This is particularly true for low MBH masses, for which we are not able to reach  $a_0 \approx R_{inf}$ , as most simulated orbits starting from large  $a_0$  result in reinitialisations, thus requiring a very large number of runs to find the EMRI-to-DP ratio within a reasonably small error. We plan on solving this issue in a follow-up work, as pushing the exploration to larger  $a_0$  values is fundamental in order to understand whether cliffhanger EMRIs allow  $S(a_0)$  to reach a plateau  $S(\infty) \neq 0$  or not.

*Acknowledgements.* We thank Ismail Qunbar and Nicholas C. Stone for useful discussions and willingness to share data with us. DM acknowledges that this publication was produced while attending the PhD program in Space Science and Technology at the University of Trento, Cycle XXXIX, with the support of a scholarship financed by the Ministerial Decree no. 118 of 2nd March 2023, based on the NRRP – funded by the European Union – NextGenerationEU – Mission 4 “Education and Research”, Component 1 “Enhancement of the offer of educational services: from nurseries to universities” – Investment 4.1 “Extension of the number of research doctorates and innovative doctorates for public administration and cultural heritage” – CUP E66E23000110001. MB acknowledges support provided by MUR under grant “PNRR - Missione 4 Istruzione e Ricerca – Componente 2 Dalla Ricerca all’Impresa – Investimento 1.2 Finanziamento di progetti presentati da giovani ricercatori ID:SOE\_0163” and by University of Milano-Bicocca under grant “2022-NAZ-0482/B”. AS acknowledges financial support provided under the European Union’s H2020 ERC Consolidator Grant “Binary Massive Black Hole Astrophysics” (B Massive, Grant Agreement: 818691).

## References

- Alexander, T. 2017, *ARA&A*, 55, 17  
Amaro-Seoane, P. 2018, *Liv. Rev. Relativ.*, 21, 4  
Amaro Seoane, P. 2022, *Handbook of Gravitational Wave Astronomy*, 17  
Amaro-Seoane, P., & Preto, M. 2011, *Classical Quantum Gravity*, 28, 094017  
Amaro-Seoane, P., Gair, J. R., Freitag, M., et al. 2007, *Classical Quantum Gravity*, 24, R113  
Amaro-Seoane, P., Sopuerta, C. F., & Freitag, M. D. 2013, *MNRAS*, 429, 3155  
Amaro-Seoane, P., Audley, H., Babak, S., et al. 2017, ArXiv e-prints [arXiv:1702.00786]  
Amaro-Seoane, P., Andrews, J., Arca Sedda, M., et al. 2023, *Liv. Rev. Relativ.*, 26, 2  
Arcodia, R., Merloni, A., Nandra, K., et al. 2021, *Nature*, 592, 704  
Arcodia, R., Liu, Z., Merloni, A., et al. 2024, *A&A*, 684, A64  
Babak, S., Gair, J., Sesana, A., et al. 2017, *Phys. Rev. D*, 95, 103012  
Bahcall, J. N., & Wolf, R. A. 1977, *ApJ*, 216, 883  
Barack, L., & Cutler, C. 2004, *Phys. Rev. D*, 69, 082005  
Barack, L., & Cutler, C. 2007, *Phys. Rev. D*, 75, 042003  
Bar-Or, B., & Alexander, T. 2016, *ApJ*, 820, 129  
Binney, J., & Tremaine, S. 2008, *Galactic Dynamics: Second Edition*  
Blanchet, L. 2014, *Liv. Rev. Relativ.*, 17, 2  
Bode, J. N., & Wegg, C. 2014, *MNRAS*, 438, 573  
Böker, T., Laine, S., van der Marel, R. P., et al. 2002, *AJ*, 123, 1389  
Bondani, S., Bonetti, M., Broggi, L., et al. 2024, *Phys. Rev. D*, 109, 043005  
Bonetti, M., Haardt, F., Sesana, A., & Barausse, E. 2016, *MNRAS*, 461, 4419

