The Astrophysical Journal, 820:129 (29pp), 2016 April 1 doi:10.3847/0004-637X/820/2/129

STEADY-STATE RELATIVISTIC STELLAR DYNAMICS AROUND A MASSIVE BLACK HOLE

Ben Bar-Or and Tal Alexander
Department of Particle Physics & Astrophysics, Weizmann Institute of Science, P.O. Box 26, Rehovot 76100, Israel
Received 2015 August 6; accepted 2016 February 18; published 2016 March 30

ABSTRACT

A massive black hole (MBH) consumes stars whose orbits evolve into the small phase-space volume of unstable orbits, the "loss cone," which take them into the MBH, or close enough to interact strongly with it. The resulting phenomena, e.g., tidal heating and disruption, binary capture and hyper-velocity star ejection, gravitational wave (GW) emission by inspiraling compact remnants, or hydrodynamical interactions with an accretion disk, can produce observable signatures and thereby reveal the MBH, affect its mass and spin evolution, test strong gravity, and probe stars and gas near the MBH. These continuous stellar loss and resupply processes shape the central stellar distribution. We investigate relativistic stellar dynamics near the loss cone of a non-spinning MBH in steady state, analytically and by Monte Carlo simulations of the diffusion of the orbital parameters. These take into account Newtonian mass precession due to enclosed stellar mass, in-plane precession due to general relativity, dissipation by GW, uncorrelated two-body relaxation, correlated resonant relaxation (RR), and adiabatic invariance due to secular precession, using a rigorously derived description of correlated post-Newtonian dynamics in the diffusion limit. We argue that general maximal entropy considerations strongly constrain the orbital diffusion in steady state, irrespective of the relaxation mechanism. We identify the exact phase-space separatrix between plunges and inspirals, and predict their steady-state rates. We derive the dependence of the rates on the mass of the MBH, show that the contribution of RR in steady state is small, and discuss special cases where unquenched RR in restricted volumes of phase-space may affect the steady state substantially.

Key words: black hole physics – galaxies: nuclei – stars: kinematics and dynamics

1. INTRODUCTION

1.1. Background

The strong interactions of stars with a massive black hole (MBH) in a galactic center lead to a variety of extreme phenomena and provide mass for the growth and evolution of the MBH. The small phase-space volume of orbits, whose periapsis lies close enough to the MBH to lead to a strong interaction, is called the loss cone (Frank & Rees 1976), since in most cases the interaction destroys the star, either immediately (for example, by a direct plunge through the event horizon or by tidal disruption outside it; Rees 1988), or gradually, after the orbit decays by some dissipation mechanism (for example, by the emission of gravitational waves, GW, Hils & Bender 1995; tidal heating of the star by the MBH, Alexander & Morris 2003; or drag against a massive accretion disk, Ostriker 1983). Even when the star survives the encounter, for example, in a tidal scattering event (Alexander & Livio 2001), the restricted set of orbits that allow such near-misses lie very close to the loss-cone phase-space. Since stars typically do not survive long on loss-cone orbits, the key questions are how and at what rate are these orbits repopulated by new stars. The stellar dynamical study of this question is known as loss-cone theory. A main repopulation channel is by dynamical relaxation mechanisms, which randomize stable orbits and causes them to diffuse in phase-space into the loss cone.[1] The close interaction event rates in the steady state of dynamically relaxed systems are of particular interest, both because these can be derived from first principles independently of initial conditions and because these correspond, statistically, to the cases most likely to be observed.

[1] Other possibilities include, for example, in situ star formation or galaxy mergers.

Past studies of the loss cone can be broadly categorized by four criteria: whether they deal with processes that lead to immediate stellar destruction (infall) or a gradual one (inspiral); whether they are strictly Newtonian or also include general relativity (GR), fully or perturbatively; whether they include only slow non-coherent two-body relaxation (i.e., non-resonant relaxation, NR; Chandrasekhar 1944) or also fast coherent relaxation (known as resonant relaxation, RR; Rauch & Tremaine 1996, see Section 2); and finally by the calculation methods employed, i.e., whether they are analytical, or based on the diffusion approximation either by direct numerical solutions of the Fokker–Planck (FP) equations or by Monte Carlo (MC) methods, or whether they employ direct *N*-body simulations.

Early studies focused on the infall rates of tidal disruption events in the Newtonian approximation, using analytic and FP-based methods (Frank & Rees 1976; Young et al. 1977; Cohn & Kulsrud 1978; Shapiro & Marchant 1978). These studies were subsequently updated and generalized to include some deviations from spherical symmetry (Magorrian & Tremaine 1999; Syer & Ulmer 1999; see the review by Alexander 2012). The prospect of detecting low-frequency GWs from compact remnants spiraling into MBHs (extreme mass ratio inspiral (EMRI) events) with long-baseline space-borne GW detectors (Amaro-Seoane et al. 2007) motivated studies of the inspiral rates for EMRIs using FP-based methods in the NR-only limit with perturbative GR (Hils & Bender 1995; Sigurdsson & Rees 1997; Miralda-Escudé & Gould 2000; Freitag 2001, 2003; Ivanov 2002; Hopman & Alexander 2005, 2006b; see the review by Sigurdsson 2003). The effects of the MBH spin on the EMRI rates were also considered (Amaro-Seoane et al. 2013).

A unifying framework relating plunge and inspiral processes was formulated by Alexander & Hopman (2003) and used to

estimate infall and inspiral event rates in the Galactic Center in the NR-only limit: infall by direct plunge and tidal disruption, inspiral by GW emission and tidal heating (Alexander & Morris 2003), and tidal scattering events (Alexander & Livio 2001). Different attempts to estimate the infall and inspiral rates have yielded a broad, uncertain range of values that spans several orders of magnitudes (Sigurdsson 2003; Alexander 2012).

Fast relaxation by RR can be effective on the small spatial scales where most EMRIs originate and, importantly, where stellar orbits are observed in the Galactic Center and can thus provide empirical constraints (Hopman & Alexander 2006a). This realization motivated a re-evaluation of relaxation processes and their impact on dynamics very close to MBHs. An approximate comparison of the relative rates of RR and NR suggested that the branching ratio between plunges and inspirals depends strongly on the efficiency of RR, which was then poorly understood (Hopman & Alexander 2006a; Eilon et al. 2009). This added a yet larger uncertainty to EMRI rate estimates. A key question is the physical origin and characteristics of the quenching mechanism that perturbs the near-Keplerian symmetry generating RR and causes the orbits to drift in phase-space from their initial values.

Initial analysis of RR in the relativistic context (Rauch & Tremaine 1996) indicated that rapid GR precession on very eccentric orbits likely plays a key role in quenching RR. Importantly, GR quenching can prevent RR from rapidly pushing all of the stars into plunge orbits, thereby allowing slow inspiral to produce detectable periodic GW signals (EMRIs; Hopman & Alexander 2006a). In these earlier studies, the deterministic GR precession was treated as an effective stochastic perturbation of the Keplerian orbits.

The first indications that precession cannot be treated that way, and that the long-timescale behavior of RR is not well described as a Markov process (random walk), were uncovered in post-Newtonian small N-body simulations of direct plunge and GW inspiral events (Merritt et al. 2011, hereafter MAMW11). These revealed oscillatory orbital behavior at high eccentricities that appeared to act as a barrier against further evolution to even higher eccentricities, which can lead to infall or inspiral. MAMW11 dubbed this dynamical phenomenon the Schwarzschild Barrier (SB), and showed that the oscillations are well approximated by the simple *ansatz* of assuming two-body GR dynamics in the presence of a randomly oriented fixed force vector (Alexander 2010) representing the residual force due to the background stars. While the effect appeared to be related to the EMRI-preserving RR quenching predicted by Hopman & Alexander (2006a), its magnitude seemed much stronger than anticipated, in that it not only damped the RR torques, but actually appeared to prevent the orbits from interacting closely with the MBH at all. The larger-scale relativistic N-body simulations of (Brem et al. 2014, hereafter BAS14) confirmed that GR precession quenches RR roughly on the scale of the SB, and concluded that the resulting EMRI rates are consistent with those predicted by assuming dynamics driven only by NR.

The SB phenomenon was subsequently explained rigorously in terms of the adiabatic invariance (AI) of the angular momentum by fast GR precession against the coherent RR torques when the precession period is shorter than the typical RR coherence time (the η-formalism, Bar-Or & Alexander 2014; see the review by Alexander 2015). By describing the RR torques due to the background stars in terms of a correlated noise field, it is possible to formulate an effective FP description for RR that takes into account AI and to derive the corresponding effective diffusion coefficients (DCs), whose form and behavior depend critically on the assumed temporal smoothness of the noise model. The continuous orbital evolution of the stellar background suggests that the physically correct form of the stochastic torques is that of a smooth (infinitely differentiable, C^{∞}) noise. The AI is maximal for smooth noise. In that case, its dynamical effect can be described as a faster-than-exponential suppression of the diffusion coefficients below some critical angular momentum limit. The vanishing phase-space density past this limit grows so slowly ($\sim\log t$) that the limit can be considered as an effective barrier fixed in time. While this limit is not a true barrier, or a reflecting one, it does effectively divide phase-space into a region where RR can be efficient and a region where it cannot. As we show below, the unavoidable presence of the competing process of NR substantially limits the significance of AI in the dynamics of the loss cone on long timescales (of the order of the NR relaxation time).

This study focuses mainly on the implications of NR and RR around an MBH for loss rates. However, these dynamical mechanisms are also relevant for understanding and modeling other processes around MBHs and, in particular, the Galactic MBH, SgrA*. Although the inner Galactic Center contains a relatively small and manageable number of stars by the standards of current Newtonian N-body codes, it is still very challenging to simulate it directly, both because of the extreme dynamical range introduced by the high MBH to star mass ratio and because of the added complexity of the GR equations of motion. The impact of MBH spin and RR on orbital tests of GR in the Galactic Center were studied with post-Newtonian, small N-body simulations (Merritt et al. 2010). A study of the implications of RR for the formation mechanisms of the of stars orbiting SgrA* either resorted to large Newtonian-only, N-body simulations (Perets et al. 2009) or substantially underestimated the efficiency of AI in quenching RR by using an MC scheme based on the simple fixed force *ansatz* with non-differentiable (C^0) noise to study the implications of SB for stars in the Galactic Center (Antonini & Merritt 2013; Antonini 2014).

1.2. Objectives and Overview

The objectives of this study are as follows: to integrate the recent insights about the role correlated noise plays in determining the properties of RR and its formulation as an effective diffusion process (Bar-Or & Alexander 2014) together with the known properties of NR; to derive a rigorous computational framework for calculating the steady-state phase-space density near the relativistic loss cone and the resulting loss rates; and to use this framework for a systematic study of the dependence of the results on the various physical mechanisms involved in the dynamics, i.e., mass (Newtonian) precession, GR precession, GW dissipation, the RR noise model, and coherence time. The ultimate objective is to provide well-defined estimates of the infall and inspiral rates (including, but not limited to, direct plunges, tidal disruptions, and EMRIs) and their scaling with the properties of the galactic nucleus (MBH mass and stellar density). These can then inform design decisions about planned surveys and experiments and serve as benchmarks for more detailed future studies.

We focus our study on a simplified galactic nucleus containing a stationary non-spinning (Schwarzschild) MBH surrounded by a Keplerian, spherically symmetric (in the time-averaged sense), power-law cusp of single-mass stars (the background cusp). Direct relativistic N-body simulations generate, by construction, the correct dynamics but are presently limited by computational costs to unrealistically small N, which generally cannot be scaled up to astrophysically relevant values since different dynamical mechanisms scale differently with N (e.g., Heggie & Hut 2003). Moreover, they allow little freedom to switch on/off the various physical mechanisms that affect the outcome. It is therefore difficult to disentangle their contributions and interpret the results.

Here, we follow a different approach. We represent the evolution of the system in the realistic large-N limit as a superposition of diffusion processes. This enables us to isolate and study the effect of the different dynamical mechanisms, and thereby obtain an analytic description of the system.

We calculate the loss-cone phase-space density and loss rates by two complementary methods. We show that the diffusion in phase-space is well approximated as a separable process: fast diffusion in angular momentum superposed on a slow diffusion in energy. We then use this separation of timescales to derive analytically the steady-state properties of the system. We also solve the diffusion in energy and angular momentum phase-space numerically through MC simulations, which are statistics-limited but have the advantage of flexibility in introducing additional dynamical effects and constraints. We cross-validate these two calculation methods, compare the MC results to the N-body loss rates of MAMW11 and BAS14, and reproduce the AI effects of Bar-Or & Alexander (2014) in the absence of NR.

This paper is organized as follows. In Section 2, we review the dynamical process of relaxation near a MBH in a galactic nucleus. We present a unified framework for describing both non-coherent two-body relaxation and coherent RR. We discuss the role of secular processes in the emergence of adiabatic invariance in the long-term orbital evolution of the system. In Section 3, we describe the structure and properties of phase-space near the loss cone and derive analytic estimates for the steady-state distribution and loss rates. We start, in Section 3.1, by formulating the diffusion equations which govern the evolution of the system. In Section 3.2, we describe the diffusion process in terms of the streamlines of the probability flow, which provide a powerful visual representation of the dynamics and guides us in Sections 3.3 and 3.4 in solving the steady-state distribution and loss rates (at this stage, without RR or GW dissipation). In Section 3.5, we show that GW dissipation separates the probability flow into two distinct regions in phase-space: a region where stars can inspiral into the MBH while emitting GWs and a region where stars plunge directly into it. The inspiral event rate is then calculated exactly by locating the separatrix that demarcates the two regions (Appendix A). Finally, in Section 3.6, we show that RR has a small impact on the steady-state density and loss rates, and provide a method to quantify its effect. In Section 4, we briefly describe our MC procedure for modeling orbital evolution in phase-space (a full description is provided in Appendix B; the derivation of the NR DCs is summarized in Appendix C; the effective RR DCs are derived in detail in Bar-Or & Alexander 2014). In Section 4.1, we validate the implementation of RR and the emergence of AI in angular momentum-only simulations against the analytic results of Bar-Or & Alexander (2014), and show that AI is very efficiently suppressed by NR on long timescales. We also show that over short timescales, AI induces the "Schwarzschild Barrier" phenomenon seen in angular momentum and energy phase-space (Merritt et al. 2011), and demonstrate that this dynamical feature is erased over long timescales. We compare in Section 4.2 the MC code against the small N-body loss rates of MAMW11 and BAS14, and show that the MC and N-body give consistent results (the derivation of steady-state rate estimates from MC simulations is described in Appendix B.3). In Section 5, we explore the robustness of the MC-derived rates under various dynamical approximations and assumptions. In Section 5.1, we estimate the rates for the prototypical target of low-frequency GW, the Galactic Center. In Section 5.2, we derive analytically, and confirm with the MC simulations, the weak scaling of the rates with the MBH mass. We discuss and summarize our results in Section 6. In Section 6.1, we focus on the role of the principle of maximum entropy (Appendix E) as a guiding principle in the derivation of the DCs. We argue that RR typically does not play a major role in the steady-state dynamics of the loss cone. We illustrate this analysis by presenting a fine-tuned, idealized counter-example where RR may substantially affect the loss rates: the interaction of icy planetesimals with a massive circumnuclear accretion disk. We conclude by discussing the limitations of our analysis in Section 6.2.

2. RELAXATION AROUND AN MBH

Here, we present an overview of dynamical relaxation around a central MBH and, in particular, of RR using two complementary approaches. The first, which is closer to the conventional description of the subject (e.g., Rauch & Tremaine 1996; Hopman & Alexander 2006a), contrasts stochastic two-body relaxation in the impulsive regime with global coherent RR. This serves to introduce the basic terms and ideas, and connect the present work to past studies. The second approach focuses rather on the commonalities of these two limits of relaxation, and connects the standard description of NR as a diffusion process to the recent representation of RR as an effective diffusion process, which is a key tool used here to investigate loss-cone dynamics in the large-N limit.

2.1. Two-body Relaxation

Consider a spherical stellar system composed of stars of mass $M_\star$ orbiting a central massive object $M_\bullet$ such that $Q = M_\bullet/M_\star \gg 1$, and focus on a test star at radius r from the center, where the local stellar number density is $n_\star(r)$. The average net force exerted on the test star by the $dN_\star(<b) \sim n_\star b^2 db$ background stars in a thin, small shell around it with radius $b \ll r$ and width $db \ll b$ is zero. However, the Poisson fluctuations in the positions of these $dN_\star$ discrete masses generate a residual specific force of magnitude $\sqrt{\langle F^2\rangle} \sim \sqrt{dN_\star}\, GM_\star/b^2$. This force persists in direction and magnitude until the stars generating it move substantially. For a random stellar velocity field with velocity dispersion $\sigma^2 \sim GM_\bullet/r$, this coherence time is $T_c^{\rm NR} \sim b/\sigma \ll r/\sigma \sim P$, where P is the orbital period. Since $T_c^{\rm NR} \ll P$, the net encounter is impulsive—a collision—and so, in the case of NR, the coherence time is the short collision time. The change in velocity due to the residual force over time $T_c^{\rm NR}$ is $\delta v \sim \sqrt{\langle F^2\rangle\, T_c^{\rm NR}}$. Over times $t > T_c^{\rm NR}$, these

Table 1
A Unified Framework for Relaxation Processes (See Text)

Process	Effective Particles	Averaged Quantity	Conserved Quantities	Coherence Time T_c	Residual Force Magnitude $\sqrt{\langle F^2\rangle}$	Relaxation Time $v^2/(\langle F^2\rangle T_c)$
NR	Points	None	None	$T_c^{\rm NR} < P$	$\sim\sqrt{N_\star}GM_\star\sqrt{Q}/a^2$[b]	$\sim Q^2P/(N_\star\log Q)$[b]
sRR	Ellipses	Mean anomaly	E	$P < T_c^{\rm sRR} \sim T_M < T_p$[a]	$\sim\sqrt{N_\star}GM_\star/a^2$	$\sim QP$
vRR	Annuli	Argument of periapsis	E, J	T_p [(a)] $< T_c^{\rm vRR} \sim T_{\rm sq}$	$\sim\sqrt{N_\star}GM_\star/a^2$	$\sim QP/\sqrt{N_\star}$

Notes.
[a] T_p is the precession period.
[b] Integrated over all impact parameters.

impulses add non-coherently and the accumulated change in velocity per unit time is $\langle\Delta v^2\rangle_t \sim (G^2M_\star^2 n_\star/\sigma)db/b$. Integration over all shells from $b_{\rm min}$ to $b_{\rm max} < r$ (assuming $n_\star$ is constant over this range) yields the NR diffusion timescale

$$T_{\rm NR} \sim \sigma^2/\langle\Delta v^2\rangle_t \sim \sigma^3/(G^2M_\star^2 n_\star \log\Lambda),$$

where $\Lambda = b_{\rm max}/b_{\rm min}$ is the Coulomb factor. These local changes in the test star's velocity lead to changes in orbital energy and angular momentum at a rate $\langle\Delta E^2\rangle/E^2 \sim \langle\Delta J^2\rangle/J_c^2 \sim 1/T_{\rm NR}$, where $J_c = \sqrt{GM_\bullet a}$ is the circular angular momentum and a is the semimajor axis (sma).

2.2. Resonant Relaxation

In a nearly symmetric potential, where the background orbits are nearly fixed on timescales of $T_c^{\rm RR} \gg P$, the test star interacts over a long period of time with all of the phase-averaged background orbits, instead of only instantaneously with the segment of the orbit closest to it. That is, the interaction with the background residual forces is non-local. The coherence time is set by the fastest dynamical mechanism that effectively randomizes the background orbits. Here, the relevant possibilities are Newtonian mass precession due to enclosed stellar mass, in-plane precession due to general relativity, or the mutual residual forces themselves. As in the NR case, the discrete number of stars on the scale of the test star's orbit, $N_\star(<r)$, gives rise to random fluctuations in the specific force on it, $\sqrt{\langle F^2\rangle} \sim \sqrt{N_\star(<r)}GM_\star/r^2$, which persist on timescales of $T_c^{\rm RR}$. However, unlike in the NR case, $T_c^{\rm RR} \gg P$, and so the mean anomaly, and hence the orbital energy, is adiabatically conserved (see below) and the residual force affects only the orbital angular momentum. On timescales longer than $T_c^{\rm RR}$, the accumulated changes in angular momentum due to the residual forces add non-coherently.