- Bortolas, E., & Mapelli, M. 2019, *MNRAS*, **485**, 2125
- Bortolas, E., Mapelli, M., & Spera, M. 2017, *MNRAS*, **469**, 1510
- Bortolas, E., Ryu, T., Broggi, L., & Sesana, A. 2023, *MNRAS*, **524**, 3026
- Broggi, L. 2024, Ph.D. Thesis, Università degli Studi di Milano-Bicocca, <https://hdl.handle.net/10281/459238>
- Broggi, L., Bortolas, E., Bonetti, M., Sesana, A., & Dotti, M. 2022, *MNRAS*, **514**, 3270
- Broggi, L., Stone, N. C., Ryu, T., et al. 2024, *Open J. Astrophys.*, **7**, 48
- Brown, L. D., Cai, T. T., & DasGupta, A. 2001, *Stat. Sci.*, **16**, 101
- Bulirsch, R., & Stoer, J. 1966, *Numerische Mathematik*, **8**, 1
- Carollo, C. M., Stiavelli, M., de Zeeuw, P. T., & Mack, J. 1997, *AJ*, **114**, 2366
- Chandrasekhar, S. 1943, *ApJ*, **97**, 255
- Chen, X., Sesana, A., Madau, P., & Liu, F. K. 2011, *ApJ*, **729**, 13
- Cohn, H., & Kulsrud, R. M. 1978, *ApJ*, **226**, 1087
- Colpi, M., Danzmann, K., Hewitson, M., et al. 2024, ArXiv e-prints [arXiv:2402.07571]
- Côté, P., Piatek, S., Ferrarese, L., et al. 2006, *ApJS*, **165**, 57
- Cutler, C., Kennefick, D., & Poisson, E. 1994, *Phys. Rev. D*, **50**, 3816
- Dehnen, W. 1993, *MNRAS*, **265**, 250
- Eddington, A. S. 1916, *MNRAS*, **76**, 572
- Ferrarese, L., & Merritt, D. 2000, *ApJ*, **539**, L9
- Fragione, G., & Sari, R. 2018, *ApJ*, **852**, 51
- Franchini, A., Bonetti, M., Lupi, A., et al. 2023, *A&A*, **675**, A100
- Freitag, M., & Benz, W. 2001, *A&A*, **375**, 711
- Gair, J. R., Barack, L., Creighton, T., et al. 2004, *Classical Quantum Gravity*, **21**, S1595
- Gair, J. R., Tang, C., & Volonteri, M. 2010, *Phys. Rev. D*, **81**, 104014
- Gallo, E., & Sesana, A. 2019, *ApJ*, **883**, L18
- Gebhardt, K., Bender, R., Bower, G., et al. 2000, *ApJ*, **539**, L13
- Ghosh, R., & Chakravarti, K. 2024, ArXiv e-prints [arXiv:2406.02454]
- Giersz, M. 1998, *MNRAS*, **298**, 1239
- Giustini, M., Miniutti, G., & Saxton, R. D. 2020, *A&A*, **636**, L2
- Gültekin, K., Richstone, D. O., Gebhardt, K., et al. 2009, *ApJ*, **698**, 198
- Hénon, M. 1971, *Ap&SS*, **13**, 284
- Hills, J. G. 1988, *Nature*, **331**, 687
- Hils, D., & Bender, P. L. 1995, *ApJ*, **445**, L7
- Hopman, C., & Alexander, T. 2005, *ApJ*, **629**, 362
- Hopman, C., & Alexander, T. 2006, *ApJ*, **645**, L133
- Jaranowski, P., & Schäfer, G. 1997, *Phys. Rev. D*, **55**, 4712
- Joshi, K. J., Rasio, F. A., & Portegies Zwart, S. 2000, *ApJ*, **540**, 969
- Kaur, K., & Stone, N. C. 2024, ApJ submitted [arXiv:2405.18500]
- Kaur, K., Rom, B., & Sari, R. 2024, ApJ, submitted [arXiv:2406.07627]
- Kormendy, J., & Ho, L. C. 2013, *ARA&A*, **51**, 511
- Kormendy, J., & Richstone, D. 1995, *ARA&A*, **33**, 581
- Laghi, D., Tamanini, N., Del Pozzo, W., et al. 2021, *MNRAS*, **508**, 4512
- Levin, Y. 2007, *MNRAS*, **374**, 515