2.3. A Unified Description of Relaxation

This description of NR and RR highlights the fact that in both cases evolution is driven by short-timescale coherent processes that effectively contribute as stochastic processes on longer timescales. The hierarchy of coherence times offers a unifying framework (summarized in Table 1) for describing both relaxation processes in terms of orbital averaging, and for describing the dynamics in the full range between the two limits where the test star's precession is either statistically similar to that of the background or much faster than it in terms of adiabatic invariance.

We begin by making the simplifying assumptions that the background is independent of the test star and that the temporal correlations in the background dominate over phase-space correlations (Bar-Or & Alexander 2014). First, consider the case where the test star is statistically indistinguishable from the background stars (this was the focus of early works on RR). In this case, the test star and the background stars have similar orbital and precession periods. Therefore, on timescales longer than the orbital or precession period, the Hamiltonian can be double-averaged over the corresponding angle and the associated action is conserved.

The treatment of two-body relaxation is based on the approximation that when two stars interact impulsively (as in the case where the impact parameter b is much smaller than the sma a of the orbit around the MBH), the interaction time is effectively limited to the crossing time of the closest approach. Since during the interaction the force on the test star is nearly constant, the duration of the interaction (the collision timescale) is the coherence timescale $T_c^{\rm NR} \sim b/\sigma$ of two-body relaxation. As long as only interactions with small enough $b < a$ are considered so that $T_c^{\rm NR} < P(a)$, the Hamiltonian cannot be orbit-averaged, and therefore the interaction can change any of the orbital elements. Since in this case the interaction time is shorter than any other relevant timescale in the system, individual collisions are uncorrelated. In this limit, the process is Markovian (Nelson & Tremaine 1999; Bar-Or et al. 2013) and can therefore be described as diffusion in phase-space. However, interactions with a large impact parameter $b > a$ (i.e., soft encounters) such that $T_c^{\rm NR} \sim b/\sigma > P(a)$ are no longer impulsive. They cannot be described as occurring instantaneously and locally between two point particles, and therefore can no longer be described by the standard two-body relaxation formalism. In particular, the interactions are no longer Markovian because the test star is repeatedly affected by the same perturbing star.

When the test star is statistically indistinguishable from the background, then RR can be viewed as describing two-body interactions in this extreme soft limit. Since $T_c^{\rm RR} > P$, the Hamiltonian can be double-averaged over both the orbit of the test star and the orbits of the background stars. The averaged Hamiltonian is then independent of the mean anomaly, and so the Keplerian energy (sma) is conserved. The contribution of these soft encounters to energy relaxation is therefore negligible, and this is approximated as a cutoff on the maximal impact parameter via the Coulomb logarithm term (Bar-Or et al. 2013). The double-averaged Hamiltonian no longer describes point particles, but rather the interaction between Keplerian ellipses ("mass wires") which mutually torque each other and exchange angular momentum, but not energy. The force on the test ellipse by the background ellipses remains constant as long as their orbital orientations remain fixed (i.e., over the coherence time $T_c^{\rm RR}$). In analogy to the case of point-point two-body relaxation, this coherence time can be

considered as the interaction ("collision") time, which is determined by the fastest process that can reshuffle the background orbital orientations. Since the background stars are not typically on relativistic orbits, the dominant shuffling process is the retrograde in-plane drift of the argument of periapsis, ω, due to the enclosed stellar mass inside the orbits on the mass-precession timescale $T_M \sim QP/N_\star$ (e.g., Hopman & Alexander 2006a). As long as there are no competing processes with timescales shorter than T_M that could randomize the residual forces of the orbit–orbit interactions, these forces will torque the orbits and change their angular momentum in a coherent ($\propto t$) fashion. Therefore, on timescales longer than T_M, the Markovian assumption holds and the diffusion rate is $\langle\tau^2\rangle_{\rm sRR} T_M/J_c^2 \sim (QP)^{-1}$. This regime of RR is sometimes called "scalar RR" since it can change the magnitude of the angular momentum as well as its direction.

On timescales substantially longer than the precession period T_p, the Hamiltonian can also be double-averaged over the argument of periapsis, and the interaction is then between mass annuli (Kocsis & Tremaine 2015). In this case, the collision (coherence) time is the self-decoherencing (or self-quenching) time $T_c^{\rm vRR} = T_{\rm sq} \sim J_c^2/\langle\tau^2\rangle \sim QP/\sqrt{N_\star}$ on which the annuli are re-shuffled by their mutual torques. The residual torque $\langle\tau^2\rangle_{\rm vRR}$ is now also averaged over the argument of periapsis, which leads to some cancellation of the torques, and therefore $\langle\tau^2\rangle_{\rm vRR} < \langle\tau^2\rangle_{\rm sRR}$. The diffusion rate is $\langle\tau^2\rangle_{\rm vRR} T_c^{\rm vRR}/J_c^2 \sim \sqrt{N_\star}\,(QP)^{-1}$. This regime of RR is sometimes called "vector RR" since the axial symmetry of the mass annuli allows the residual torque to change only the direction of the angular momentum, but not its magnitude.[2]

Generally, the higher the degree of reshuffling (i.e., higher resulting symmetry), the longer it takes to achieve it, and hence the longer the associated coherence timescale. The higher the symmetry, the more degrees of freedom (phases) that can be averaged out of the Hamiltonian, the more symmetric the resulting effective potential, and hence the smaller the magnitude of the residual forces and the more restricted their effect. It is found that the countervailing effects of longer coherence times with smaller torques lead in balance to faster relaxation at higher symmetry, $T_r \propto v^2/(\langle F^2\rangle T_c)$, see Table 1.

We now turn to the case where the timescales of the test star and the background are different, which was not treated rigorously until recently (Bar-Or & Alexander 2014), following the discovery in N-body simulations (MAMW11) of an abrupt transition in phase-space to a different dynamical regime where orbital evolution is governed by deterministic rather than stochastic processes. In this case, the precession of the test star is due to the combined prograde GR in-plane precession with period $T_{\rm GR}$ and retrograde Newtonian precession due to the stellar mass enclosed inside the orbit. Since the GR precession rate $T_{\rm GR}^{-1}$ diverges as $1/j^2$, where $j = \sqrt{1-e^2}$ and e is the eccentricity, eccentric stars with j smaller than some critical value j_0 will precess much faster than the background, i.e., $T_{\rm GR} < T_c$. In this case, the Hamiltonian can be averaged over both the mean anomaly of the background and of the test star, as well as over the argument of periapsis ω of the test star. As a result, the test star's j is adiabatically conserved. The transition between unconstrained RR-driven diffusion at $j \gg j_0$ to strict adiabatic invariance at $j < j_0$ was calculated for several models of the effective background "noise" (residual torques) in terms of effective DCs by Bar-Or & Alexander (2014), who showed that for a smoothly varying background the transition at j_0 is extremely sharp, with the j-diffusion coefficients suppressed exponentially with the argument $(T_c/T_{\rm GR})^2 \propto (j_0/j)^4$.

In this study, we are primarily interested in processes that can deflect stable orbits into very eccentric unstable ones which can then interact strongly with the MBH. We therefore average over the orbital orientations and henceforth consider only scalar RR, referring to it simply as RR.

3. THE PHASE-SPACE OF THE LOSS CONE

We describe the dynamics of the loss cone in the Keplerian approximation where the gravitational potential of the stars is assumed to be negligible relative to that of the central MBH. In that limit, the Keplerian orbital energy is $E = GM_\bullet/2a$, where the stellar dynamical convention $E = -E_{\rm true} > 0$ for bound orbits is adopted. The orbital angular momentum is parameterized by $j = J/J_c(a)$. We marginalize the dynamics over the orbital angles and consider the evolution in the (a, j) phase-space. We assume a stationary non-spinning MBH of mass $M_\bullet$ that is surrounded by an isotropic power-law cusp of stars of mass $M_\star$, each with a number density profile $n_\star = n_0(r/r_0)^{-\alpha}$. The mass ratio is denoted as $Q = M_\bullet/M_\star$. The cusp is assumed to extend between $a_{\rm min}$ and $a_{\rm max}$. The inner boundary at $a_{\rm min}$ is an absorbing boundary, set at the innermost stable circular orbit. The outer boundary at $a_{\rm max}$ is the interface to the galaxy around the central cusp, which provides an effectively infinite reservoir of stars to replace those that are lost into the MBH or evaporate back to the galaxy. The Newtonian gravitational dynamics are described in terms of DCs, whose functional form and normalization are calculated assuming the background cusp and an isotropic distribution of angular momentum. The DCs, together with additional non-Newtonian processes such as an absorbing boundary at the last stable orbit (LSO) and GW dissipation of energy and angular momentum, are then used to generate the dynamics of test particles and to derive their steady-state loss rates and phase-space density.

Stars reach the MBH by crossing the LSO loss line in (a, j) phase-space (Figure 1) at $j_{lc}(a) = \sqrt{16 r_g/a}$, where $r_g = GM_\bullet/c^2$ (this value of $J_{lc} = J_c j_{lc} = 4 r_g c$ is exact for a zero-energy orbit). In addition, it is useful to define in the statistical sense the locus of "no-return" for GW inspiral (EMRI). Conventionally, this is defined by a comparison of timescales as the locus where the time to spiral into the MBH by the emission of GWs, $t_{\rm GW}(a, j)$, is shorter than the time needed to scatter across the LSO line by NR, $(j - j_{lc})^2 T_J(a, j)$, where $T_J \sim T_{\rm NR}$ is the J-diffusion timescale (Appendix A). Note that the GW timescale line shown in Figure 1 is calculated both using the common simplification $t_{\rm GW} = j^2 T_J$ and with the more accurate form $t_{\rm GW} = (j - j_{lc})^2 T_J$, which yields an arc-like shape that peaks well below the point where the approximate power-law GW line intersects the LSO line. The maximal sma along these lines, $a_{\rm GW}$, is then interpreted as the critical sma for EMRIs, below which phase-space trajectories cannot (statistically) avoid crossing the GW line and becoming EMRIs. The EMRI rate scales as $\propto a_{\rm GW}$ (Equation (32)). In Section 3.5 and Appendix A, we formulate a more rigorous criterion for the GW line by identifying the exact separatrix between phase-space streamlines that plunge directly into the MBH and those that inspiral into it (see Figure 4). This

[2] Both the scalar and vector RR torques scale as $\tau \propto \sqrt{N_\star} GM_\star/r$; their differences lie in the order of unity proportionality factors and the symmetry constraints on the torque orientation. Scalar RR can change both the magnitude and the direction of the angular momentum, and is thus a more general (less symmetric) case of RR than vector RR.

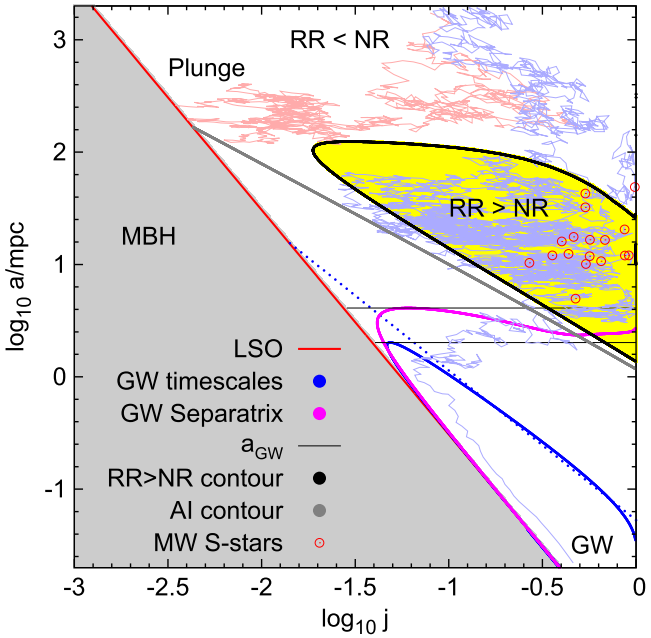

Figure 1. The (a, j) phase-space of the loss, with the various critical lines and regions, for a model of the Milky Way model with $Q = 4 \times 10^5$, mass-precession coherence time, and a Gaussian noise model (Section 5.1). Orbits in the gray area below the LSO line (red) are unstable and promptly plunge into the MBH event horizon (plunge track example in light red line). Where RR diffusion is faster than NR diffusion (yellow region), RR dominates the dynamics. The S-stars observed near the MBH of the MW (red circles; Gillessen et al. 2009b) lie in the RR-dominated region. AI suppresses RR torquing below the AI line (gray). Inside the phase-space region delimited by the GW line (blue), GW dissipation is faster than NR J scattering and orbits spiral into the MBH by the emission of GWs (inspiral track example in light blue line). The critical sma for EMRIs, a_{GW} (thin black line), corresponds to the maximum of the GW curve; below it, stars become EMRIs before they cross the LSO. The approximate power-law GW line, with the often-assumed simplification $j_{lc} \to 0$ (dotted blue line), substantially over-estimates a_{GW}. The exact separatrix streamline (magenta) provides a more accurate estimate of a_{GW} than either of the timescale-based GW lines.

results in an intermediate value of a_{GW}. These three estimates of the GW line are plotted in Figure 1 for reference, and correspond to different EMRI rate predictions. It should be emphasized that the GW line does not enter the MC procedure directly, but is an emergent property. Here, we use the separatrix method for analytic rate estimates, which accurately reproduce the MC results (Figure 17).

Here, we derive analytic estimates for the steady-state distribution and the flux of stars through the loss cone, quantify the contribution of RR to the loss rates, and validate our estimates through MC simulations.

3.1. Diffusion Equations

On sufficiently long timescales, where relaxation can be described as a diffusion process (Bar-Or et al. 2013; Bar-Or & Alexander 2014), the evolution of the probability density function, $n(E, J, t)$, in (E, J) phase-space can be describe by an FP equation:

$$\frac{\partial n(E,J,t)}{\partial t} = -\frac{\partial S_E(E,J,t)}{\partial E} - \frac{\partial S_J(E,J,t)}{\partial J}, \quad (1)$$

where the probability current densities in the E and J "directions" are

$$S_E(E,J,t) = D_E n(E,J,t) - \frac{1}{2}\frac{\partial}{\partial E}[D_{EE} n(E,J,t)] - \frac{1}{2}\frac{\partial}{\partial J}[D_{EJ} n(E,J,t)], \quad (2)$$

and

$$S_J(E,J,t) = D_J n(E,J,t) - \frac{1}{2}\frac{\partial}{\partial J}[D_{JJ} n(E,J,t)] - \frac{1}{2}\frac{\partial}{\partial E}[D_{EJ} n(E,J,t)], \quad (3)$$

and where D_{EE}, D_E, D_{JJ}, D_J, and D_{EJ} are the DCs that describe the combined effect of NR and RR.

3.2. Probability Flow in Phase-space

The effects of the various physical mechanisms are more clearly apparent in the flow patterns in phase-space. Since the physical flow is stochastic, it is more useful to describe it in terms of the flow of the probability density. In steady state, the FP equation (Equation (1)) can be written as a continuity equation of a compressible flow:

$$\frac{\partial}{\partial E}[n(E,J)v_E] + \frac{\partial}{\partial J}[n(E,J)v_J] = 0, \quad (4)$$

with effective velocities $v_E = S_E/n(E,J)$ and $v_J = S_J/n(E,J)$. The two-dimensional flow in phase-space $\boldsymbol{v} = (v_J, v_E)$ can be visualized by the streamlines[3], $v_J/dJ = v_E/dE$, which are derived below from the steady-state solution of the FP Equation (28).

Using the streamlines, in Figures 2–3, we show the effects of the various physical mechanisms. The probability current densities are determined by the distribution function (DF) of the background stars through the DCs and by the DF of the test stars. We begin by assuming that a relaxed cusp will be approximately isotropic (i.e., $f(E,J) \propto E^{1/4}$; Bahcall & Wolf 1976, hereafter BW76). The existence of a loss cone introduces a logarithmic correction, so that the DF is of the form $f(E,J) \propto E^{1/4}\log(J/J_{lc})/\log(J_c/J_{lc})$ (see Equation (28), Figures 6, 7, and Hopman & Alexander 2005). Since the DCs are not strongly affected by this small anisoptropy, it is convenient to assume that the DF of the background is isotropic. Therefore, in the calculation of the streamlines, we use a DF of the form $f(E,J) \propto E^{1/4}$ for the background and a DF of the form $f(E,J) \propto E^{1/4}\log(J/J_{lc})/\log(J_c/J_{lc})$ for the test stars.

The flow at a point in phase-space is considered to be j-dominated when the streamlines are approximately horizontal (i.e., $|\dot{j}|/j \gg |\dot{a}|/a$) and a-dominated when the streamlines are approximately vertical (i.e., $|\dot{j}|/j \ll |\dot{a}|/a$). As shown in Figures 2–4, the flow is j-dominated, apart from two restricted regions in phase-space ($J \to J_c$ and the GW-dominated region). Therefore, the full flow field can be separated into two one-dimensional flows, a fact that will be used below to simplify the analytic treatment. Since RR drives stars only in the j direction, this separation is enhanced in the phase-space region where

[3] The streamlines are immutable under coordinates transformation and therefore do not depend on the specific choice of coordinate system.

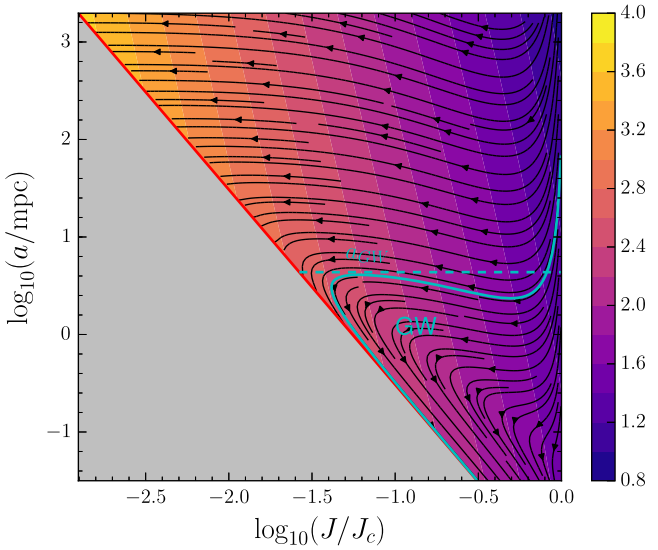

Figure 2. Streamlines of the phase-space flow $\bar{v} = (\dot{j}/j, \dot{a}/a)$. All of the dynamical effects, apart from resonant relaxation, are included (i.e., two-body relaxation, GW emission, GR precession, and mass precession). The DCs are calculated for a Milky Way-like model (isotropic cusp $f(E) \propto E^{1/4}$ with an MBH of $4 \times 10^6 M_\odot$, a mass ratio of $M_\bullet/M_\star = 5 \times 10^5$, and total stellar mass $M_\star(r_h) = 2M_\bullet$, where $r_h = 2$ pc). In addition, the probability current densities are calculated assuming the DF $f(E, J) \propto E^{1/4}(2J/J_c^2)\log(J/J_{lc})/\log(J_c/J_{lc})$, which is the steady-state solution in the presence of a loss cone. The color map describes the magnitude of the DCs. The solid cyan line is the GW separatrix. The dashed cyan line indicates the critical sma for EMRIs.

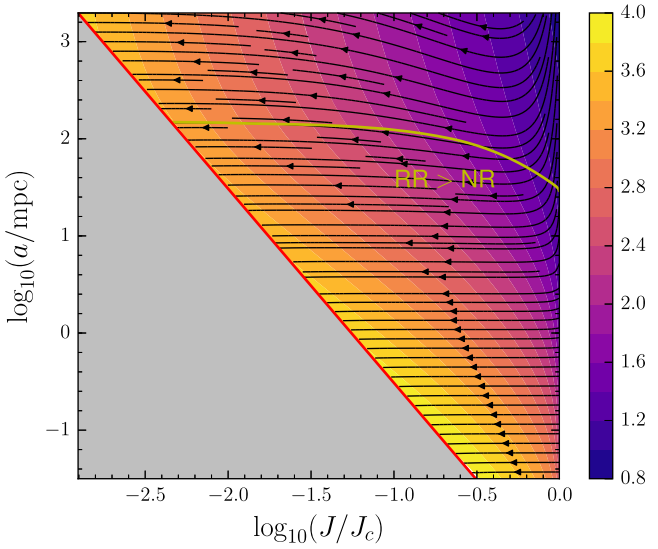

Figure 3. Streamlines of the phase-space flow $\bar{v} = (\dot{j}/j, \dot{a}/a)$. All dynamical effects apart from GR precession are included (same cusp model as in Figure 2). In this case, RR dominates the dynamics all the way to the loss cone. Note that this scenario leads to strong depletion of the cusp near the MBH (see for example Figure 15 bottom). This means that in that region, our assumption of a power-law steady-state cusp does not hold. The solid black line marks the region where RR is effective (stronger than NR). The color map describes the strength of the DCs.