- Linial, I., & Metzger, B. D. 2023, *ApJ*, **957**, 34
- MacLeod, C. L., & Hogan, C. J. 2008, *Phys. Rev. D*, **77**, 043512
- Maggiore, M. 2007, *Gravitational Waves: Volume 1: Theory and Experiments*
- Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, *AJ*, **115**, 2285
- Mazzolari, G., Bonetti, M., Sesana, A., et al. 2022, *MNRAS*, **516**, 1959
- Merritt, D. 2013a, *Dynamics and Evolution of Galactic Nuclei*
- Merritt, D. 2013b, *Classical Quantum Gravity*, **30**, 244005
- Merritt, D., Alexander, T., Mikkola, S., & Will, C. M. 2011, *Phys. Rev. D*, **84**, 044024
- Metzger, B. D., Stone, N. C., & Gilbaum, S. 2022, *ApJ*, **926**, 101
- Miller, M. C., Freitag, M., Hamilton, D. P., & Lauburg, V. M. 2005, *ApJ*, **631**, L117
- Miniutti, G., Saxton, R. D., Giustini, M., et al. 2019, *Nature*, **573**, 381
- Misner, C. W., Thorne, K. S., & Wheeler, J. A. 1973, *Gravitation* (San Francisco: W.H. Freeman and Company)
- Naoz, S., Rose, S. C., Michaely, E., et al. 2022, *ApJ*, **927**, L18
- Neumayer, N., Seth, A., & Böker, T. 2020, *A&ARv*, **28**, 4
- Nicholl, M., Pasham, D. R., Mummery, A., et al. 2024, *Nature*, **634**, 804
- Pan, Z., & Yang, H. 2021, *Phys. Rev. D*, **103**, 103018
- Pattabiraman, B., Umbreit, S., Liao, W.-K., et al. 2013, *ApJS*, **204**, 15
- Peebles, P. J. E. 1972, *ApJ*, **178**, 371
- Peters, P. C. 1964, *Phys. Rev.*, **136**, 1224
- Qunbar, I., & Stone, N. C. 2024, *Phys. Rev. Lett.*, **133**, 141401
- Raveh, Y., & Perets, H. B. 2021, *MNRAS*, **501**, 5012
- Richardson, L. F. 1911, *Philos. Trans. R. Soc. London Ser. A*, **210**, 307
- Rom, B., Linial, I., Kaur, K., & Sari, R. 2024, *ApJ*, **977**, 7
- Rubbo, L. J., Holley-Bockelmann, K., & Finn, L. S. 2006, *ApJ*, **649**, L25
- Ryan, F. D. 1997, *Phys. Rev. D*, **56**, 1845
- Sadeghian, L., & Will, C. M. 2011, *Classical Quantum Gravity*, **28**, 225029
- Schäfer, G., & Jaranowski, P. 2024, *Liv. Rev. Relativ.*, **27**, 2
- Schödel, R., Eckart, A., Alexander, T., et al. 2007, *A&A*, **469**, 125
- Schödel, R., Merritt, D., & Eckart, A. 2009, *A&A*, **502**, 91
- Servin, J., & Kesden, M. 2017, *Phys. Rev. D*, **95**, 083001
- Shapiro, S. L., & Marchant, A. B. 1978, *ApJ*, **225**, 603
- Spinnato, P. F., Fellhauer, M., & Portegies Zwart, S. F. 2003, *MNRAS*, **344**, 22
- Toubiana, A., Sberna, L., Caputo, A., et al. 2021, *Phys. Rev. Lett.*, **126**, 101105
- Tremaine, S., Richstone, D. O., Byun, Y.-I., et al. 1994, *AJ*, **107**, 634
- Turner, M. L., Côté, P., Ferrarese, L., et al. 2012, *ApJS*, **203**, 5
- Vasiliev, E. 2017, *ApJ*, **848**, 10
- Vasiliev, E., & Merritt, D. 2013, *ApJ*, **774**, 87
- Will, C. M. 2012, *Classical Quantum Gravity*, **29**, 217001
- Wilson, E. B. 1927, *J. Am. Stat. Assoc.*, **22**, 209
- Zhang, F., & Amaro Seoane, P. 2024, *ApJ*, **961**, 232