$D_{jj}^{\mathrm{RR}}(a, j) > D_{jj}^{\mathrm{NR}}(a, j)$, and RR governs the dynamics (see Figures 4–3).

As shown in Figure 3, in the absence of GR precession (i.e., mass precession-only) interior to some sma, RR dominates the dynamics all the way to the loss cone. In this hypothetical case, the loss rate could be high enough to actually empty the cusp close to the MBH. This would then invalidate the assumption

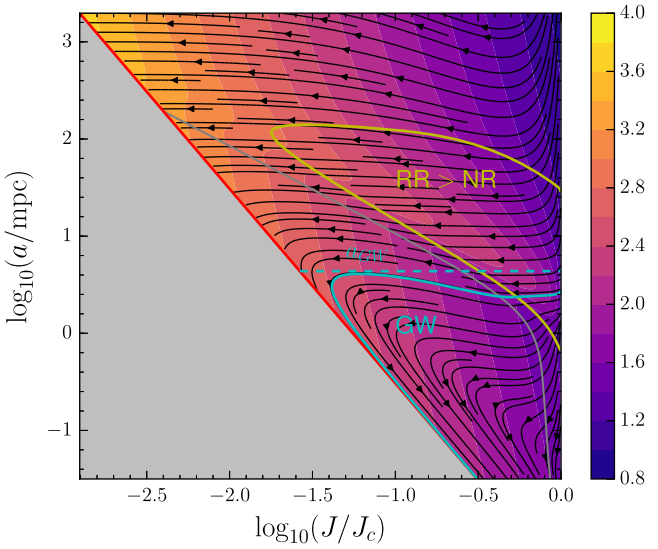

Figure 4. Streamlines of the phase-space flow $\bar{v} = (\dot{j}/j, \dot{a}/a)$. All dynamical effects are included (same cusp model as in Figure 2). The solid black line marks the region where RR is effective (stronger than NR). The gray line is the locus beyond which RR is totally ineffective due to adiabatic invariance. The color map describes the strength of the DCs. The solid cyan line is the GW separatrix. The dashed cyan line indicates the critical sma for EMRIs.

of a single power-law cusp. However, in realty, GR precession does play a crucial role in determining the dynamics and the steady state. In fact, due to GR precession, RR is totally quenched by the adiabatic invariance (AI) of the angular momentum. This happens for stars with angular momenta below the locus where the precession frequency $\nu_p(a, j) = |\nu_{\mathrm{GR}} + \nu_M|$ equals the coherence frequency $2\pi/T_c$, where ν_{GR} and ν_M are the GR and mass-precession frequencies (see AI curve in Figure 4). Only above the AI line can RR be effective. Figure 4 shows a closed contour, somewhat above the AI line, where RR is faster than NR and therefore dominates the dynamics. As we show in Section 3.6, the fact the RR is not effective near the loss lines means that RR does *not* play an important role in setting the steady state and the loss rates.

3.3. Steady-state Distribution and Loss-cone Flux

We assume that the system relaxes in J much faster than it relaxes in E. This assumption can be justified by noting that $J \in [J_{lc}, J_c(E)]$ is bound, whereas E is unbound. We therefore assume that the relaxation process is separable: on short timescales, stars exchange only angular momentum but not energy and reach their steady-state J distribution at fixed E, and only on a much longer timescales do they reach a global steady state in E. This assumption of local equilibrium (i.e., in each energy bin the J distribution is relaxed) is further supported by the pattern of the probability current densities (Equations (2), (3)) described by the streamlines shown in Figure 5 for the NR-only case. The inclusion of RR only strengthens this separability. This demonstrates that the motion in the E direction (a direction) occurs only at $j \to 1$, whereas it is almost entirely in the j direction at $j < 1$. The validity of this assumption is verified by the excellent match between our analytic predictions and the result of the MC simulations, which do not assume separability a priori, as shown in Figures 6–8. In this section, we use this separability assumption

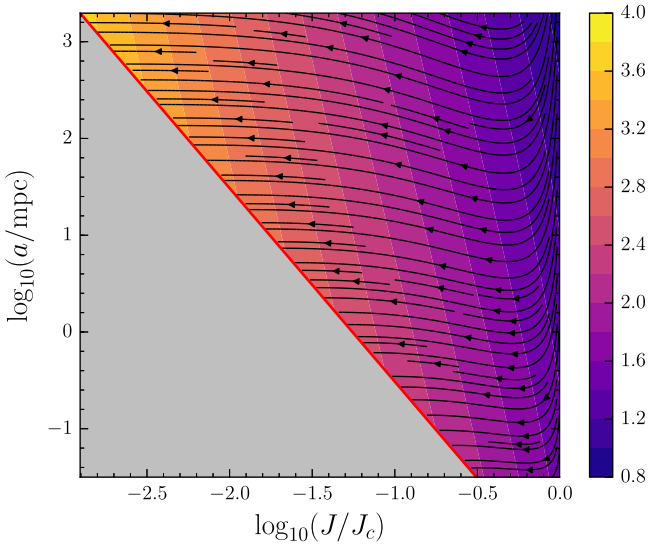

Figure 5. Streamlines of the phase-space flow $\bar{v} = (\dot{j}/j, \dot{a}/a)$. Resonant relaxation and GW emissions are not included (same cusp model as in Figure 2). The color map describes the strength of the DCs.

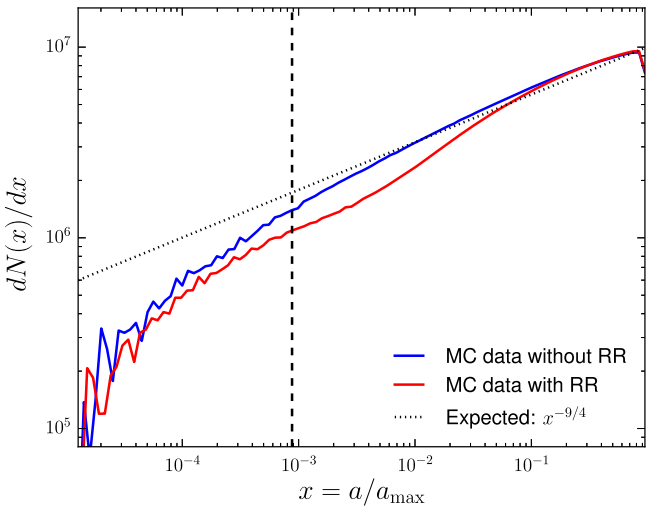

Figure 6. Energy distribution as a function of energy shows good agreement between the analytic BW76 cusp solution and MC results. The critical sma, $a_{\rm GW}(E_{\rm GW})$ as defined by the separatrix (see Section 3.5), is shown by a dashed line.

and the fluctuation–dissipation relation (Appendix E) to derive the steady-state (E, J) distribution and the loss-cone flux.

In the limit where there is no energy exchange between stars, the FP equation can be written as (Bar-Or & Alexander 2014)

$$\frac{\partial n(E,J,t)}{\partial t} = \frac{1}{2}\frac{\partial}{\partial J}\left\{JD_{JJ}(E,J)\frac{\partial}{\partial J}\left[\frac{1}{J}n(E,J,t)\right]\right\} = -\frac{\partial S_J(E,J,t)}{\partial J}, \quad (5)$$

which generally follows from the maximum entropy principle and, in fact, provides a necessary test for the validity of the DCs (Appendix E). In the absence of a loss cone (i.e., $J_{lc} \to 0$), the steady-state probability current density $S_J(E, J, t)$ is zero, and the local equilibrium distribution is isotropic:

$$n(E,J) = 2J/J_c^2(E)n(E). \quad (6)$$

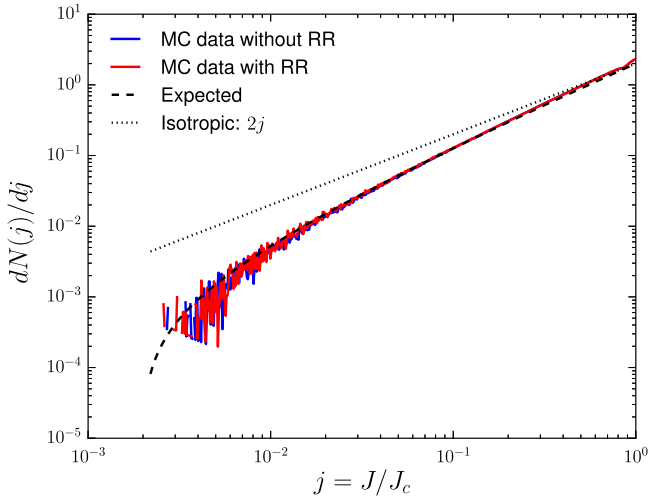

Figure 7. Angular momentum distribution in a bin centered around $a = 0.4a_{\max}$. The MC data show the expected logarithmically suppressed distribution (Equation (28)) compared with the isotropic one.

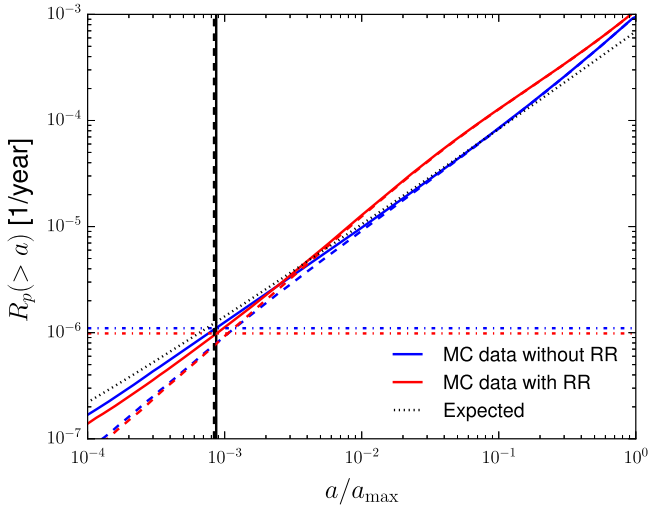

Figure 8. The plunge rate as a function of semimajor axis. The MC data (without RR) agree with the analytic expression (Equation (48)). The MC results demonstrate that the contribution of RR to the plunge rate is small. The critical sma, $a_{\rm GW}$ (vertical dashed line) as defined by the separatrix (see 3.5), agrees very well with the sma (vertical solid line) where the inspiral rate, $R_i^{\rm tot}$ (dashed–dotted lines), equals the plunge rate $R_p^{\rm no\ GW}$ in the absence of GW emission.

For a finite J_{lc}, it follows from the separability assumption that the probability current density is non-zero and independent of J, $S_J(E, J) \approx S_J(E, J_{lc})$. Therefore, from Equation (5), we obtain

$$JD_{JJ}(E,J)\frac{\partial}{\partial J}\left[\frac{1}{J}n(E,J,t)\right] = -2S_J(E,J_{lc}). \quad (7)$$

By integrating over J and using the normalization $n(E) = \int_{J_{lc}}^{J_c} n(E,J)dJ$, we obtain

$$S_J(E,J_{lc}) = -n(E)\Big/\int_{j_{lc}}^{1}\frac{1-j^2}{d_{jj}(E,j)}\frac{dj}{j}, \quad (8)$$

and

$$n(E, J) = \frac{2J}{J_c^2}\frac{n(E)}{\int_{j_{lc}}^{1}\frac{1-j^2}{d_{jj}(E,j)}\frac{dj}{j}}\int_{j_{lc}}^{j}\frac{1}{j' d_{jj}(E, j')}dj', \quad (9)$$

where $d_{jj}(E, j) = D_{JJ}(E, J)/J_c^2$.

We now assume that the system achieves local equilibrium in J at any E (Equation (9)) and derive the E-only diffusion equation. Integrating Equation (1) over J from J_{lc} to J_c, we obtain the total probability current density gradient (loss-rate) per unit energy:

$$\frac{\partial n(E)}{\partial t} = -\int_{J_{lc}}^{J_c}\frac{\partial S_E}{\partial E}dJ - S_J(E, J_c) + S_J(E, J_{lc}). \quad (10)$$

It is convenient to transform these expressions to (E, j), since for $j = J/J_c$ the current density in the j direction is zero at the boundary ($j = 1$). Generally, under the coordinate change $\boldsymbol{x} \to \boldsymbol{x}'$ (here $(E, J) \to (E, j)$), the probability density currents, $S_i \to S_i'$, transform as

$$S_i'(\boldsymbol{x}') = \left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}'}\right|\frac{\partial x_i'}{\partial x_k}S_k(\boldsymbol{x}). \quad (11)$$

Thus,

$$S_j(E, j) = S_J(E, J) - j\frac{\partial J_c}{\partial E}S_E(E, J), \quad (12)$$

and since $S_j(E, j = 1) = 0$, we have

$$\frac{\partial n(E)}{\partial t} = -\int_{J_{lc}}^{J_c}\frac{\partial S_E}{\partial E}dJ - \frac{\partial J_c}{\partial E}S_E(E, J_c) + S_J(E, J_{lc}), \quad (13)$$

and

$$\frac{\partial n(E)}{\partial t} = -\frac{\partial}{\partial E}\int_{J_{lc}}^{J_c}S_E dJ + S_J(E, J_{lc}). \quad (14)$$

Equation (2) and the fact that $D_{EJ}(E, J_c) = D_{EE}(E, J_c)\partial J_c/\partial E$ (Appendix C) allow us to obtain the J-averaged FP equation for the energy probability density $n(E)$,

$$\frac{\partial n(E, t)}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial E^2}[\bar{D}_{EE}n(E)] - \frac{\partial}{\partial E}[\bar{D}_E n(E)] + S_J(E, J_{lc}), \quad (, 15)$$

in the presence of a loss term $S_J(E, J_{lc})$, resulting from the flux of stars through the loss cone, per unit energy[4], with the J-averaged diffusion coefficients:

$$\bar{D}_E(E) = n^{-1}(E)\int_{J_{lc}}^{J_c}D_E(E, J)n(E, J)dJ, \quad (16)$$

$$\bar{D}_{EE}(E) = n^{-1}(E)\int_{J_{lc}}^{J_c}D_{EE}(E, J)n(E, J)dJ. \quad (17)$$

In steady state, $n(E)$ is obtained by solving Equation (15), given $S_J(E, J_{lc})$,

$$\frac{dR_P}{dE} = -S_J(E, J_{lc}) = \frac{\partial}{\partial E}\left\{\frac{1}{2}\frac{\partial}{\partial E}[\bar{D}_{EE}n(E)] - \bar{D}_E n(E)\right\}, \quad (18)$$

where $R_p(E) = -\int_{J_{lc}}^{J_c}S_E(E, J)dJ$ is the cumulative number of stars lost through the loss cone per unit time. Note that in steady state, the probability current density in the J direction equals the probability current density gradient in E (from continuity considerations: the density carried by the J current at fixed E and lost through j_{lc} is balanced by the E gradient of the total E current).

3.4. Steady-state Distribution for Two-body Relaxation

We now show that in the case of two-body relaxation, the solution of the energy FP equation (Equation (18)) with non-zero flux can be approximated analytically to derive the steady-state density distribution and the plunge rate. Since the plunge rate is small compared to the relaxation rate, the energy distribution asymptotes to the zero-flux (i.e., no plunges) BW76 power-law solution.

For two-body relaxation, $d_{jj}^{\rm NR}(E, j)$ asymptotes to a finite value $d_{\rm NR}^0(E) = d_{jj}^{\rm NR}(E, j = 0)$ as $j \to 0$.[5] Since most of the contribution to the current density, $S_J(E, J_{lc})$, reflects the value of d_{jj} at small j (Equation (8)), it can be approximated by $d_{jj}^0(E)$, so that

$$S_J(E, J_{lc}) \approx -n(E)\frac{2d_{\rm NR}^0(E)}{\log(J_c^2/J_{lc}^2) - 1 + J_{lc}^2/J_c^2} \approx -n(E)\frac{d_{\rm NR}^0(E)}{\log(J_c/J_{lc})}, \quad (19)$$

and

$$n(E, J) \approx n(E)\frac{2J}{J_c^2}\frac{\log(J^2/J_{lc}^2)}{\log(J_c^2/J_{lc}^2) - 1 + j_{lc}^2} \approx n(E)\frac{2J}{J_c^2}\frac{\log(J^2/J_{lc}^2)}{\log(J_c^2/J_{lc}^2)}. \quad (20)$$

Since $d_{\rm NR}^0$ scales as $\bar{D}_{EE}/E^2 = T_E^{-1}$ (Appendix C), it is convenient to represent $S_J(E, J_{lc})$ explicitly in terms of the energy relaxation time, T_E, as $S_J(E, J_{lc}) = -n(E)\chi(E)/T_E(E)$, where

$$\chi(E) = T_E d_{\rm NR}^0/\log(J_c/J_{lc}) = 2T_E d_{\rm NR}^0/\log(E_{lc}/E), \quad (21)$$

expresses the logarithmic suppression of the flux due to the decreasing size of the loss cone away from the MBH, and where $E_{lc} \equiv GM_\bullet/32r_g$ corresponds to the limit $J_c = J_{lc}$ for $J_{lc} = 4r_g c$ and for Keplerian energy. Note that this is not the true innermost stable circular orbit, but rather is a formal extrapolation of the approximations adopted here used for normalization only. Here, we are interested in stars with $E \gg E_{lc}$, where $\partial \log \chi/\partial \log E \propto \chi \propto 1/\log(E_{lc}/E)$ is small.

[4] This generalizes the simpler situation where stars are only destroyed once they reach some high energy threshold where the loss is expressed instead by a boundary condition (see BW76).

[5] Since $d_{jj}^{\rm NR} = 2jd_j^{\rm NR}$ for $j \to 0$ (e.g., Shapiro & Marchant 1978; Appendix C) and $2jd_j^{\rm NR} = \partial_j j d_{jj}^{\rm NR}$ (Appendix E).

The cumulative plunge rate $R_P = -\int S_J(E, J_{lc})dE$ can then be approximated as

$$R_p \approx \frac{1}{2}N(E)\frac{\chi(E)}{T_E} + \frac{3}{4}\int_E^{E_{\max}} \frac{\chi(E')}{T_E(E')}N(E')\frac{dE'}{E'}, \quad (22)$$

where $N(E)$ is the number of stars with energy larger than E and we used $T_E \propto P/N(E)$.

For an infinite isotropic cusp $n(E) \propto E^{p-5/2}$ where $0 < p < 1/2$ (to ensure the DCs are finite), the J-averaged DCs are (Bar-Or et al. 2013)

$$\bar{D}_E = -\frac{4(1-4p)\log\Lambda}{1+p}\frac{N(E)}{Q^2P(E)}E \propto E^{p+1}, \quad (23)$$

$$\bar{D}_{EE} = \frac{16(3-2p)\log\Lambda}{(1+p)(1-2p)}\frac{N(E)}{Q^2P(E)}E^2 \propto E^{p+2}, \quad (24)$$

and the plunge rate is

$$R_p \approx \frac{3-2p}{3-4p}N(E)\frac{\chi(E)}{T_E}. \quad (25)$$

Since χ is nearly constant in E for $E \ll E_{lc}$, it can be approximated in that limit by evaluating it at $E_{\min} = GM_\bullet/2a_{\max}$. In that case, Equation (18) has an analytical solution,

$$\frac{(4p-3)(1-4p)}{4(3-2p)} = \chi, \quad (26)$$

which connects the current to the power-law exponent of the cusp. The physical branch of the solution is[6]

$$p = \frac{1}{4}(-\sqrt{\chi^2+8\chi+1} - \chi + 2) \approx \frac{1}{4}(1-5\chi). \quad (27)$$

Since for $p = 1/4$, $d_{jj}^0 T_E \approx 1/8$ (Appendix C), it follows that $\chi \approx 1/[8\log(1/j_{lc})] \ll 1$, and so the BW76 solution $p = 1/4$ (i.e., $\chi \to 0$) is a reasonable approximation in the $E \to E_{\min}$ limit, where $j_{lc} \ll 1$. Thus, in steady state, the energy (or sma) distribution is $n(E) \propto E^{-9/4}$ (or $n(a) \propto a^{1/4}$). Using Equation (20), we obtain the (E, J) steady-state distribution

$$n(E, J) \approx \frac{5}{4}N(E_{\min})\frac{2J}{J_c^2}\frac{\log(J/J_{lc})}{\log(J_c/J_{lc})}(E/E_{\min})^{-9/4}, \quad (28)$$

and steady-state plunge rate

$$R_p(a) \approx \frac{5}{32}\frac{1}{\log(1/j_{lc}(a))}\frac{N(a)}{T_E(a)}, \quad (29)$$

where the energy relaxation time is (Bar-Or et al. 2013)

$$T_E(a) = \frac{1}{64}Q^2\frac{P(a)}{N(a)\log Q}. \quad (30)$$

In the limit $E \to E_{lc}$, $\chi(E)$ can no longer be approximated as fixed or small and the power-law solution breaks down. We now argue that this power-law approximation is valid for almost the entire range of MBH masses and their host galactic nuclei. We define $a_{\rm BW}$ as the minimal sma where χ and $\partial\log\chi/\partial\log E$ are still small enough for this approximation to hold. Since $\partial\log\chi/\partial\log E \propto \chi \propto 1/\log(1/j_{lc})$, where $j_{lc} \propto \sqrt{r_g/a}$, it follows that $a_{\rm BW} \propto M_\bullet$. We are interested in resolving the dynamics close to the MBH in order to obtain a reliable estimate for the EMRI event rate. As we show in Section 3.5, the rate is determined by the dynamics near the critical GW sma, $a_{\rm GW}$. Thus, it is sufficient to show that $a_{\rm BW} < a_{\rm GW}$, since then the power-law approximation is valid over the entire range of relevant radii. The MC results shown in Figures 6 and 8 for the case of $M_\bullet = 4\times10^6\,M_\odot$ demonstrate that our analytic estimates (Equations (28), (29)) for the steady-state energy distribution and plunge rate reproduce the simulated results at least down to $a_{\rm GW}$. This also holds for more massive MBHs. The M/σ relation and the scalings $a_{\rm BW} \propto M_\bullet$ and $a_{\rm GW} \propto (\log Q)^{-4/5} r_h$ (Equation (60)) imply $a_{\rm BW}/a_{\rm GW} \propto (\log Q)^{4/5} M_\bullet^{2/\beta}$, and therefore $a_{\rm BW} \lesssim a_{\rm GW}$ up to $M_\bullet \sim 10^9\,M_\odot$.

3.5. EMRI Event Rates

So far, we have ignored the contribution of GW emission to the dynamics. Compact objects can withstand the tidal field of the MBH. When on eccentric orbits, their orbital decay by the emission of GWs can be faster than the diffusion of angular momentum due to the stochastic perturbations of the stellar background. In that case, they inspiral gradually all the way down to the innermost stable circular orbit (ISCO) as EMRIs (Peters & Mathews 1963; Peters 1964; Gair et al. 2006), instead of plunging directly into the MBH with $J < J_{lc}$ (Figure 1). The GW signatures of plunges and inspirals are very different. The low mass of the compact objects generates weak signals, well below the noise. Plunges result in short, very hard to detect broad spectrum GW flares. In contrast, EMRIs are of special interest since their quasi-periodic signal can be integrated and detected against the noise if the waveform is approximately known.

In the absence of GWs (Figure 5), the streamlines are approximately constant in a. In contrast, GW emission diverts the streamlines to tracks that are almost parallel to the loss cone (i.e., nearly constant J) in the phase-space region where GW dominates the dynamics (Figure 2). The outermost inspiraling streamline separates phase-space into two distinct regions. Above this separatrix all streamlines are plunges, while below it all streamlines are inspirals (Figure 4). The continuity equation (Equation (4)) implies that the probability current in steady state is constant along a streamline bundle. Since the streamlines in the GW-dominated region below the separatrix originate in phase-space regions where the density is much higher, the small depletion due to EMRI losses is not expected to affect the density at the origin of the streamlines. The EMRI rate can therefore be estimated by identifying the terminal point (a_p, j_{lc}) of the plunge streamline (without GW) corresponding to the separatrix. This is the effective critical sma for EMRIs, $a_{\rm GW}$. The EMRI rate is then obtained by integrating the differential plunge rate in the absence of GW emission from $a_{\rm isco}$ to $a_{\rm GW}$,

$$R_i^{\rm tot} = -\int_{E_{\rm GW}}^{E_{\rm isco}} S_J(E, J_{lc})dE = R_p(a_{\rm GW}). \quad (31)$$

[6] This generalizes the analytic BW76 solution ($p = 1/4$, $\chi = 0$), which applies for a power-law DF in steady state with a constant E current (which then must be zero). Here, the "leakage" of stars through the loss cone at all E ($\chi > 0$) implies a non-constant current, which allows flatter cusp solutions with $p < 1/4$.

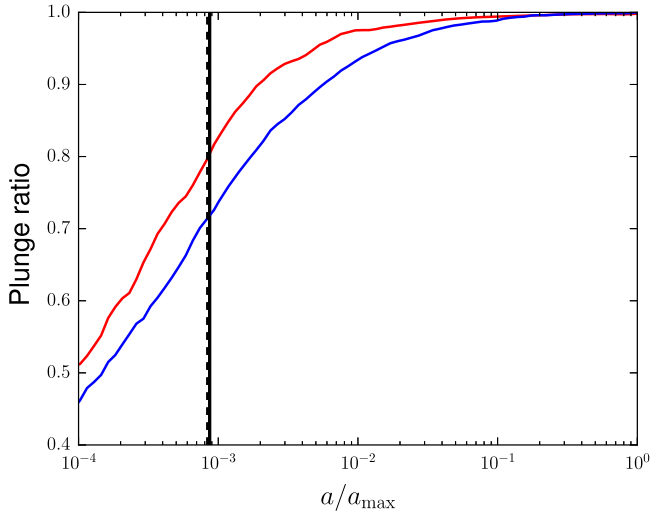

Figure 9. Plunge ratio $R_p^{\rm GW}/R_p^{\rm no\,GW}$, where $R_p^{\rm GW}$ and $R_p^{\rm no\,GW}$ are the cumulative plunge rates obtained with and without GW dissipation. The critical sma, $a_{\rm GW}$ (vertical dashed line) as defined by the separatrix (see 3.5), agrees very well with the sma (vertical solid line) where the inspiral rate, $R_i^{\rm tot}$, equals the plunge rate $R_p^{\rm no\,GW}$ in the absence of GW emission.

Thus, for a BW76 cusp, the EMRI rate is

$$R_i^{\rm tot} = \frac{5}{32}\frac{1}{\log\left(1/j_{lc}\left(a_{\rm GW}\right)\right)}\frac{N\left(a_{\rm GW}\right)}{T_E\left(a_{\rm GW}\right)} = \frac{a_{\rm GW}}{a_{\max}}\frac{\log\left(1/j_{lc}\left(a_{\max}\right)\right)}{\log\left(1/j_{lc}\left(a_{\rm GW}\right)\right)}R_p^{\rm no\,GW}\left(a_{\max}\right), \quad (32)$$

which is approximately linear in $a_{\rm GW}$. The value of $a_{\rm GW}$ is determined by solving the streamline equation $dE/dJ = S_E/S_J$ with the boundary condition that the streamline trajectory reaches J_{lc} at $E_{\max}$. The probability current densities are estimated by assuming that S_J is constant in J and S_E results only from GW dissipation. This means that in the absence of GWs, that streamline is constant in E and $a_{\rm GW}$ can be estimated by taking the value of a at $J \gg J_{lc}$. The exact value of $a_{\rm GW}$ depends on the GW emission approximation used and is calculated in Appendix A. As shown in Figure 8, this definition of $a_{\rm GW}$ is indeed a good approximation to the sma where the plunge rate (without GW) is equal to the inspiral rate and can be used to predict the inspiral rate in the MC simulations (Figure 17). As expected, inside $a_{\rm GW}$, the plunge rate with GW decreases relative to the plunge rate without GW (See Figure 9).

3.6. Effect of RR

Due to the long coherence time, RR is a much more effective process than two-body relaxation. However, in regions of phase-space where in-plane GR precession is faster than the coherence time, j becomes an adiabatic invariant and the RR process is quenched. RR is therefore limited to a small region of phase-space (see Figure 4). The locus where in-plane precession quenches RR by AI (Section 4.1) defines the outer envelope of the region where RR may be efficient relative to NR.

In the relativistic regime where the precession time is shorter than the coherence time, RR cannot be described as a simple diffusion process. The stochastic orbital evolution of the test star is determined by the temporal correlations of the background. As shown in Bar-Or & Alexander (2014), the leading-order phase-averaged Hamiltonian yields stochastic equations of motion where the effect of the stellar background is described by a noise vector in angular momentum space, $\boldsymbol{\eta}(t)$. This then corresponds to an effective diffusion term (Bar-Or & Alexander 2014),

$$D_{JJ}^{\rm RR} = J_c^2\nu_j^2 S_\eta\left(\nu_p\left(j\right)\right), \quad (33)$$

and an associated (parametric) drift term,

$$D_J^{\rm RR} = \frac{1}{2J}\frac{\partial}{\partial J}JD_{JJ}^{\rm RR}. \quad (34)$$

The RR DCs depends on the power spectrum $S_\eta(\nu_p)$ of the noise $\boldsymbol{\eta}(t)$ at the precession frequency of the test orbit, which combines the net effect of both the mass and the GR precession. The magnitude of the RR diffusion is proportional to the normalized residual torque in the $\boldsymbol{J}$ direction, $\nu_j = J_c^{-1}\sqrt{\langle\tau_J^2\rangle}$ (see Appendix D),

$$\nu_j \approx 0.28\sqrt{1-j}\sqrt{N\left(a/2\right)}\nu_r/Q. \quad (35)$$

The noise $\boldsymbol{\eta}(t)$ is an explicit function of the orbital elements of the background stars, and can therefore be measured in principle by N-body simulations. However, this has yet to be done. Here, for simplicity, we limit ourselves to a specific smooth noise with Gaussian power, $S_\eta(\nu) = 2T_c\exp\left(-T_c\nu^2/\pi\right)$, where

$$T_c = \sqrt{\pi/2}\,\nu_p^{-1}\left(2a,\sqrt{1/2}\right). \quad (36)$$

The corresponding RR diffusion coefficient is then (Bar-Or & Alexander 2014)

$$D_{JJ}^{\rm RR} = 2J_c^2 T_c\nu_j^2 e^{-4\pi j_0^4/j^4}, \quad (37)$$

where $j_0(a)$ is the AI locus where the GR precession frequency, $\nu_{\rm GR}(a,j)$, equals the coherence time

$$j_0 = \sqrt{T_c\nu_{\rm GR}\left(a,j=1\right)/2\pi}. \quad (38)$$

As we show in Section 5.1, our results do not depend strongly on the noise model, as long as it is correlated (i.e., smooth) on timescales smaller than T_c.

The region where RR dominates the dynamics even on long timescales is where the ratio of the second-order DCs[7] exceeds unity, i.e., $D_{jj,\rm RR}/D_{jj,\rm NR} > 1$ (see Figure 4). The typical phase-space configuration is shown in Figure 1: the region where RR dominates is detached from the loss lines; NR is required for the stars to evolve toward them, and therefore the slow NR timescale remains the bottleneck for the loss rates which are mostly unaffected by RR (see Figure 17). This can be shown formally by re-estimating the probability current density in the presence of RR.

The combined diffusion coefficients are

$$D_J = D_J^{\rm NR} + D_J^{\rm RR}, \quad (39)$$

$$D_{JJ} = D_{JJ}^{\rm NR} + D_{JJ}^{\rm RR}, \quad (40)$$

[7] The transformation of the DC from J to $j = J/J_c$ is $D_{jj,\rm NR} = D_{JJ,\rm NR}/J_c^2 + (j^2/4)D_{EE}/E^2 + jD_{EJ}/J_cE$ (Appendix C).

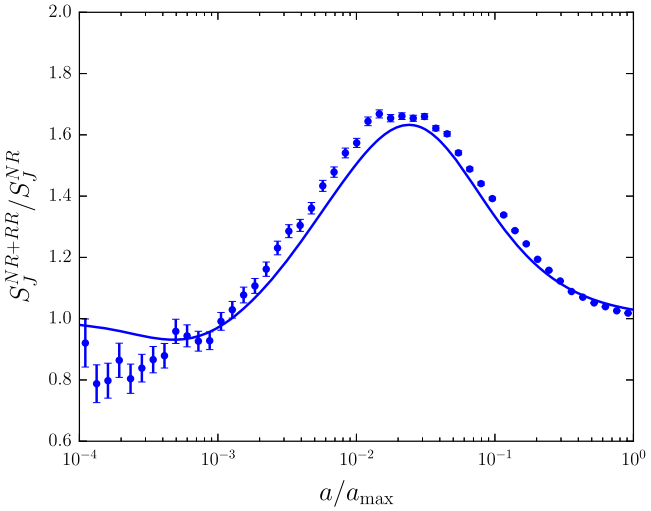

Figure 10. Ratio between the probability current density in J with and without RR, as function of a (Equation (43)), which expresses the differential plunge rate. The limited increase in the ratio over its asymptotic value of 1 in the $a \to 0$ and $a \to \infty$ limits expresses the fact that the contribution of RR to the plunge rate is small.

and using Equation (8), the flux is given by

$$S_J^{\rm NR+RR}(E, J_{lc}) = -n(E) \Big/ \int_{j_{lc}}^{1} \frac{1-j^2}{D_{JJ}^{\rm NR}/J_c^2 + D_{JJ}^{\rm RR}/J_c^2} \frac{dj}{j}. \tag{41}$$

Since $D_{JJ}^{\rm RR}$ rises up to some maximal value before it sharply drops as it approaches $j \to j_0$, we can approximate the RR DC as

$$d_{jj}^{\rm RR} \equiv D_{JJ}^{\rm RR}/J_c^2 \approx \begin{cases} d_{\rm RR}^0 (1-j) e^{-4\pi j_0^4/j^4} & j \geqslant j_m, \\ 0 & j < j_m \end{cases} \tag{42}$$

where $d_{\rm RR}^0 = 2T_c \nu_j^2 \, (j=0)$ and the maximum of $d_{\rm RR}(E, j)$ occurs at j_m, given by $j_m^5/j_0^5 = 16\pi(1-j_m)/j_0 \approx 16\pi(1-j_0)/j_0$. The differential flux is therefore given by

$$\frac{S_J^{\rm NR+RR}(E)}{S_J^{\rm NR}(E)} = \left[\int_{j_{lc}}^{j_m} \frac{1-j^2}{d_{jj}^{\rm NR}} \frac{dj}{j} + \int_{j_m}^{1} \frac{1-j^2}{d_{jj}^{\rm NR} + d_{jj}^{\rm RR}} \frac{dj}{j} \right]^{-1} \approx \left[1 - \frac{\chi_{\rm RR}}{1+\chi_{\rm RR}} \frac{\log(16\pi(1-j_0)j_0^4)}{5\log(j_{lc})} \right]^{-1}, \tag{43}$$

where $\chi_{\rm RR} = d_{\rm RR}^0 e^{-4\pi j_0^4}/d_{\rm NR}^0$. As shown in Figure 10, this analytic approximation reproduces the MC results.

The small effect of RR on the loss rates can be estimated by integrating Equation (43) over the relevant region:

$$R_p^{\rm NR+RR}(E) = -\int_E^{E_{\max}} S_J^{\rm NR+RR}(E')dE'. \tag{44}$$

4. MONTE CARLO MODELS

We complement and validate our analytic study of the relativistic loss cone by numerically evolving the FP equation in both E and J using a MC procedure (described in Appendix B). Unlike the analytic treatment, this procedure does not assume that the evolution in J can be decoupled from that in E. The advantages of the MC method over direct N-body simulations are the high degree of flexibility it offers for isolating and studying the different mechanisms that affect the dynamics of the loss cone, the ease of including additional physical effects and of modifying the initial and boundary conditions, and importantly, its scalability to systems with a realistically large number of stars. In Section 5 below we employ the MC procedure to calculate the phase-space density and rates of relaxed galactic nuclei, and in particular, that of a Milky Way-like nucleus with a $M_\bullet = 4 \times 10^6 \, M_\odot$ MBH and $N_\star \sim \mathcal{O}(10^7)$ stars on its radius of influence, $r_h \sim 2$ pc. Such nuclei are considered archetypal for future space-borne missions to detect low-frequency GWs from inspiraling compact objects.

We begin here by validating the MC procedure. First, we study the dynamics in the restricted case where E remains fixed, which allows a direct test of the impact of adiabatic invariance on the long-term dynamics (Section 4.1). We then compare rate results from our MC procedure in both E and J with the currently available results from direct post-Newtonian N-body simulations of small-N systems (Section 4.2).

4.1. j-only Monte Carlo Simulations

The maximal entropy limit (Appendix E) provides a basic test for the physical validity of the DCs and of the MC procedure used to evolve the Fokker–Planck equation. The probability density of a closed system with zero net angular momentum must asymptote to the maximal entropy solution $dN/dj = 2j$. Experimentation shows that this is a sensitive test of both the functional form of the DCs and the details of the MC procedure, in particular, the implementation of the boundary conditions. We verify the maximal entropy limit in Section 4.1.1. In the absence of NR (for example on timescales $\ll T_{\rm NR}$), a relativistic system that is subject to RR with a smooth background noise should display adiabatic invariance (AI) in the form of a sharp drop in the phase-space density below some small value of j where the GR precession period falls below the coherence time (Bar-Or & Alexander 2014). RR with non-smooth background noise is not expected to display such an AI barrier. We demonstrate that our MC procedure reproduces this behavior in Section 4.1.2. Finally, we study the realistic case where NR smears the RR-generated AI in Section 4.1.3, and also show how this smearing appears in the unrestricted case where both a and j evolve.

Since the $j \to 0$ limit is of special interest, it is efficient to use logarithmic bins to collect statistics on the phase-space density. In that case, it is more useful to represent the density as $dN/d\log j$,[8] for which the maximal entropy solution is $dN/d\log j = 2j^2$ (or $dN/d\log_{10} j = 2\log(10) j^2$).

4.1.1. NR Only

Figure 11 shows the j PDF at $a = a_{\max}/4$ for an $\alpha = 7/4$ cusp with $a_{\max} = 10^4 a_{\min}$ after time $t = 100 T_E$ ($T_E = [E^2/\Delta(E^2)]_{a_{\max}}$). Near-complete convergence is already reached at $T \gtrsim T_E$, (Section 4.1.3). The convergence to the expected maximal entropy solution is apparent, although a bias toward a somewhat steeper slope for $j \lesssim 0.1$ is observed. In

[8] The decreasing size of the bins in linear space leads to a misleading graphical representation of dN/dj when the statistics are low, as is the typical case at low j, since the normalized bin density $\Delta N/\Delta j$ diverges for $\Delta N \geqslant 1$ as $\Delta j \to 0$.

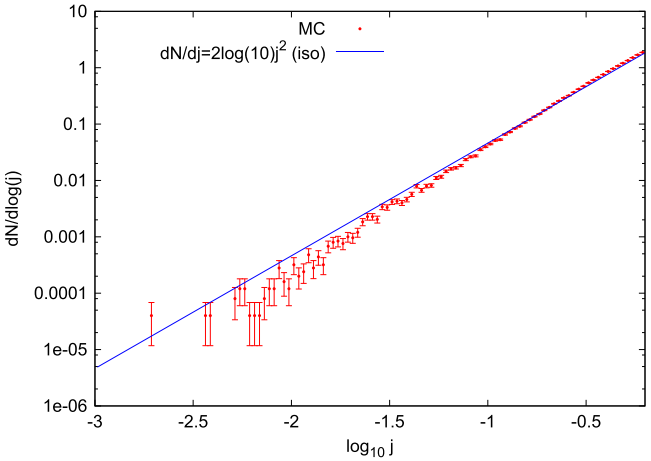

Figure 11. Convergence of an MC j-only NR diffusion simulation to the maximal entropy solution $dN/d\log_{10} j = 2\ln 10 j^2$.

addition to the $\alpha = 7/4$ case shown in Figure 11, simulations with other values of α or of $a_{\max}/a_{\min}$ confirm that the maximal entropy solution generally holds.

4.1.2. RR Only

Figure 12 shows the j PDF for the RR-only case. The MC code reproduces the AI barrier for smooth noise (Figure 12 top) at $j_0 = \sqrt{T_c \nu_{\rm GR}(a, j=1)/2\pi}$, where the PDF rapidly drops. j_0 is effectively independent of time (Bar-Or & Alexander 2014) and constrains the minimal angular momentum accessible to the system. In contrast, when the noise is not smooth (Figure 12 bottom), there is no such barrier. While at any finite time there exists a scale $j_b(t)$ where the PDF falls below the maximal entropy solution, thereby mimicking an instantaneous barrier, $j_b(t)$ evolves rapidly on the RR diffusion timescale to $j_b \to 0$ and the PDF asymptotes to the maximal entropy configuration for all physically relevant values of j within a few diffusion timescales (see Bar-Or & Alexander 2014, Figure 2). This is a demonstration that the asymptotic limit of maximum entropy does not depend on the nature of the relaxation process.