## Appendix A: Diffusion coefficients

In this work, we consider a stellar-mass BH of mass  $m_\bullet$  and velocity magnitude  $v$  having a sequence of two-body encounters with background bodies, grouped into populations (identified by the index  $\alpha$ ) based on their masses  $m_\alpha$ . We decompose its velocity change as the sum of  $\Delta v_\parallel$ , parallel to the initial line of motion, and  $\Delta v_\perp$ , in the plane orthogonal to the original direction of motion.

We use the notation  $D[X]$  to indicate the expectation value of the generic quantity  $X$  per unit of time. The diffusion coefficients we use in this work are (Chandrasekhar 1943; Binney & Tremaine 2008; Merritt 2013a):

$$D[\Delta v_\parallel] = -16\pi^2 G^2 \ln \Lambda \sum_\alpha m_\alpha (m_\bullet + m_\alpha) \mathcal{I}_{2,\alpha}(r, v), \quad (\text{A.1})$$

$$D[\Delta v_\perp] = 0, \quad (\text{A.2})$$

$$D[(\Delta v_\parallel)^2] = \frac{32\pi^2}{3} G^2 \ln \Lambda v \sum_\alpha m_\alpha^2 (\mathcal{I}_{4,\alpha}(r, v) + \mathcal{J}_{1,\alpha}(r, v)) \quad (\text{A.3})$$

$$D[(\Delta v_\perp)^2] = \frac{32\pi^2}{3} G^2 \ln \Lambda v \times \sum_\alpha m_\alpha^2 (3\mathcal{I}_{2,\alpha}(r, v) - \mathcal{I}_{4,\alpha}(r, v) + 2\mathcal{J}_{1,\alpha}(r, v)). \quad (\text{A.4})$$

Here  $\ln \Lambda$  is the Coulomb logarithm, which we set to 10 as an order of magnitude approximation (Spinnato et al. 2003; Merritt 2013b), while the functions  $\mathcal{I}_{k,\alpha}$  and  $\mathcal{J}_{k,\alpha}$  are defined as

$$\mathcal{I}_{k,\alpha}(r, v) = \int_0^v dv' \left(\frac{v'}{v}\right)^k f_\alpha(r, v'), \quad (\text{A.5})$$

$$\mathcal{J}_{k,\alpha}(r, v) = \int_v^\infty dv' \left(\frac{v'}{v}\right)^k f_\alpha(r, v'), \quad (\text{A.6})$$

with  $f_\alpha$  being the distribution function of objects in population  $\alpha$  on the six-dimensional phase-space of position and velocity. Assuming that the distribution of scatterers is spherical, and approximating it as locally isotropic in velocity,  $f_\alpha$  depends only on the magnitudes of its arguments  $r$  and  $v$ .

In practice, the diffusion coefficients tell us how the velocity of the stellar-mass BH changes per unit time as it moves with velocity  $v$  through the distributions of background bodies at distance  $r$  from their centres. Notably, the perpendicular variation of the velocity  $D[\Delta v_\perp]$  vanishes since the stellar-mass BH is equally likely to be deflected along any of the infinite directions perpendicular to its motion, while the parallel variation  $D[\Delta v_\parallel] < 0$  describes the effect of dynamical friction (Chandrasekhar 1943).

At Newtonian order,  $v = \sqrt{2(\psi - \mathcal{E})}$  for bound orbits. Under this approximation, we can express  $\mathcal{I}_{k,\alpha}$  and  $\mathcal{J}_{k,\alpha}$  as functions of  $\psi(r)$  and  $\mathcal{E}(r, v)$ :

$$\mathcal{I}_{k,\alpha}(\psi, \mathcal{E}) = \frac{1}{\sqrt{2(\psi - \mathcal{E})}} \int_{\mathcal{E}}^\psi d\mathcal{E}' \left(\frac{\psi - \mathcal{E}'}{\psi - \mathcal{E}}\right)^{(k-1)/2} f_\alpha(\mathcal{E}'), \quad (\text{A.7})$$

$$\mathcal{J}_{k,\alpha}(\psi, \mathcal{E}) = \frac{1}{\sqrt{2(\psi - \mathcal{E})}} \int_0^\mathcal{E} d\mathcal{E}' \left(\frac{\psi - \mathcal{E}'}{\psi - \mathcal{E}}\right)^{(k-1)/2} f_\alpha(\mathcal{E}'). \quad (\text{A.8})$$

Here  $f$  depends only on  $\mathcal{E}$  as we assume that the distribution function is ergodic, meaning it depends on  $r$  and  $v$  only through the Hamiltonian:  $f(r, v) = f(H(r, v))$ .

The ergodic distribution function corresponding to a spherical density profile can be computed using Eddington's formula (Eddington 1916; Merritt 2013a):

$$f_\alpha(\mathcal{E}) = \frac{1}{\sqrt{8\pi^2}} \left( \frac{1}{\sqrt{\mathcal{E}}} \frac{dn_\alpha}{d\psi} \Big|_{\psi=0} + \int_0^\mathcal{E} \frac{d\psi}{\sqrt{\mathcal{E}-\psi}} \frac{d^2 n_\alpha}{d\psi^2} \right). \quad (\text{A.9})$$

We can simplify this expression by observing that  $\psi = 0$  when  $r \rightarrow \infty$ . For a Dehnen density profile, at infinite distance we have that  $n_\alpha \propto \psi^4$ , since  $n_\alpha \propto r^{-4}$  and  $\psi \propto r^{-1}$ . Thus,

$$\frac{dn_\alpha}{d\psi} \Big|_{\psi=0} = 4\psi^3 \Big|_{\psi=0} = 0, \quad (\text{A.10})$$

which gives us the simpler expression:

$$f_\alpha(\mathcal{E}) = \frac{1}{\sqrt{8\pi^2}} \int_0^\mathcal{E} d\psi g(\psi), \quad (\text{A.11})$$

where

$$g(\psi) = \frac{1}{\sqrt{\mathcal{E}-\psi}} \frac{d^2 n_\alpha}{d\psi^2}. \quad (\text{A.12})$$

Finally, we can observe that:

$$\frac{d^2 n_\alpha}{d\psi^2} = \frac{dr}{d\psi} \frac{d}{dr} \left( \frac{dr}{d\psi} \frac{d}{dr} n_\alpha \right) = \frac{1}{(\psi')^2} n_\alpha'' - \frac{\psi''}{(\psi')^3} n_\alpha', \quad (\text{A.13})$$

where the prime symbol indicates derivation with respect to  $r$ .

In practice, to find  $f_\alpha$  we have to integrate  $g(\psi)$  over  $\psi$  between 0 and  $\mathcal{E}$ . We perform this numerically by evaluating the function  $g(\psi)$  at discrete values  $\psi_i$  inside the range  $[0, \mathcal{E}]$ . Thus, to use the expression for the second derivative of  $n_\alpha$  that we just found, for each of these values we first numerically find the  $r_i$  such that  $\psi(r_i) = \psi_i$  by solving Eq. (5). Then, we can compute  $\psi'(r_i)$ ,  $\psi''(r_i)$ ,  $n'(r_i)$ , and  $n''(r_i)$ , as their analytic expressions can be easily obtained from Eqs. (3), (4), and (5).

In order to estimate the orbit-averaged relaxation timescale, we compute the orbit averaged diffusion coefficient  $D_o[\Delta R]$  as

$$D_o[\Delta R] = \lim_{R \rightarrow 0} \frac{D_{RR}(\mathcal{E}, R)}{4\pi^2 J_c^2(\mathcal{E}) P(\mathcal{E}, R)}, \quad (\text{A.14})$$

where  $D_{RR}$  is one of the coefficients of the orbit-averaged Fokker-Planck equation in the flux conservation form. It can be computed as

$$D_{RR} = \frac{256}{3} \pi^4 G^2 \log \Lambda R J^2 \times \sum_\alpha m_\alpha^2 \int_{r_-}^{r_+} \frac{v dr}{v_r} \left[ \left( \frac{2}{v_t^2} + \frac{2v^2}{v_c^4} - \frac{4}{v_c^2} \right) \mathcal{J}_{1,\alpha} + \left( \frac{3}{v_t^2} - \frac{3}{v^2} \right) \mathcal{I}_{2,\alpha} + \left( \frac{3}{v^2} - \frac{1}{v_t^2} + \frac{2v^2}{v_c^4} - \frac{4}{v_c^2} \right) \mathcal{I}_{4,\alpha} \right], \quad (\text{A.15})$$

where  $v_c = J_c/r$  is the circular velocity at a given energy.