4.1.3. NR+RR

The presence of NR erases the AI cutoff in the j PDF on timescales approaching or exceeding the NR timescale (quantified by the energy diffusion timescale T_E). This is demonstrated in Figure 13, which shows a sequence of j-only MC simulations that include both NR and RR. All of the simulation runs had a fixed duration $T_{\rm sim} = 100T_E$, which kept the number of binned points, and hence the statistical sampling fluctuations, fixed (the MC values are sampled every $\Delta t = 1\ T_E$). However, the effective NR timescale was artificially extended to $T'_E = X_D T_E$ $(X_D \geqslant 1)$, so that $T_{\rm sim} = (100/X_D)T'_E$, and X_D was varied from $X_D = 10^3$ ($T_{\rm sim} = 0.1T'_E$) down to $X_D = 1$ ($T_{\rm sim} = 100T'_E$): the larger X_D, the less significant NR is over $T_{\rm sim}$. Figure 13 shows results for $T_{\rm sim} = 0.1T'_E$, $T_{\rm sim} = T'_E$, and $T_{\rm sim} = 100T'_E$. The AI cutoff is substantially smeared already when $T_{\rm sim} = 0.1T'_E$ (compare Figure 13 top left with Figure 12 right) and the AI remains only as a moderate steepening of the slope below j_0 for $T_{\rm sim} = T'_E$. For $T_{\rm sim} = 100T_E$, the j PDF is almost indistinguishable from the case of NR only.

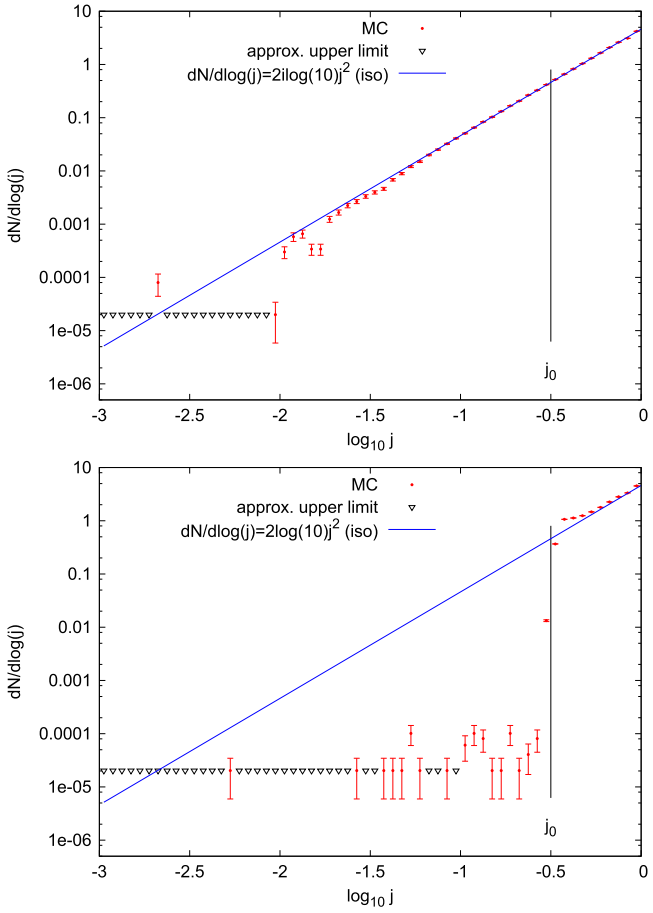

Figure 12. Adiabatic invariance and the convergence of j–RR to the maximal entropy solution $dN/d\log_{10} j = 2\ln 10 j^2$ for two different background noise models: non-smooth noise (C^0) with an exponential ACF (top) and smooth noise (C^∞) with a Gaussian ACF (bottom). The simulations were evolved to $T_{\rm sim} = 100T_E$. The predicted characteristic position of the AI front at j_0 is marked by the vertical line. An approximate upper limit on the density (1 count per bin) was estimated for empty bins (triangles).

This trend is also evident in the general case where both a and j are free to evolve, as shown in Figure 14. On timescales of the order of the RR relaxation time, but shorter than the NR timescale, the stellar trajectories are bound by the AI line, as was qualitatively observed in the N-body simulations of MAMW11. However, on longer timescales, NR drives stellar diffusion across the AI line and beyond. The existence of a *persistent* SB with a locus of $a_{\rm SB}(j) \propto j^{-2/(5-\alpha)}$ (which is not the locus of the AI line), as suggested by MAMW11 (Equation (35) there), is neither supported by our analysis nor observed in our MC simulations. The MAMW11 MC simulations that attempted to model the effects found in the N-body simulations assumed a simplified Hamiltonian model based on a non-smooth noise. As discussed here, such simulations could not have produced a persistent barrier. We interpret the truncation of the PDF observed by MAMW11 toward low j as being due to the transient suppression of the PDF below $j_b(T_{\rm sim})$ (see Bar-Or & Alexander 2014, Figure 2), which appears as a sharp cutoff when the measured PDF is based on finite number statistics.

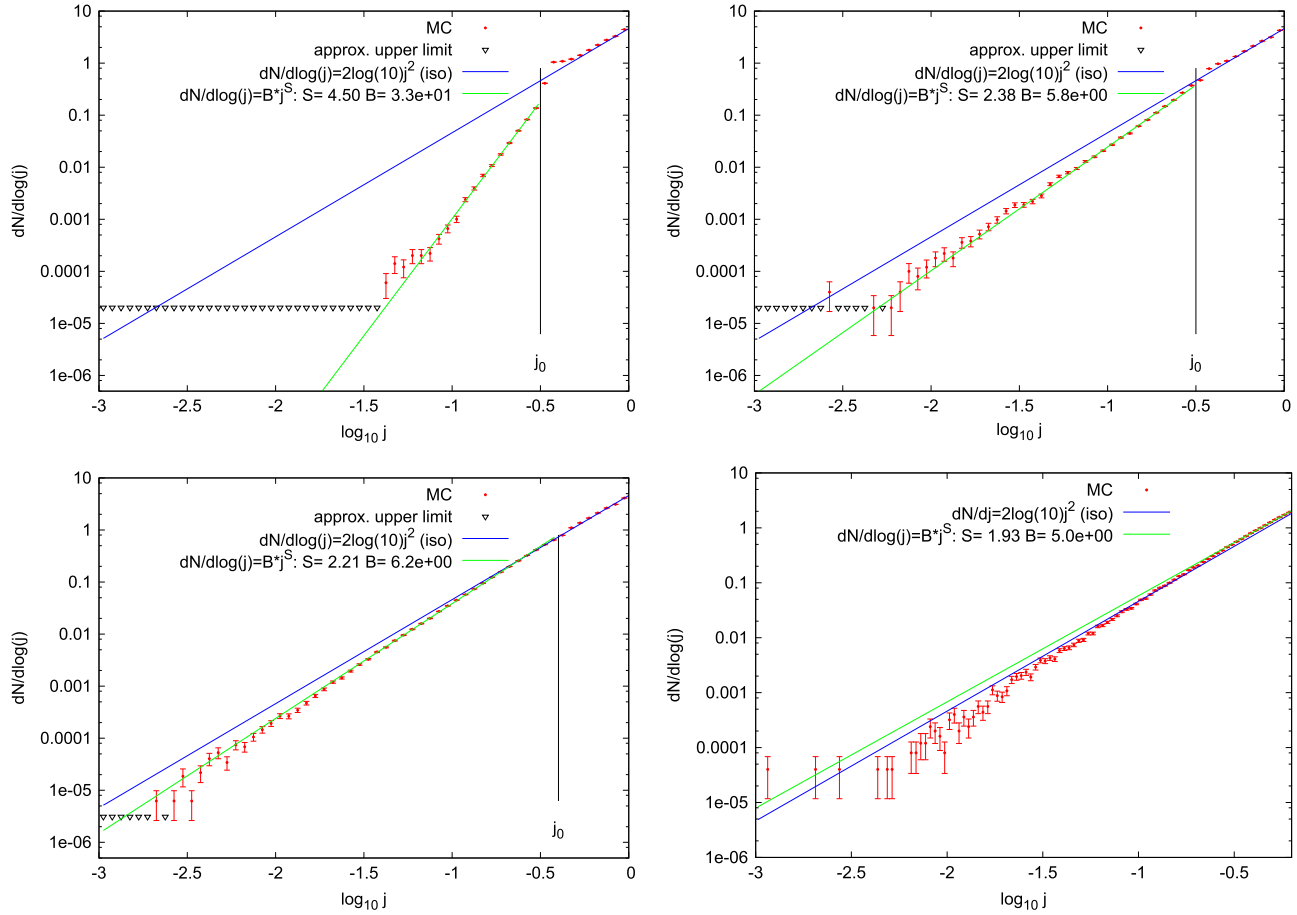

Figure 13. Suppression of adiabatic invariance by NR and the convergence of j-only MC simulations with NR and RR with a Gaussian ACF noise to the maximal entropy solution $dN/d\log_{10}j = 2\ln 10 j^2$. All of the MC runs lasted a fixed time $T_{\rm sim} = 100T_E$, where T_E is the energy diffusion timescale at $a_{\rm max}$, but the NR relaxation time was artificially extended to $T'_E = X_D T_E$ ($X_D \geqslant 1$) so that $T_{\rm sim} = (100/X_D)T'_E$ (the larger X_D, the less significant is NR). The predicted characteristic location of the AI barrier at j_0 is marked by the vertical line. The best-fit power law to the j-PDF slope is also shown (green line). An approximate upper limit on the density (1 count per bin) was estimated for empty bins (triangles). Top left: $T_{\rm sim} = 0.1T'_E$. Top right: $T_{\rm sim} = T'_E$. Bottom left: $T_{\rm sim} = 100T'_E = 100T_E$. Bottom right: NR-only MC simulation for comparison.

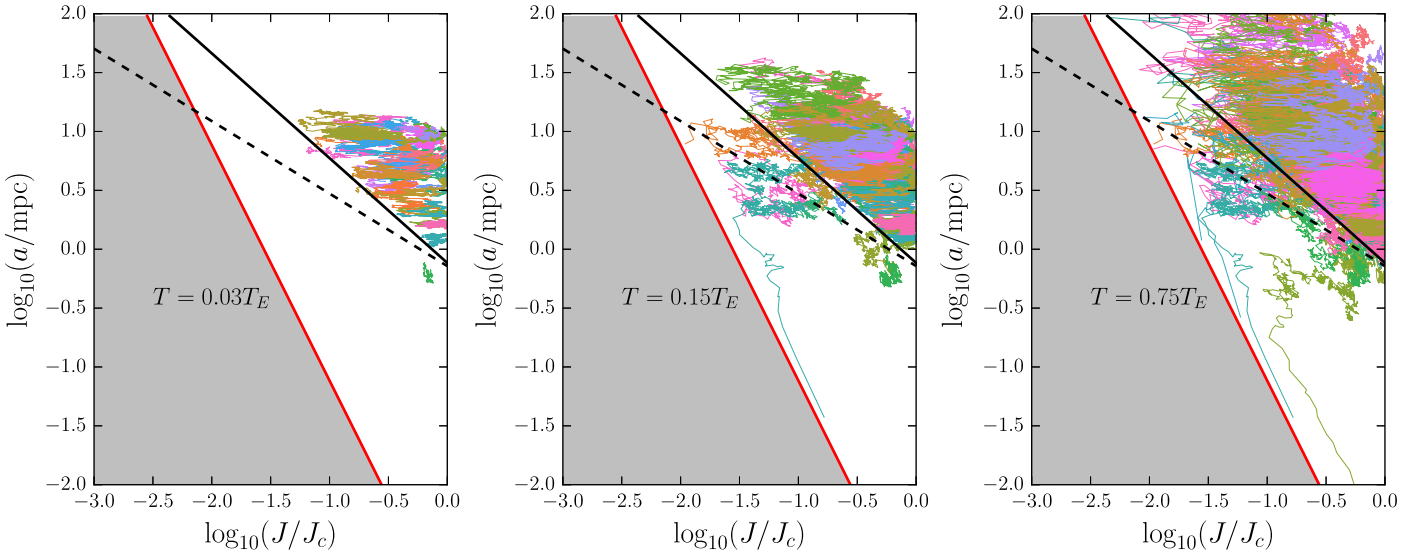

Figure 14. MC snapshots of the trajectories of individual stars at different times, corresponding to increasing fractions of the energy relaxation timescales T_E at $a_{\rm max}$. On short timescales, $t \ll T_E$, stars do not cross the AI line (solid line), while on longer timescales, $t \to T_E$, NR progressively drives stellar diffusion across the AI line to the entire available phase-space. The MC simulation assumes an MBH of $10^6\,M_\odot$ and a cusp of 50 stars of $50M_\odot$, each with initial conditions drawn from an isotropic cusp with $\alpha = 7/4$ and $a_{\rm max} = 10$ mpc. We also plot for comparison the locus of the Schwarzschild barrier for this cusp model (dashed line), as suggested by MAMW11.

4.2. Comparison with N-body Simulations

Matching the results of MC simulations to the results from direct N-body simulations is not straightforward. The MC procedure enforces boundary conditions at $a_{\max}$ and assumes an approximate steady-state background, whereas the N-body simulations of MAMW11 and BAS14 provide $a_{\max}$ only as initial conditions (the cluster subsequently expands) and allow the stellar number to fall with time as stars are lost into the MBH. In addition, the MAMW11 and BAS14 models all have an initial $n_\star(r) \propto r^{-2}$ cusp, which is away from the BW76 steady-state configuration of a single-mass population, $n_\star(r) \propto r^{-7/4}$. Thus, these N-body simulations always remain out of steady state due to relaxation, expansion, and stellar loss. The loss rates of the MC (when its parameters are matched to initial state of the N-body simulations) should therefore be compared to the *initial* loss rates of the N-body simulations. To further reduce this incompatibility, we modified our MC procedure to reproduce the initial conditions of the N-body simulations by introducing the test stars into the interior of the cusp according to an $n_\star \propto r^{-2}$ probability density.

Figure 15 shows the phase-space density derived from MC simulations for the MAMW11 cusp model (see details below), which is used here for comparison with N-body results. Results are shown for the case of Newtonian dynamics without RR, and for the case where all of the dynamical effects are switched on (Newtonian dynamics, GR, NR, and RR). Also shown are the endpoints in phase-space of a representative fraction of the plunge and inspiral events. Figure 15 and Table 2 demonstrate that GR precession plays a critical role in making EMRIs possible: in its absence, RR remains unquenched at all low values of $j \geqslant j_{lc}$, and therefore stars are rapidly driven to plunging orbits before they can reach the EMRI line and inspiral by the emission of GW. However, when GR precession is included, the region where RR dominates over NR is restricted to regions that are far from the loss lines (black contours, Figure 15 top right). This creates a bottleneck in the flow from $a_{\max}$ and $j \sim \mathcal{O}(1)$ to the loss lines, where the orbital evolution is driven by slow NR, and therefore the effect of RR on the loss rates in steady state is not large. This near-independence of the steady-state loss rates from RR was analyzed in Section 3. The MC results also show that mass precession cannot play a similar role, since it becomes efficient only for $j \to 1$ and $a \to r_h$ orbits.

Computational costs limited the MAMW11 simulations to a non-realistic cusp of only $N_\star = 50$ heavy objects of mass $M_\star = 50\,M_\odot$, each around an MBH of mass $M_\bullet = 10^6\,M_\odot$, extending from $a_{\min} = 10^{-3}$ mpc to $a_{\max} = 10$ mpc.[9] The stars were initially set on stable orbits, isotropic in orientation and eccentricity. GR was introduced to the equations of motion perturbatively, up to post-Newtonian (PN) order PN2.5. Orders PN1 and PN2 contribute only to the in-plane (Schwarzschild) periapsis precession, while order PN2.5 contributes only to dissipative GW emission. By selectively switching on or off the various PN terms, the N-body simulations tested the cases of Newtonian gravity (all PN terms switched off), no GR precession (only the PN2.5 term switched on), or full perturbative GR (all PN terms switched on). Table 2 compares the loss rates for the corresponding MC and N-body

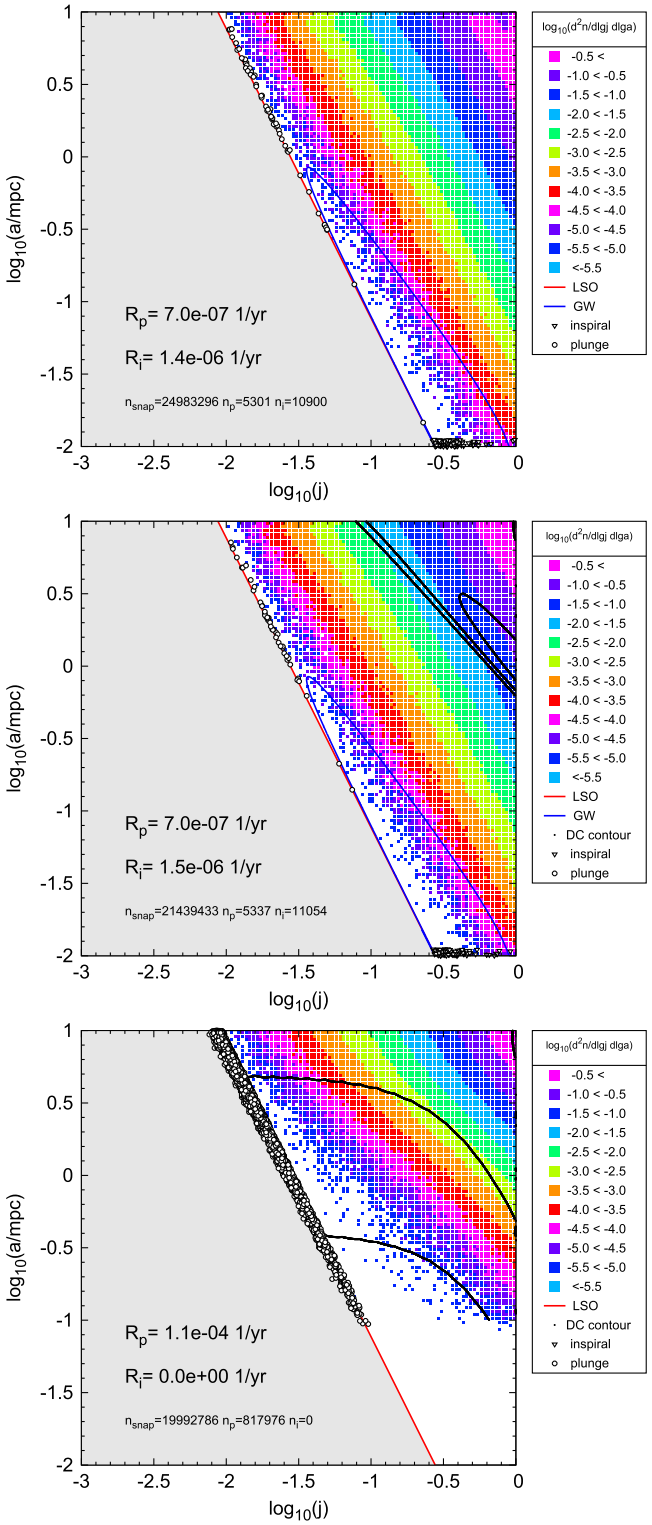

Figure 15. The 2D loss-cone phase-space density in a MAMW11-like cusp model with mass-precession coherence time and Gaussian noise calculated by MC simulations. Top: only NR. Middle: full model with all of the dynamical processes. Bottom: same as the full model, with GR precession switched off. The black contours denote the loci where RR dominates NR, at $D_2^{\rm RR}/D_2^{\rm NR} = 1$, 10, 100. To avoid clutter, only 1% of the plunge and inspiral terminal track points (circles and triangles) are displayed.

[9] These constraints lead to atypical dynamical properties in this model. (1) Because $N_\star$ is so small that $\sqrt{N_\star} \sim \mathcal{O}(N_\star)$, the difference between the mass-precession coherence time and the self-quenching coherence time is small, and the two are hard to discriminate between. (2) The NR relaxation time is $T_{\rm NR} \sim Q^2 P/N_\star \log Q$, while the mass-precession RR timescale is $T_{\rm RR} \sim QP$. Therefore, $T_{\rm RR}/T_{\rm NR} \ll 1$ everywhere in the cusp, that is, RR is atypically efficient.