## Appendix B: From $\delta v_{\parallel}$ and $\delta v_{\perp}$ to $\delta \mathcal{E}$ and $\delta J$

The aim of this appendix is to show how we can express the change in energy and angular momentum of the orbiting stellar-mass BH as a function of its velocity change in the parallel and perpendicular directions relative to its original motion.

At first, the velocity of the stellar-mass BH is

$$\mathbf{v} = v_r \hat{\mathbf{e}}_r + v_t \hat{\mathbf{e}}_t, \quad (\text{B.1})$$

where  $\hat{\mathbf{e}}_r$  and  $\hat{\mathbf{e}}_t$  respectfully represent the radial and tangential unit vectors of the orbit.

The variation of velocity  $\delta \mathbf{v}$  due to a stochastic velocity kick is

$$\delta \mathbf{v} = \left( \delta v_{\parallel} \frac{v_r}{v} - \delta v_{\perp} \sin \theta \frac{v_t}{v} \right) \hat{\mathbf{e}}_r + \left( \delta v_{\parallel} \frac{v_t}{v} + \delta v_{\perp} \sin \theta \frac{v_r}{v} \right) \hat{\mathbf{e}}_t + \delta v_{\perp} \cos \theta \hat{\mathbf{e}}_3, \quad (\text{B.2})$$

where  $\hat{\mathbf{e}}_3$  is the unit vector in the direction perpendicular to the orbital plane and  $\theta$  is the angle between a direction perpendicular to the motion of the stellar-mass BH and  $\hat{\mathbf{e}}_3$ . In practice, in our implementation this angle is always uniformly selected in the range  $[0, 2\pi)$ , thus we do not properly define the direction of  $\delta v_{\perp}$  here.

We can now compute the change in energy given  $\delta v_{\parallel}$  and  $\delta v_{\perp}$ :

$$\begin{aligned} \delta \mathcal{E} &= -\frac{(\mathbf{v} + \delta \mathbf{v})^2}{2} + \frac{v^2}{2} = -\frac{(\delta \mathbf{v})^2}{2} + \mathbf{v} \cdot \delta \mathbf{v} \\ &= -v \delta v_{\parallel} - \frac{1}{2} (\delta v_{\parallel})^2 - \frac{1}{2} (\delta v_{\perp})^2. \end{aligned} \quad (\text{B.3})$$

The vector change of  $\mathbf{J}$  is instead

$$\delta \mathbf{J} = r \hat{\mathbf{e}}_r \times \delta \mathbf{v} = -r \delta v_{\perp} \sin \theta \hat{\mathbf{e}}_t + r \left( \frac{v_t}{v} \delta v_{\parallel} + \frac{v_r}{v} \delta v_{\perp} \cos \theta \right) \hat{\mathbf{e}}_3, \quad (\text{B.4})$$

from which follows that  $J = |\mathbf{J}|$  changes as

$$\begin{aligned} \delta J &= |\mathbf{J} + \delta \mathbf{J}| - |\mathbf{J}| \\ &= J \left( \frac{\delta v_{\parallel}}{v} + \delta v_{\perp} \frac{v_r}{v v_t} \sin \theta + \frac{1}{2} \left( \frac{\delta v_{\perp} \cos \theta}{v_t} \right)^2 \right) + \mathcal{O}((\delta v)^3), \end{aligned} \quad (\text{B.5})$$

where we used  $J = r v_t$ .