Table 2
The Plunge and Inspiral Rates in MAMW11-like Cusp Models

Method[a]	Processes[b]	T_c[c]	Noise[d]	Plunge[e]	Inspiral[e]
MC	All	SQ	E	0.96	1.8
MC	All	SQ	G	0.71	1.5
MC	All	M	E	0.92	1.8
MC	All	M	G	0.71	1.4
MC	No RR	⋯	⋯	0.68	1.5
MC	No GR	M	G	110	⋯
MC	No GR prec.	M	G	110	0
MC	No mass prec.	M	G	0.71	1.4
NB1	With GR ($t \to 0$)			∼0.2	∼0.9
NB1	No GR ($t \to 0$)			>20	⋯
NB1	No GR prec. ($t \to 0$)			>20	<1
NB2	With GR			0.5 ± 0.1	0.8 ± 0.1
NB2	No GR			28 ± 2	⋯
NB2	No Gr prec.			26 ± 2	4.3 ± 0.6

Notes.
[a] Method: MC—Monte Carlo, NB—*N*-body (1: MAMW11, 2: BAS14).
[b] Processes: all includes NR, RR, GW (Gair et al. 2006), mass prec., GR prec.
[c] Coherence time: M—Mass prec., SQ—Self-quenching.
[d] Noise model: G—Gaussian noise, E—Exponential noise.
[e] Event rates in units of 10^{-6} yr^{-1}.

Table 3
The Plunge and Inspiral Rates in Milky Way-like Cusp Models

$M_\star$[a]	Processes[b]	T_c[c]	Noise[d]	Plunge[e]	Inspiral[e]
1	No RR	⋯	⋯	730	3.1
1	GW1	SQ	W	16000	0.0
1	GW1	SQ	E	860	3.3
1	GW1	SQ	G	880	2.3
1	GW1	M	W	930	0.0
1	GW1	M	E	840	3.2
1	GW1	M	G	840	3.2
10	No RR	⋯	⋯	610	2.8
10	GW1	SQ	W	6060	0.0
10	GW1	SQ	E	760	1.9
10	GW1	SQ	G	690	2.4
10	GW1	M	W	800	0.0
10	GW1	M	E	730	2.0
10	GW1	M	G	730	2.5
10	GW2	M	G	730	1.2
10	GW3	M	G	740	1.1

Notes.
[a] Stellar mass in $M_\odot$.
[b] GW approximations: GW1 Gair et al. (2006), GW2 Peters (1964), GW3 Hopman & Alexander (2006a).
[c] Coherence time: M—Mass prec., SQ—Self-quenching.
[d] Noise model: W—White, E—Exponential, G—Gaussian.
[e] Event rates in units of 10^{-6} yr^{-1}.

simulations, as well as for an artificial model that can only be realized by the MC method, a Newtonian case where RR is switched off. The MAMW11 loss rates were reported as a function of time, and so it is possible to extrapolate to $t \to 0$ and obtain lower or upper limits on the rates. The BAS14 rates were reported only in the average.

Table 2 shows that the MC loss rates for the full GR models are quite similar, irrespective of the RR noise model, and that they are also similar to the rates predicted in the artificial case where RR is switched off. The MC loss rates are somewhat higher than those derived from the *N*-body simulations. The MC model with smooth (Gaussian) noise provides a better fit to the *N*-body results, while the coherence models are virtually indistinguishable, with only a marginal preference for mass precession.

The MC loss rates are higher by factors of 1.5–3.5 than the *N*-body loss rates. As we now argue, this is generally consistent in sign and magnitude with the expected rate differences due to the differences between the computational techniques and the physical assumptions, as well as with the small numbers of statistical fluctuations in the MAMW11 rates (by more than a factor of 2, see their Figure 2).

Compared to the MAMW11 *N*-body simulations, the MC plunge rate is 3.4 higher and the inspiral rate is 1.7 higher; compared to the large-*N* BAS14 simulations, the MC plunge rate is 1.5 ± 0.3 higher and the inspiral rate is $1.8^{+0.2}_{-0.2}$ higher. Since the incompatibility of the MC and *N*-body treatment of the boundary condition at $a_{\max}$ results in a lower stellar density in the *N*-body simulations, the fact that the *N*-body loss rates are systematically lower is to be expected.[10] We note that the MC boundary conditions are a better representation of the infinite stellar reservoir of the host galaxy than the vacuum boundary conditions of the *N*-body simulations. The same systematic trends in the rates are also seen in models where GR is switched off (i.e., no GR precession, no GW).

An additional difference between these studies is the plunge criterion. MAMW11 used $r < 8r_g$, BAS14 used $r < 6r_g$, and here the criterion was based directly on the angular momentum, $J < J_{lc}$ (where J was evaluated in the Keplerian limit). As noted by Gair et al. (2006), plunging orbits (i.e., parabolic orbits with $J = J_{lc} = 4GM/c$) correspond to Keplerian orbits with periapsis $r_{lc} = 8r_g$, or to relativistic orbits with periapsis $r_{lc} = 4r_g$. This can explain some of the systematic differences in the rates, since when r_{lc} *over-estimates* the true value, stars that should inspiral plunge prematurely, thereby biasing the rates to too-high plunge rates and too-low inspiral rates. Conversely, when r_{lc} *under-estimates* the true value, a too-high inspiral rate will follow. Our approximate angular momentum plunge criterion for parabolic orbit applies generally in both the Keplerian and relativistic regimes (Gair et al. 2006), unlike the periapsis criteria used by the *N*-body simulations. This may explain why the MC inspiral rates are somewhat higher than the *N*-body results, which assume a too-high value of r_{lc} for the relativistic regime.

When only GW is switched on but GR precession is switched off (i.e., no GR quenching, Figure 15), the MC inspiral rate is zero, in agreement with MAMW11. However, BAS14 find a discrepant non-zero inspiral rate for this case. We believe that this is an erroneous result that is also due to their biased plunge criterion, which is underestimated in the Keplerian regime ($r_{lc} = 6r_g$ instead of the correct $r_{lc} = 8r_g$), and therefore allows inpiral events to occur where none should.

[10] This effect can be reproduced in the MC simulations. A suite of MAMW11-like MC models with $a_{\max}$ increasing from 10 mpc (the initial $a_{\max}$ in the MAMW11 models) to 50 mpc (the extent of the expansion in those models, see MAMW11 Figure 5) shows a factor 1.7 decrease in the total loss-rate due the longer survival times in the larger available phase-space volume (see Equation (69)).

5. LOSS RATES

5.1. The Galactic Center Test Case

The MBH in the center of the Milky Way and the stars and compact objects around it is a system of particular relevance, both because it is uniquely accessible to observations, and can therefore place constraints on dynamical models and theories (Alexander 2005), but also because planned space-borne GW detectors with $\mathcal{O}(10^6$ km) baseline will be optimally sensitive to GWs emitted by a mass orbiting a 10^6–$10^7 M_\odot$ MBH near the LSO (Amaro-Seoane et al. 2007). Therefore, although it remains an open question whether the Galactic center (GC) is surrounded by a high-density relaxed cusp of stellar remnants (see review by Alexander 2011), and although the rate of GW events from the GC itself is expected to be small (but see Freitag 2003), the Galactic MBH SgrA* represents the archetypal cosmic GW target.

We adopt here a simplified model of the GC consisting of an MBH mass of $M_\bullet = 4 \times 10^6\, M_\odot$, surrounded by a steady-state $\alpha = 7/4$ cusp of equal-mass stars of either $M_\star = 1\, M_\odot$ or $10\, M_\odot$, extending from $a_{\rm min} = 2 \times 10^{-4}$ pc ($a_{\rm in} = 0.1 a_{\rm min}$, see Appendix B) to $a_{\rm out} = a_{\rm max} = r_h = 2$ pc with a total stellar mass of $M_\star N_\star(<r_h) = 2M_\bullet$ inside the radius of influence. Test stars are injected into the nucleus with initial sma $a_0 = a_{\rm out}$ and with isotropic j_0.

Figure 16 shows the steady-state configuration and loss rates for a GC model with $M_\star = 10\, M_\odot$ and a smooth background noise whose coherence is set by mass precession. As expected from the fact that the RR-dominated region in phase-space is well separated from the loss lines, the steady-state phase-space density is very close to the simple NR-only solution.

Table 3 explores the implications of varying some of the assumed processes for the loss rates: the mass of the cusp stars ($1\, M_\odot$ or $10\, M_\odot$), the nature of relaxation (NR only, or NR and RR) the noise model (white, exponential, Gaussian), the coherence mechanism (mass precession, self-quenching), or the GW dissipation approximation.

The uncorrelated (white) noise model for the resonant torques, which is equivalent to the assumption that $\nu_{\rm GR} \to 0$, results in very high plunge rates as strong RR rapidly drains the cusp, and as a result the EMRI rate drops to zero. In contrast, for all other RR models, irrespective of assumptions about the nature of the noise or the coherence mechanism, GR precession suppresses RR to the extent that rates are very similar to those derived for the non-physical "NR-only" model: a plunge rate of $\Gamma_p \sim$(6–9)$\times\ 10^{-4}\,{\rm yr}^{-1}$ and an inspiral rate of $\Gamma_i \sim$(1–3)$\times\ 10^{-6}$ years. We find that the more sophisticated GW dissipation estimate of Gair et al. (2006) results in inspiral rates that are a factor $\sim$2 higher than the estimates by Peters (1964) or Hopman & Alexander (2006a). We conclude that to within a factor of $\sim$2, our rate estimates for relaxed steady-state cusps are robust to variations of the physical assumptions.

5.2. Scaling with the MBH Mass

The MC simulations can be used to validate a simple analytic model for estimating the loss rates and their dependence on the parameters of the galactic nucleus, which is based on identifying critical values of the sma, a_c, below which the probability of a star to cross the loss line is $\mathcal{O}(1)$ (Lightman & Shapiro 1977; Hopman & Alexander 2005). The loss rate is then $\Gamma \propto N(<a_c)/T_{\rm NR}(a_c)$, where the proportionality factor includes the suppression of the density near the loss line. For plunge events $a_c \sim \mathcal{O}(r_h)$ (Lightman & Shapiro 1977), while for GW inspiral $a_c \sim a_{\rm GW} \ll r_h$, that is, the maximum of the GW line (Section 3). Figure 1 shows that the region of phase-space where RR dominates the dynamics is well separated from the loss lines, is well below r_h, and well above $a_{\rm GW}$. The timescale relevant for estimating is therefore that of NR and not RR.

In order to estimate the integrated cosmic rates of EMRIs or tidal disruption flares, it is necessary to scale the loss rates by the parameters of the host galaxy, in particular, the MBH mass. Here, we adopt a simplified one-parameter sequence of galactic nuclei where the free parameter is $M_\bullet$, which together with several additional fixed parameters define the sequence. The $M_\bullet$ scaling is based on the empirical $M_\bullet/\sigma$ relation $M_\bullet = M_0(\sigma/\sigma_0)^\beta$, where σ is the stellar velocity dispersion outside the MBH radius of influence $r_h = \eta_h GM_\bullet/\sigma^2$, which

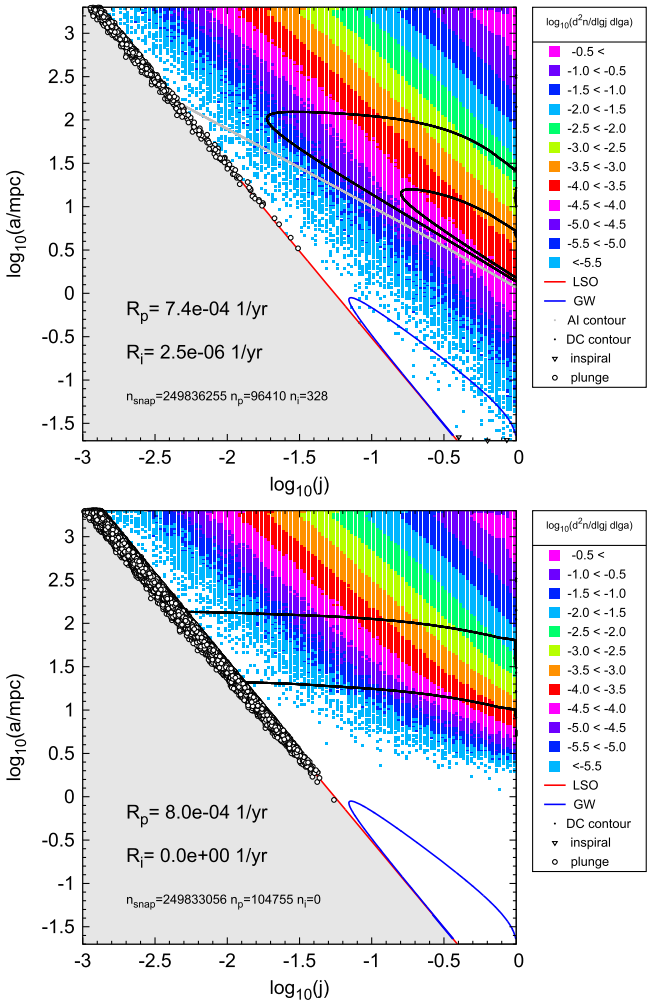

Figure 16. The 2D loss-cone phase-space density in a Milky Way-like cusp model with mass-precession coherence time and Gaussian noise, calculated by high phase-space resolution MC simulations. Stars/stellar mass BHs of $10\, M_\odot$ are assumed. Top: GR precession included. Mass precession limits the efficiency of RR beyond $\sim$100 mpc, while GR precession limits RR below the AI locus (gray line). RR is faster than NR only well away from the loss cone, inside the black contours (equally fast at the outer contour, 10 times faster at the inner contour). Bottom: when GR precession is artificially switched off, RR remains effective all the way down to the loss cone and is faster than NR below $a \sim 100$ mpc. As a result, stars are driven to plunge trajectories well before they can lose enough energy by NR to reach the GW loss-line. A central, strongly depleted cavity is formed and the EMRI rate is completely suppressed.

encloses a stellar mass of the order of the mass of the MBH $N_\star(r_h) = \mu_h Q$. The power-law parameter $\beta \sim 4$–5 and the normalization M_0/σ_0^β are determined empirically (Ferrarese & Merritt 2000; Gebhardt et al. 2000). It then follows that $r_h = \eta_h (M_\bullet/M_0)^{1-2/\beta} G M_0/\sigma_0^2$ (Alexander 2011).

Using this parameterization and the approximation that the steady-state distribution is given by a BW76 cusp, the total plunge and inspiral rates can be estimated from Equations (29) to (32),

$$R_p^{\rm tot} \approx \frac{10\mu_h^2}{\eta_h^{3/2}} \frac{\sigma_0^3}{2\pi G M_0} Q_0^{3/\beta-1} \times \frac{\log Q}{\log\left[\sqrt{\eta_h} Q_0^{-1/\beta} c/4\sigma_0\right]}, \quad (45)$$

and

$$R_i^{\rm tot} \approx A_{\rm GW} \frac{10\mu_h^{6/5}}{\eta_h^{3/2}} \frac{\sigma_0^3}{2\pi G M_0} Q_0^{3/\beta-1} \times \frac{(\log Q)^{1/5}}{\log\left[\sqrt{\eta_h A_{\rm GW}} Q_0^{-1/\beta} c/4\sigma_0 (\mu_h \log Q)^{-2/5}\right]}, \quad (46)$$

where $Q_0 = M_\bullet/M_0$, $a_{\rm GW} = A_{\rm GW}(\mu_h \log Q)^{-4/5} r_h$, and $A_{\rm GW}$ is a numerical factor that depends on the GW dissipation approximation (Appendix A).

In our MC simulations, for simplicity, we adopted $\beta = 4$, $\mu_h = 2$, $\eta_h = 1$, $M_0 = 5.4 \times 10^6\, M_\odot$, and $\sigma_0 = 100$ km s^{-1}. Thus,

$$r_h = 2 (M_{\bullet,\rm MW})^{1/2} \text{ pc}, \quad (47)$$

where $M_{\bullet,\rm MW} = M_\bullet/4 \times 10^6 M_\odot$ is the MBH mass scaled to the mass of the Galactic MBH. The rates as a function of the MBH mass $M_\bullet$ and the mass ratio $Q = M_\bullet/M_\star$ are then

$$R_p^{\rm tot} \approx 3 \times 10^{-4} M_{\bullet,\rm MW}^{-1/4} \times \frac{\log Q}{6.70 - 0.25 \log(M_{\bullet,\rm MW})} \text{ yr}^{-1}, \quad (48)$$

and

$$R_i^{\rm tot} \approx 5 \times 10^{-6} M_{\bullet,\rm MW}^{-1/4} \times \frac{(\log Q)^{1/5}}{4.65 - 0.25 \log(M_{\bullet,\rm MW}) - 2\log(\log Q)/5}, \quad (49)$$

where we used the value $A_{\rm GW} \approx 0.029$, corresponding to the GW dissipation approximation of Gair et al. (2006; Appendix A). As shown in Figure 17, these analytic approximations are in agreement with the results of the MC simulations over several orders of magnitude of $M_\bullet$.

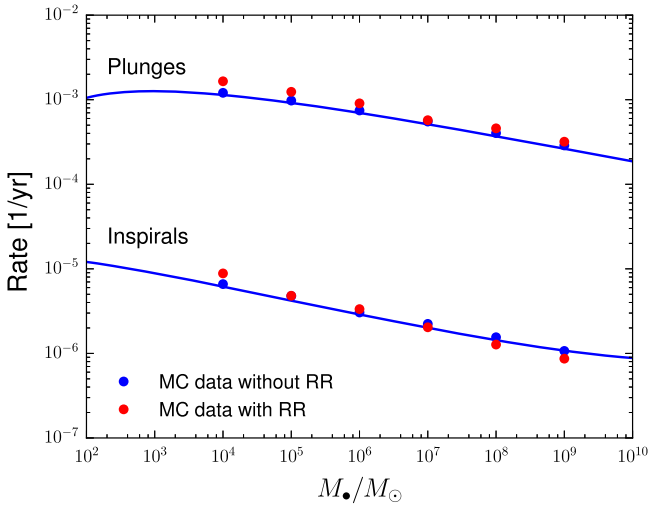

Figure 17. Total plunge and inspiral rates as a function of MBH mass. The MC simulations (circles) agree with the analytic approximations for dynamics without RR (solid lines), Equations (48), (49). Simulations with RR show that the contribution of RR is small: the discrepancy between the rates with and without RR does not exceed ~30% over five orders of magnitude in $M_\bullet$.

6. DISCUSSION AND SUMMARY

The determination of the steady state of galactic nuclei is a fundamental open question in stellar dynamics with many implications and ramifications, and has been the focus of numerous numerical and analytical studies. In particular, current estimates of loss rates vary over several orders of magnitude due to theoretical and empirical uncertainties. Previous studies either used post-Newtonian N-body simulations, which are limited to small N, or did not include the relevant relativistic physics (Section 1). Building on recent progress in the formal description of RR as a correlated diffusion process (the $\boldsymbol{\eta}$-formalism; Bar-Or & Alexander 2014), we obtain an MC procedure and analytic expressions for the steady-state distribution and loss rates in galactic nuclei, taking into account two-body relaxation, RR, mass precession, and the GR effects of in-plane precession and GW emission. By cross-validating the analytic estimates and the MC results with a high degree of accuracy, and without the introduction of any free fit parameters, we are able to confirm our analysis and interpretation of the dynamics of the loss cone in the context of our underlying assumptions.

6.1. Discussion of Main Results

The advantage of modeling RR using the $\boldsymbol{\eta}$-formalism, compared to previous attempts by other approaches (Rauch & Tremaine 1996; Hopman & Alexander 2006a; Gürkan & Hopman 2007; Madigan et al. 2011; MAMW11; Antonini & Merritt 2013; Hamers et al. 2014), is that it allows us to derive the FP equation rigorously from the stochastic leading-order relativistic three-dimensional (3D) Hamiltonian. The resulting effective DCs, which are thus derived from first principles, are then guaranteed to obey the fundamental fluctuation–dissipation relation and the correct 3D maximal entropy solution (-Binney & Tremaine 2008, Section 7.4.3; Appendix E).