More detailed steps can be found in Broggi (2024).<sup>13</sup>

## Appendix C: From $\Delta \mathcal{E}$ and $\Delta J$ to $\Delta v$

The aim of this appendix is to show how we can kick the stellar-mass BH to recover a given change in energy and angular momentum. Crucially, we wish to achieve this only by changing the velocity of the stellar-mass BH, while keeping its position the same.

At first, the velocity of the stellar-mass BH is

$$\mathbf{v} = v_r \hat{\mathbf{e}}_r + v_t \hat{\mathbf{e}}_t, \quad (\text{C.1})$$

where  $\hat{\mathbf{e}}_r$  and  $\hat{\mathbf{e}}_t$  respectfully represent the radial and tangential unit vectors of the orbit. The most generic velocity kick we can give to the stellar-mass BH is

$$\Delta \mathbf{v} = \Delta v_r \hat{\mathbf{e}}_r + \Delta v_t \hat{\mathbf{e}}_t + \Delta v_3 \hat{\mathbf{e}}_3, \quad (\text{C.2})$$

<sup>13</sup> The expression for  $\delta J$  reported in Broggi (2024) mistakenly reports an extra term proportional to  $\delta v_{\parallel} \delta v_{\perp}$ .

where  $\hat{\mathbf{e}}_3$  is the unit vector in the direction perpendicular to the orbital plane. Thus, the new velocity of the stellar-mass BH  $\mathbf{v}_1$  is

$$\mathbf{v}_1 = (v_r + \Delta v_r) \hat{\mathbf{e}}_r + (v_t + \Delta v_t) \hat{\mathbf{e}}_t + \Delta v_3 \hat{\mathbf{e}}_3. \quad (\text{C.3})$$

In Newtonian dynamics,  $\mathbf{J} = \mathbf{r} \times \mathbf{v}$ . Thus,

$$\mathbf{J} = r v_t \hat{\mathbf{e}}_3 \quad (\text{C.4})$$

and

$$\mathbf{J}_1 = \mathbf{r} \times \mathbf{v}_1 = -r \Delta v_3 \hat{\mathbf{e}}_t + r (v_t + \Delta v_t) \hat{\mathbf{e}}_3, \quad (\text{C.5})$$

where  $\mathbf{J}_1$  represents the angular momentum after the change in velocity. Moreover, the energy will change from

$$\mathcal{E} = -\frac{v^2}{2} + \psi(r) \quad (\text{C.6})$$

to

$$\mathcal{E}_1 = -\frac{v_1^2}{2} + \psi(r). \quad (\text{C.7})$$

Now, we write the following system:

$$\begin{cases} \mathcal{E}_1 = \mathcal{E} + \Delta \mathcal{E} \\ J_1 = J + \Delta J \\ \frac{\mathbf{J}_1 \cdot \mathbf{J}}{J_1 J} = \cos \beta \end{cases}, \quad (\text{C.8})$$

where  $\beta$  is the angle between vectors  $\mathbf{J}_1$  and  $\mathbf{J}$ . Using the relations found above, we can rewrite the system as

$$\begin{cases} -\frac{1}{2} \sqrt{(v_r + \Delta v_r)^2 + (v_t + \Delta v_t)^2 + (\Delta v_3)^2} = -\frac{1}{2} v^2 + \Delta \mathcal{E} \\ r \sqrt{(\Delta v_3)^2 + (v_t + \Delta v_t)^2} = J + \Delta J \\ \frac{r(v_t + \Delta v_t)}{J + \Delta J} = \cos \beta \end{cases}. \quad (\text{C.9})$$

In order to solve this system, we need to know the value of  $\beta$ . In practice, during our simulations we randomly drew it from the interval  $[0, 2\pi)$ , thus assuming an isotropic distribution.

Finally, after a sequence of substitutions, the system above can be written as

$$\begin{cases} \Delta v_r = \sqrt{v^2 - 2\Delta \mathcal{E} - \left( \frac{J + \Delta J}{r} \right)^2} - v_r \\ \Delta v_t = \frac{(J + \Delta J) \cos \beta}{r} - v_t \\ \Delta v_3 = \frac{(J + \Delta J) \sin \beta}{r} \end{cases}. \quad (\text{C.10})$$