These constraints on the functional form of valid DCs are critical, since the correct steady state is the result of a fragile near cancellation of two large opposing currents (the diffusion and drift); even small deviations from this relation (e.g., due to approximations, empirical fits, or reduction to lower dimensions) will result in large errors. For example, Hamers et al. (2014) obtained the RR DCs from numerical simulations using an assumed functional form, $D_{jj} \propto \sqrt{1-j^2}$, based on the fit of Gürkan & Hopman (2007)[11] and on the ad hoc expression $D_j = j^{-1} D_{jj}$, which is inconsistent with the fluctuation–dissipation relation, and therefore leads to an invalid steady-

[11] We obtain a more accurate expression for D_{jj} (Appendix D) which fits torques measured in static wires simulations very well, over the entire range $j \in [0, 1]$.

sate solution. This was then partially remedied by Merritt (2015a), who treated separately the Newtonian ($j \to 1$) and relativistic ($j \to 0$) regimes. In the Newtonian regime, the Hamers et al. (2014) data was re-fitted to DCs that effectively satisfy the fluctuation–dissipation relation, which means that in the absence of a loss cone, the dynamics asymptote to the maximal entropy limit $n(j) = 2j$. However, in the relativistic limit $j \to 1$, where the simulation statistics are poorer due to the smaller phase-space volume, Merritt (2015a) used analytic DCs based on the Hamiltonian model of MAMW11, which represented the stochastic background by an ad hoc dipole pseudo-potential and a recipe for switching its direction every coherence time. This recipe corresponded to the η-formalism's "Steps" or "Exponential ACF" noise (depending on the exact switching procedure), which both converge to the same form in the $j \to 0$ limit (Bar-Or & Alexander 2014, Figure 1). As shown by Bar-Or & Alexander (2014, Equation (42)), in that limit $D_{jj} \approx j^4/\tilde{T}_c$ and $D_j \approx (5/2)j^3/\tilde{T}_c$, where $\tilde{T}_c = 0.5 T_c \nu_{\rm GR}^2 (j=1)/\nu_j^2 (j=0)$. This indeed satisfies the fluctuation–dissipation relation, as any Hamiltonian model is guaranteed to do. These DCs are different from those derived by Merritt (2015a), $D_{jj} \approx j^4/\tilde{T}_c$ and $D_j \approx 2j^3/\tilde{T}_c$, who implicitly forced the solution to two-dimensional (2D) in-plane motion by setting $\sin i = 1$ in the derivation (Merritt 2015a, Equations (C.8)–(C.9)). Therefore, these derived DCs satisfy the 2D fluctuation–dissipation relation $2D_j = \partial D_{jj}/\partial j$, rather than the correct 3D one, $2jD_j = \partial j D_{jj}/\partial j$. These DCs therefore imply the steady-state solution $n(j) = $ const in the relativistic regime (assuming no loss cone); this is *not* the correct solution for 3D orbital motion (the correct one is $n(j) = 2j$). We note that the concatenation of two diffusion solutions, a 3D one for the Newtonian regime and a 2D one for the relativistic regime, may create an artificial discontinuity in the dynamical behavior at the interface.

We have shown that the representation of stochastic dynamics near a MBH in terms of the streamlines of the probability flow provides a powerful tool for analyzing the loss fluxes, and leads to the identification of the exact separatrix between plunges and inspirals. We show that the typical sma of this separatrix, $a_{\rm GW}$, yields an excellent analytical estimate for the inspiral rates found in our MC simulations (Figure 17). This remedies the ambiguity in the identification of the critical sma, which was used to estimate rates in previous studies. We also explored the effect of different GW dissipation approximations on $a_{\rm GW}$ and the resulting GW inspiral rate, and found that the rates are robust to within a factor 2. Nevertheless, it is worth noting that the more accurate method of Gair et al. (2006) predicts EMRI rates that are more than twice higher than those of the commonly used approximation of Peters & Mathews (1963).

We have shown that GR precession plays a critical role in the dynamics of the loss cone by efficiently quenching the RR torques. Conversely, in its absence, all of the stars would rapidly plunge into the MBH, creating a depleted central cavity (see Figure 16). This implies that GR precession is important even in systems that are effectively Newtonian, where at any given time only a very small fraction of the stars are on relativistic orbits. In particular, *N*-body simulations of stellar dynamics and stellar populations near MBHs should include GR precession, even if the questions of interest are in the non-relativistic regime, so that plunges do not compete, or limit the lifetime of the stars. It is worth noting that GR precession has not yet been tested empirically for relativistic parameters larger than $\Upsilon = 2GM_\bullet/c^2 r > 8 \times 10^{-6}$ (the double pulsar system PSR J0737-3039, Lyne et al. 2004), whereas the S-stars in the Galactic center, for example, are already observed to reach $\Upsilon = 2GM_\bullet/c^2 r > 1 \times 10^{-3}$ at periapsis (star S14; Gillessen et al. 2009a, see also Alexander 2006). Moreover, near the relativistic loss cone ($r_p = 4r_g$; Gair et al. 2006), where $\Upsilon \to 1/2$, there is to date no empirical confirmation of GR precession (Will 2006). The existence, dynamics, and loss rates of stars on such relativistic orbits can therefore probe GR precession in the strong field limit.

We have shown that the influence of RR on the steady-state loss-cone dynamics of compact objects is a small $\mathcal{O}(1)$ effect, since the loss lines for both direct plunge and GW inspiral lie well outside the region where RR is effective (e.g., Figure 1). RR does introduce a correction to the steady-state distribution and the loss rates, which is at present small in comparison to the astrophysical uncertainties, such as the stellar density and mass function, the $M_\bullet/\sigma$ relation, and deviations from spherical symmetry. Equations (45)–(49) provide useful analytic estimates for the plunge and inspiral rates per galaxy based on the NR-only approximation. The RR correction to the rates can be obtained by numerical integration of Equation (44). Using a suite of MC simulations, we have verified in the context of our assumptions that the rate estimates are robust under different assumptions about the properties of the stellar background noise (smoothness, coherence time). This is in large measure a reflection of the limited role of RR in the presence of continuous noise (Bar-Or & Alexander 2014, Figure 1), which restricts the domain where RR is effective, so that slow NR remains the bottleneck and sets the rates.

RR can substantially affect processes whose loss lines intersect the phase-space region where RR dominates ($D_2^{\rm RR} \gg D_1^{\rm RR}$). The loss lines for the tidal disruption of extended objects by the MBH, such as red giants or binaries, do lie closer to the RR line. However, it can be readily shown that neither class of objects is long-lived enough for RR-driven tidal destruction to play a dominant role in Milky Way-like galactic centers. Red giants are relatively short-lived, and the more extended and tidally susceptible they are, the shorter their lifespan. Soft stellar binaries are destroyed by three-body ionization before they are affected by RR-driven tidal separation (Alexander & Pfuhl 2014).

One class of processes where RR may have more than an $\mathcal{O}(1)$ effect is the hydrodynamical destruction (or removal) of objects through interaction with a large circumnuclear accretion disk. To demonstrate this point, we consider here as a simple, idealized example the case of a massive accretion disk of radius $R_d = 2000 r_g$ (Goodman 2003) and a population of long-lived icy planetesimals around it which are destroyed by several consecutive disk crossings (it is assumed that the number of crossings for destruction is $N_{\rm cross} P \ll T_{\rm NR}$). In that case, the critical angular momentum for destruction is $j_d = \sqrt{2GM_\bullet R_d} = 15.8 j_{lc}$, which is large enough to intersect the RR-dominated zone (Figure 18 top). As a result, the differential sma distribution of the planetesimals is depleted below $a \lesssim 100$ mpc, and strongly so below $a \lesssim 10$ mpc (Figure 18 bottom).

Although RR is typically inefficient in driving stars all the way to the loss cone, it can randomize orbits in the phase-space regions where it dominates, even when the NR timescale is longer than the system's age or the lifespan of the stars. This

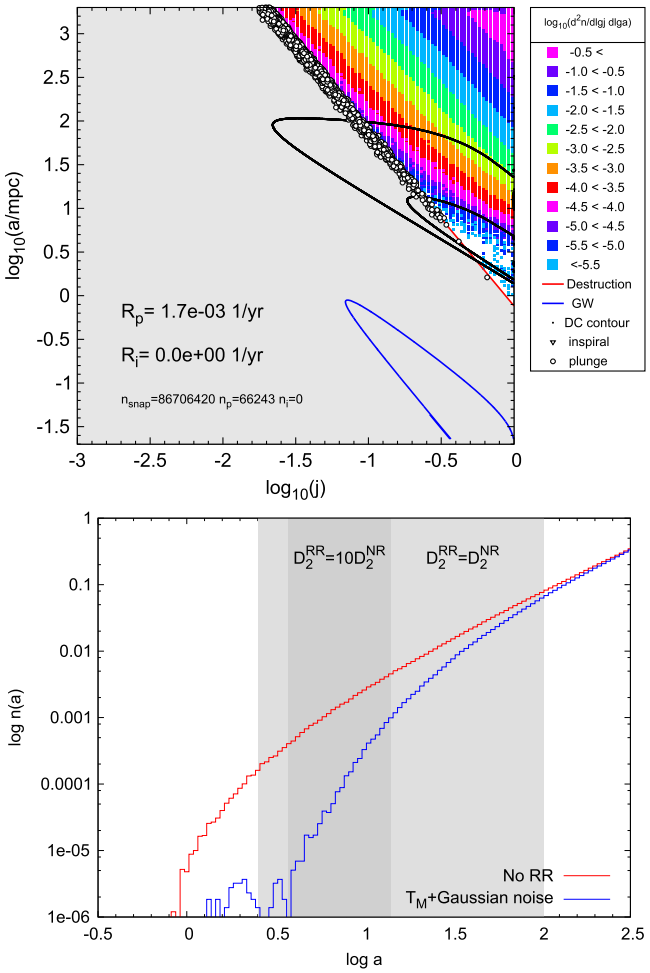

Figure 18. Suppression of the phase-space density of icy planetesimals around an MBH due to RR-driven interactions with a circumnuclear accretion disk. Top: the phase-space density and the disk interaction loss-line. The black contours delineate the region where RR dominates ($D_2^{\rm RR}/D_2^{\rm NR} = 1, 10$). Bottom: the resulting steady-state a distribution of the planetesimals for the NR-only case, and for the full dynamical model (top panel). The shaded regions denote the a intervals where the loss-line crosses regions where RR dominates.

may be a key element in solving the "paradox of youth" (Ghez et al. 2003) in the Galactic Center, where young B stars are observed on tight orbits around SgrA* (the so-called "S-star" cluster). The leading formation or migration scenarios for the S-stars predict non-isotropic initial eccentricities (see reviews by Alexander 2005, 2011). However, most of the S-stars are in the RR-dominated phase-space region (Figure 1), and so substantial evolution and isotropization of the initial eccentricities is possible. However, many of the S-stars are short-lived, and some are close in phase-space to the AI-suppressed region. A detailed analysis using the $\boldsymbol{\eta}$ formalism, which can treat this intermediate dynamical regime rigorously, has yet to be carried out.

6.2. Limitations and Caveats

The applicability and validity of our results are limited by several simplifying assumptions. We assume a non-spinning MBH surrounded by an isotropic (on average), non-rotating single Keplerian power-law cusp of single-mass stars. We assume that the dynamics are dominated by single-star interactions, that is, we neglect the possible effects of binaries or the contribution of non-stellar massive perturbers, such as gas clouds, clusters, or intermediate-mass BHs (Perets et al. 2007).

In our MC simulations, we used NR and RR DCs, which are based on a fixed isotropic, single power-law BW76 model. These DCs are therefore not self-consistent with the steady-state solution. However, we showed that the BW76 cusp is a good approximation (within a factor of two) for the steady-state solution in the relevant region (down to $a_{\rm BW} < a_{\rm GW}$ for $M_\bullet \lesssim 10^9 M_\odot$; see Section 3.4). In addition, it can be shown that for the RR DCs, the isotropic fluctuation–dissipation relation holds even for the non-isotropic case as long as the total angular momentum of the system is zero (Bar-Or & Alexander 2014). Since the fluctuation–dissipation relation results from the symmetries of the Hamiltonian, it is reasonable to assume that the same will also hold for the NR DCs. This means that small non-isotropies will only result in small magnitude changes of the flux and will have little effect on the steady-state distribution.

Finally, the RR DCs are based on simplified (single timescale) background noise models that can be treated analytically, and simple coherence models that are functions of the sma only. However, we were able to show using MC simulations that the results are largely independent of the exact noise model as long as it is continuous (i.e., not white noise).

TA acknowledges support by the I-CORE Program of the PBC and ISF (Center No. 1829/12).

APPENDIX A
THE ANALYTIC GW LINE

The phase-space of the relativistic loss cone is divided by a separatrix into an outer region where streamlines end as plunges, and an inner region where streamlines end as inspirals (Figure 4). Since the probability current along an infinitesimal bundle of streamlines is conserved (Section 3.5), it can be evaluated at any convenient point along the flow, particularly at $J \gg J_{lc}$, where GWs are negligible and the streamlines are identical to those in the absence of GWs. Therefore, the inspiral rate is estimated by locating the no-GW plunge streamline that corresponds to the separatrix and identifying the sma of its terminal point (a_p, j_{lc}) as the critical (maximal) sma for inspiral, $a_{\rm GW}$. The inspiral rate is then simply the integrated plunge rate in the absence of GW, up to $a_{\rm GW}$, i.e., $R_i^{\rm tot} = R_p^{\rm no\ GW}(a_{\rm GW})$.

The separatrix streamline, $dE/dJ = S_E/S_J$, can be evaluated by noting that the flow in the E direction is mostly due to GW dissipation,

$$S_E = n(E, J)E/t_{\rm GW}, \tag{50}$$

where $t_{\rm GW} \equiv a/|\dot{a}_{\rm GW}|$ is the GW orbital decay timescale. In the presence of GW, the innermost plunge streamline initially approaches j_{lc} from $j \sim 1$ at nearly constant a, and then turns over to lower a and runs nearly parallel to the loss-line $j_{lc}(a) = 4\sqrt{r_g/a}$ until it terminates when $j \to j_{lc}$ at some small enough terminal sma, a_i, where the orbital motion is effectively circular (i.e., inspiral, see Figure 4). For example, for parabolic orbits, the minimal possible value for a_i is obtained where $j_{lc} = 1$, i.e., for $a_i = 16r_g$. As we demonstrate below (Equation (59)), the separatrix solution is a function of a_i only

via $a_i/a_{\max} \ll 1$, and is therefore independent of the exact choice of a_i.

The flow in the J direction is approximately constant in J with a typical timescale of T_J (Equation (8)),

$$S_J \approx -\frac{n(E)}{T_J \log(J/J_{lc})}. \quad (51)$$

Since RR is negligible near $j = j_{lc}$, T_J can be approximated as $T_J(a) \approx \theta_\gamma Q^2 P(a)/(N(a)\log Q)$ (see Equation (19)), where θ_γ is a numeric pre-factor that depends on the cusp density profile ($\theta_\gamma \approx 1/8$ for a $\gamma = 7/4$ (BW76) cusp).

It then follows that the separatrix is the solution of the differential equation

$$\frac{dE}{dJ} = \frac{S_E}{S_J} \approx -2\frac{J}{J_c^2} T_J \frac{E}{t_{\rm GW}} \log\left(\frac{J}{J_{lc}}\right), \quad (52)$$

with the boundary condition $J(E_i) = J_{lc}$ (where $E_i = GM_\bullet/2a_i$).

An exact expression for the GW dissipation can be obtained only numerically. Some useful analytical approximations are available (see Gair et al. 2006 for a comparison between the different techniques). The simplest expression was obtained by Peters & Mathews (1963), who assumed point-mass objects moving on a Keplerian orbit. In this approximation, the GW timescale is

$$t_{\rm GW} \equiv \frac{a}{|\dot{a}_{\rm GW}|} = \frac{1}{2\pi}\frac{5}{64}\frac{Q}{f(e)}\frac{r_g}{a}\left(\frac{2r_p}{r_g}\right)^{7/2} P(a), \quad (53)$$

where

$$f(e) = \left(\frac{1+e}{2}\right)^{-7/2}\left(1 + \frac{73}{24}e^2 + \frac{37}{96}e^4\right). \quad (54)$$

In the limit $J \to J_c$, the periapsis r_p is related to the angular momentum by $r_p/r_g = 8J^2/J_{lc}^2$ and the streamline equation (Equation (52)) can be written in terms of $x = a/a_{\max}$ and $s = J/J_{lc}$,

$$\frac{dx}{ds} = A_D x^{\gamma-2} s^{-6} \log(s), \quad (55)$$

where the competition between the GW dissipation and NR diffusion is expressed by the parameter

$$A_D = \frac{\pi}{20} f(1) \frac{\theta_\gamma Q}{N_h \log Q} \sim \mathcal{O}(10^{-3}). \quad (56)$$

The boundary condition at the terminal sma is

$$x_1 \equiv x(s=1) = a_i/a_{\max} \ll 1, \quad (57)$$

and the plunging branch of the solution is give by

$$x(s) = \left[x_1^{3-\gamma} + \frac{A_D(3-\gamma)(s^5 - 5\log s - 1)}{25\, s^5}\right]^{1/(3-\gamma)}. \quad (58)$$

In the phase-space region where GW is negligible, the streamline defined by Equation (52) is approximately constant in E. Therefore, $a_{\rm GW}$ can be identified by the sma at $s \gg 1$ ($j \to 1$), that is,

$$a_{\rm GW} \approx a_{\max} \lim_{s\to\infty} x = \left(x_1^{3-\gamma} + \frac{(3-\gamma)}{25}A_D\right)^{1/(3-\gamma)} a_{\max}. \quad (59)$$

For a steady-state cusp (i.e., BW76, $\gamma = 7/4$), $x_1^{3-\gamma} \ll (3-\gamma)A_D/25$, and therefore

$$a_{\rm GW} \approx A_{\rm GW}\left(\frac{N_h \log Q}{Q}\right)^{-4/5} a_{\max} \quad (60)$$

where the value of $A_{\rm GW}$ depends on the specific GW dissipation approximation that is assumed. For the Keplerian approximation, $A_{\rm GW}^K = (\pi f(1)/3200)^{4/5} \approx 0.013$.

As shown by Gair et al. (2006), the Peters & Mathews (1963) estimate can be improved by using a "semi-relativistic" approximation, that is, using the fully relativistic orbit in place of the Keplerian one in the Peters & Mathews (1963) equations. In the limit $e \to 1$, this approach, used by Hopman & Alexander (2005), amounts to replacing the Keplerian r_p/r_g in Equation (54) with the relativistic one (Gair et al. 2006):

$$r_p/r_g = 8\frac{J^2}{J_{lc}^2}\left(\frac{1+\sqrt{1-J_{lc}^2/J^2}}{2}\right). \quad (61)$$

Thus, the GW separatrix (Equation (55)) is replaced by its relativistic version:

$$\frac{dx}{ds} = A_D x^{\gamma-2}\left(\frac{1+\sqrt{1-s^{-2}}}{2}\right)^{-7/2} s^{-6}\log(s), \quad (62)$$

where $a_{\rm GW}$ can be solved numerically. As shown in Figure 19, $a_{\rm GW}$ can be approximated for $\gamma = 7/4$ by Equation (60) with $A_{\rm GW}^R = 0.022$ and for the more accurate treatment of Gair et al. (2006), $A_{\rm GW} = 0.029$, which is adopted in this study.

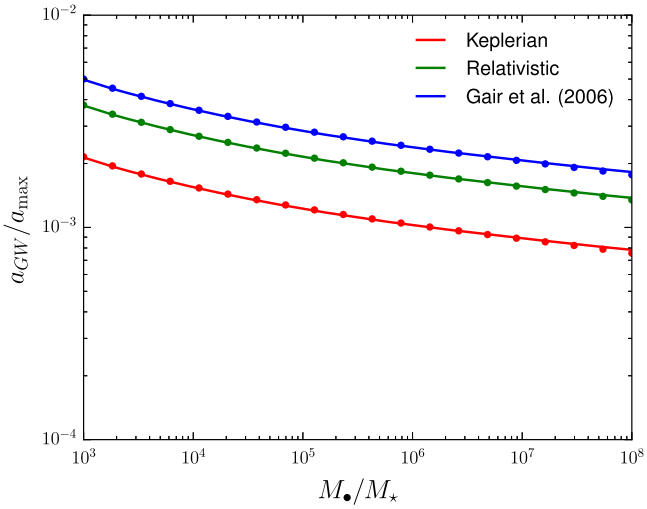

Figure 19. Semimajor axis $a_{\rm GW}$ of the outermost inspiraling streamline for a BW76 cusp ($\gamma = 7/4$) for different MBH to star mass ratios and for different approximations of the GW dissipation. The approximate expression (Equation (60)) for $a_{\rm GW}$ (lines) is fitted to values of $a_{\rm GW}$ found by numerically integrating the streamline equation (Equation (52), circles).

APPENDIX B
MONTE CARLO SIMULATIONS OF LOSS-CONE DYNAMICS

Here, we summarize the assumptions underlying our MC simulations, the details of the implementation and the derivation of the steady-state configuration and loss rates.

B.1. Assumptions and Procedure

The MC simulations in (a, j) assume a fixed background model, whose properties define the NR diffusion coefficients, an RR coherence timescale model, and a background noise model. The stellar background is approximated as a power-law cusp with an enclosed number of stars $N(a) = N_{\max}(a/a_{\max})^{3-\alpha}$, extending between $a_{\min}$ and $a_{\max}$. The accessible phase-space for the test stars extends between a reflecting boundary at $a_{\rm out} = a_{\max}$ (evaporation) and an absorbing boundary at $a_{\rm in} < a_{\min}$ (destruction). The reflective boundary at $a_{\rm out}$ ensures that the long-term distribution of test stars converges to an isotropic distribution in j, $n(j) \to 2j$, which is the assumed distribution of stars far from the MBH (note that this is not guaranteed for an absorbing outer boundary, even when the test stars are introduced into the simulation isotropically). The extension of phase-space to small $a_{\rm in}$ allows inspiral trajectories to be tracked down to a tight enough sma, where their ultimate fate as EMRIs is certain (cf Figure 15). Phase-space extends in j between a reflecting boundary at $j = 1$ and an absorbing boundary at the LSO, defined by the locus $J_{\rm LSO}(a) = J_c j_{\rm LSO} = 4r_g c$ (the LSO for a zero-energy orbit in the Newtonian approximation; in the limit $e \to 1$ of the Keplerian regime, it corresponds to a critical periapsis $r_{\rm LSO} = a(1-e) = 8r_g$).

An MC run starts by injecting a test star in some initial phase-space position (a_0, j_0). This is typically chosen randomly, either just below $a_{\max}$, with j distributed isotropically above $j_{\rm LSO}$, to simulate a star diffusing from an isotropic galaxy into the MBH's radius of influence, or isotropically in the cusp's bulk according to the power-law cusp distribution, to simulate the initial conditions of an N-body simulation (see below), or the distribution of tracer stars (e.g., red giants) that mirror the distribution of cusp stars. The star is then evolved in small time increments dt, taking into account the stochastic changes in energy and angular momentum due to NR, the random changes in angular momentum due to RR, and the deterministic changes due to GW dissipation. The changes in the dimensionless energy $x = a_{\max}/a$ and angular momentum j are calculated separately for the different processes. The change due to NR diffusion is

$$\begin{aligned} dx^{\rm NR} &= D_x^{\rm NR} dt + \gamma_1 \sqrt{D_{xx}^{\rm NR} dt}\,, \\ dj^{\rm NR} &= D_j^{\rm NR} dt + \gamma_2 \sqrt{D_{jj}^{\rm NR} dt}\,, \end{aligned} \tag{63}$$

where the random normal variates γ_1 and γ_2 are first drawn from the bi-normal distribution, and then adjusted to obey the required correlation $\rho = D_{xj}/\sqrt{|D_{xx}D_{jj}|}$ by the transformation $\gamma_2 \to \gamma_1\rho + \gamma_2\sqrt{1-\rho^2}$. The change due to the effective RR diffusion is

$$dj^{\rm RR} = D_j^{\rm RR} dt + \gamma_3 \sqrt{D_{jj}^{\rm RR} dt}\,, \tag{64}$$

where γ_3 is an independent normal variate. Finally, the star is evolved in phase-space by

$$\begin{aligned} x(t+dt) &= x(t) + dx^{\rm NR} + dx^{\rm GW}, \\ j(t+dt) &= j(t) + dj^{\rm NR} + dj^{\rm RR} + dj^{\rm GW}, \end{aligned} \tag{65}$$

where $dx^{\rm GW}$ and $dj^{\rm GW}$ are the radiative losses due to GW emission.

The star is tracked in phase-space, and its position is recorded by snapshots taken at fixed intervals $\Delta t \gg dt$. Ultimately, after surviving for a time t_s, the star leaves the system at some terminal phase-space position (a_1, j_1) as a result of one of four possible outcomes. (1) Evaporation. The star's sma increases above $a_{\rm out}$. This happens to the majority of test stars. (2) Inspiral. The star's orbit decays until it crosses $a_i = \epsilon a_{\min}$ $(\epsilon \ll 1)$. (3) Plunge. The star crosses the LSO directly, at some $a_p > a_i$. (4) The star's assumed finite lifespan is exceeded. This is relevant for burning stars that are limited by stellar evolution, or for binaries that are also limited by dynamical evaporation (Section 6.1). It is not relevant for compact remnants. Note that the branching ratio between plunges and inspirals is independent of the exact value of the sma chosen to distinguish between the two outcomes (parameterized by ϵ), since their respective terminal phase-space positions are clearly separated ($\min a_p \gg a_i$, see Figure 15; $\epsilon = 0.1$ was typically used here).

Once a star exits the system, a new star is injected (reflection at $a_{\rm out}$ is equivalent to the injection of a new star at $(a_{\rm out}, j_1)$). This is repeated over some long accumulated time $T_{\rm sim} \gg t_s$ (typically $T_{\rm sim} \sim 100T_E$, Section 3.4). All of the test stars simulated over $T_{\rm sim}$, apart from the last one, which is omitted from the analysis, reach a definite outcome. For each test star, we record its trajectory in phase-space (in coarse Δt resolution); the total time it spent evolving in phase-space, t_s; the nature of the final outcome (evaporation, inspiral, plunge or end of lifespan); and its initial and terminal phase-space positions. The procedure is repeated as needed (typically $(10^3$–$10^4) \times T_{\rm sim}$), until enough test star statistics are collected. The snapshots of the phase-space positions, the survival times, and the final outcomes are then used to estimate the steady-state configuration and loss rates, as detailed below.

B.2. Representation of Physical Processes

The forms of the RR DCs depend on the background noise model and the precession of the test star (Bar-Or & Alexander 2014). We assume three optional noise models: white noise (equivalent to no precession), C^0 noise with exponential ACF (an Ornstein–Uhlenbeck process), and smooth C^∞ noise with a Gaussian ACF. Prograde GR in-plane precession is modeled by the 1PN approximation $\nu_{\rm GR} = 3\nu_r(a)(r_g/a)/j^2$, where $\nu_r(a)$ is the Keplerian radial (orbital) frequency. Mass precession is given exactly for an $\alpha = 2$ cusp by $\nu_M(a=2, j) = -[N(a)/Q]\nu_r j/(1+j)$ (MAMW11), and for other values of α by polynomial approximations of the exact integral (Alexander 2005). The magnitude of the RR DCs reflects the strength of the RR torques, $\tau_N \simeq 0.28\sqrt{1-j}\sqrt{N(2a)}GM_\star/a$, as derived from static wire simulations (Appendix D).

The RR DCs have explicit analytic forms that are easy to evaluate (Bar-Or & Alexander 2014),

$$D_{jj}^{\rm RR}(a,j) = 2T_c(a)\nu_J^2(j)/(1+[T_c(a)\nu_{\rm prec}(a,j)]^2), \quad (C^0 \text{ exponential ACF noise}), \tag{66}$$

$$D_{jj}^{\rm RR}(a,j) = 2T_c(a)\nu_J^2(j)\exp(-[T_c(a)\nu_{\rm prec}(a,j)]^2/\pi), \quad (C^\infty \text{ Gaussian ACF noise}), \tag{67}$$

where $\nu_J = \tau_N/J_c$ is the RR torque rate and ν_p is the precession frequency. The RR drift, which is not a true drift but arises from $D_{jj}^{\rm RR}$ (parametric drift), is given by

$$D_j^{\rm RR} = \frac{1}{2j}\frac{\partial(jD_{jj}^{\rm RR})}{\partial j}. \tag{68}$$

Two optional approximate background coherence timescale models are considered, which are assumed to be functions of a only: j-averaged mass precession, $T_M = \sqrt{\pi/2}\,\nu_M^{-1}(2a, \sqrt{1/2})$, and self-quenching $T_{\rm SQ} = Q\nu_r^{-1}(a)/\sqrt{N(2a)}$, where the approximate numeric prefactors (Bar-Or & Alexander 2014, footnote 7) were evaluated here specifically for $\alpha = 7/4$, but are generally insensitive to the exact value of α. Note that GR precession is not included in background coherence models of the MC, but it is included in approximate form ($\nu_{\rm prec} = |\nu_M + \nu_{\rm GR}|$) in analytic modeling of the coherence time (Section 3).

The NR DCs (Appendix C) involve multiple integrations that are too computationally expensive to perform on the fly. We therefore calculate them exactly beforehand on a 25×25 evenly spaced logarithmic grid extending between $a_{\rm min}$ and $a_{\rm max}$ and $j_{\rm min} = 10^{-3}$ and $j_{\rm max} = 1$, and then bi-linearly interpolated to any (a, j) as needed.

For each MC step, the time step was chosen as $\Delta t = \min(\Delta t_{\rm NR}, \Delta t_{\rm RR}, \Delta t_{\rm GW})$, where the time steps for the NR and RR were chosen by criteria similar to those used by Shapiro & Marchant (1978) for NR, and the time step for GW was chosen to be a small fraction of the GW dissipation times $\min(E/\dot{E}_{\rm GW}, J/\dot{J}_{\rm GW})$.

Three alternative analytic perturbative estimates for the rate of GW dissipation of energy and angular momentum were studied: those of Peters (1964), Hopman & Alexander (2006a), and Gair et al. (2006).

B.3. Steady-state Rate Estimates

The interpretation of the MC results in terms of loss rates depends on whether the test stars represent the underlying background cluster that is generating the NR and RR perturbations, or whether they are a separate trace population that is affected by the perturbations but does not contribute to them.

B.3.1. Test Stars as Background

In statistical steady state, stars that exit the system ($a_{\rm in} < a < a_{\rm out}$) are replaced at a rate that keeps their time-averaged number N fixed. A star that evaporates from the system back to the infinite reservoir at $a > a_{\rm out}$ (reflection at $a_{\rm out}$ is treated as evaporation) is replaced by another star from the reservoir, and so there is no net current through $a_{\rm out}$ due to evaporation. The situation is different for stars that end up in the MBH, whether by plunge or inspiral. Since they are permanently removed from the system, a net current of stars through $a_{\rm out}$, from the reservoir into the system, is required to compensate for their loss. Both the stars that evaporate and those that are lost[12] contribute to the total mean number of stars in steady state. In our models, this number is an input parameter, determined by the assumed background cusp.

Let P_k be the probability (branching ratio) for outcome $k = 0, 1, 2, \ldots$, where $k = 0$ denotes evaporation and $k > 0$ denote the various loss channels, so that $\Sigma_k P_k = 1$. The MC statistics provide estimates of the branching ratios, $P_k = n_k/n_{\rm sim}$, where $n_{\rm sim}$ is the total number of test stars whose phase-space trajectories were simulated, and n_k is the number of times outcome k has been reached. This translates to event rates by requiring that the total number of stars be on average $N = \sum_k N_k = \sum_k \Gamma_k \bar{t}_k$, where Γ_k is the rate of outcome k and $\bar{t}_k = n_k^{-1}\sum_j t_k^{(j)}$ is the mean survival time in the n_k simulations that had outcome k. The rate for each channel is related to the total rate of all outcomes, Γ, by $\Gamma_k = \Gamma P_k$, where $\Gamma = N/\bar{t}_s$ and $\bar{t}_s$ is the overall mean survival time, irrespective of outcome.[13] It then follows that the event rates are

$$\Gamma_k = (N/\bar{t}_s)P_k. \tag{69}$$

The total replenishment rate that is required to keep the system in steady state is then the sum over all of the loss channels,

$$\Gamma_{\rm loss} = \sum_{k>0}\Gamma_k. \tag{70}$$

Loss-rates estimated by direct N-body simulations (MAMW11; BAS14) can in principle be compared to the rates derived by MC simulations of scaled-down nuclei where the test stars are treated as representative of the background. However, this comparison is complicated by the fact that the N-body systems are not necessarily in steady state (the initial configuration may not be the steady-state one, and/or stars lost in the course of the simulation are not replenished), and do not have fixed boundaries or boundary conditions. The comparisons discussed in Section 4.2 and Table 2 are approximate. The MC loss rates were estimated for a non-equilibrium cusp that corresponds to the initial conditions of the N-body simulations, and the MC rates were compared to the rates early in the simulations, at times where the N-body configuration is still close to its initial state.

B.3.2. Test Stars as a Trace Population

It is of interest to consider how a small population of tracer stars, which are injected into the cusp by some dynamical or evolutionary mechanism, evolves dynamically with time and is lost via the various channels. Some possible injection mechanisms are capture by tidal separation of an incoming binary (Hills 1988), in which case the injection point in phase-space is a tight eccentric orbit deep inside the cusp; the formation of a red giant when a background star evolves off the main sequence, in which case the injection point reflects the background distribution of progenitors; or the formation of a

[12] To simplify bookkeeping, a test star that would have wandered back and forth across $a_{\rm out}$ and is finally destroyed by the MBH is not counted as a single star (i.e., a single trajectory). Once it leaves the cusp ($a > a_{\rm out}$), it is considered to have evaporated. The next crossing into the cusp $a < a_{\rm out}$ is identified as the beginning of the phase-space trajectory of a new star.

[13] This follows from summation over all channels: $N = \sum_k \Gamma_k \bar{t}_k = \Gamma\sum_k P_k \bar{t}_k = \Gamma\sum_k (n_k/n_{\rm sim})(n_k^{-1}\sum_j t_k^{(j)}) = \Gamma\left(\sum_{k,j} t_k^{(j)}\right)/n_{\rm sim} = \Gamma\bar{t}_s$.

massive blue giant in a fragmenting gas disk, in which case the injection point is a low-eccentricity orbit.

When the test stars in the MC simulation represent a tracer population, the total rate of all outcomes is determined by the assumed injection rate, and is no longer related to the total number of background stars. In this case, the MC does not predict the event rates Γ_k, but rather the branching ratios P_k.

APPENDIX C
TWO-BODY DIFFUSION COEFFICIENTS

This appendix summarizes for completeness the derivation of the NR DCs in angular momentum and energy space, for stars orbiting in a spherical potential, as calculated by Shapiro & Marchant (1978), Cohn & Kulsrud (1978), and Cohn (1979) based on the earlier work of Spitzer (1962) and Spitzer & Shapiro (1972), and further generalizes that treatment by explicitly including the dependence of the DCs on the test star's mass (however, this generalization is not used in our dynamical modeling of the loss cone, which assumes for simplicity a single-mass stellar population).

Consider a star of mass m moving in a spherical potential, $\phi = \phi(r)$, with velocity $\boldsymbol{v}$. The binding energy and angular momentum are

$$E = \phi(r) - \frac{1}{2}v^2, \tag{71}$$

$$\boldsymbol{J} = \boldsymbol{r} \times \boldsymbol{v}. \tag{72}$$

Note that here E is the positively defined orbital energy and $\phi(r)$ is the positively defined potential. Due to gravitational encounters with the field stars, the star changes its velocity, $\boldsymbol{v}$, to $\boldsymbol{v}' = \boldsymbol{v} + \Delta\boldsymbol{v}$. Consider the orthonormal basis

$$\hat{\boldsymbol{v}} = \boldsymbol{v}/v, \tag{73}$$

$$\hat{\boldsymbol{J}} = \boldsymbol{J}/J = \boldsymbol{r} \times \boldsymbol{v}/|\boldsymbol{r} \times \boldsymbol{v}|, \tag{74}$$

$$\hat{\boldsymbol{w}} = \boldsymbol{v} \times \boldsymbol{J}/|\boldsymbol{v} \times \boldsymbol{J}| = \boldsymbol{v} \times (\boldsymbol{r} \times \boldsymbol{v}) = (v\hat{\boldsymbol{r}} - v_r\hat{\boldsymbol{v}})/v_t. \tag{75}$$

In this basis, the change in velocity is

$$\Delta\boldsymbol{v} = \Delta v_\parallel \hat{\boldsymbol{v}} + \Delta\boldsymbol{v}_\perp, \tag{76}$$

where

$$\Delta\boldsymbol{v}_\perp = \Delta v_J \hat{\boldsymbol{J}} + \Delta v_w \hat{\boldsymbol{w}}, \tag{77}$$

and

$$\Delta v_\perp = \sqrt{(\Delta v_J)^2 + (\Delta v_w)^2}. \tag{78}$$

The change in energy is

$$\begin{aligned}\Delta E &= -\frac{1}{2}(v'^2 - v^2) = -\frac{1}{2}(\Delta v)^2 - \boldsymbol{v}\cdot\Delta\boldsymbol{v} \\ &= -\frac{1}{2}(\Delta v_\parallel)^2 - \frac{1}{2}(\Delta v_\perp)^2 - v\Delta v_\parallel.\end{aligned} \tag{79}$$

The position vector, $\boldsymbol{r}$, is

$$\boldsymbol{r} = r(v_t/v)\hat{\boldsymbol{w}} + r(v_r/v)\hat{\boldsymbol{v}}, \tag{80}$$

where v_r and v_t are the radial and transversal velocities. Therefore, the change in the radial velocity is

$$\Delta v_r = \frac{\Delta\boldsymbol{v}\cdot\boldsymbol{r}}{r} = \frac{v_r}{v}\Delta v_\parallel + \frac{v_t}{v}\Delta v_w. \tag{81}$$

The change in the transverse velocity up to second order in $\Delta v/v$ is

$$\Delta v_t = v_t\frac{\Delta v_\parallel}{v} - v_r\frac{\Delta v_w}{v} + \frac{1}{2}\frac{\Delta v_J^2}{v_t}. \tag{82}$$

The change in the angular momentum is

$$\Delta\boldsymbol{J} = \boldsymbol{r}\times\Delta\boldsymbol{v} = J\left(\frac{\Delta v_\parallel}{v} - \frac{v_r}{v_t}\frac{\Delta v_w}{v}\right)\hat{\boldsymbol{J}} + J\frac{\Delta v_J}{v}\left(\frac{v_r}{v_t}\hat{\boldsymbol{w}} - \hat{\boldsymbol{v}}\right), \tag{83}$$

and the change in angular momentum magnitude (up to second order in $\Delta v/v$) is

$$\Delta J = J\frac{\Delta v_\parallel}{v} - rv_r\frac{\Delta v_w}{v} + \frac{1}{2}\frac{r^2}{J}\Delta v_J^2. \tag{84}$$

Using Equations (79) and (84), we can obtain the local (orbital phase dependant) DCs in terms of the velocity DCs,

$$\langle\Delta E\rangle = -\frac{1}{2}\langle(\Delta v_\parallel)^2\rangle - \frac{1}{2}\langle(\Delta v_\perp)^2\rangle - v\langle\Delta v_\parallel\rangle, \tag{85}$$

$$\langle(\Delta E)^2\rangle = v^2\langle(\Delta v_\parallel)^2\rangle, \tag{86}$$

$$\langle\Delta J\rangle = \frac{J}{v}\langle\Delta v_\parallel\rangle + \frac{r^2}{4J}\langle(\Delta v_\perp)^2\rangle, \tag{87}$$

$$\langle(\Delta J)^2\rangle = \frac{J^2}{v^2}\langle(\Delta v_\parallel)^2\rangle + \frac{1}{2}\left(r^2 - \frac{J^2}{v^2}\right)\langle(\Delta v_\perp)^2\rangle, \tag{88}$$

$$\langle\Delta E\Delta J\rangle = -J\langle(\Delta v_\parallel)^2\rangle, \tag{89}$$

where we omitted higher-order terms in $\Delta v/v$, and used $\langle\Delta v_w\rangle = \langle\Delta v_J\rangle = 0$ and $\langle\Delta v_J^2\rangle = \langle\Delta v_w^2\rangle = \langle\Delta v_\perp^2\rangle/2$.

The local velocity DCs are (Binney & Tremaine 2008)

$$\langle\Delta v_\parallel\rangle = -\kappa\frac{m+m_a}{m_a}\int_0^v dv_a\frac{v_a^2}{v^2}f_a(v_a), \tag{90}$$

$$\langle\Delta v_\parallel^2\rangle = \frac{2}{3}\kappa\left[\int_0^v dv_a\frac{v_a^4}{v^3}f_a(v_a) + \int_v^\infty dv_a v_a f_a(v_a)\right], \tag{91}$$

$$\begin{aligned}\langle\Delta v_\perp^2\rangle = \frac{2}{3}\kappa\Bigg[&\int_0^v dv_a\left(\frac{3v_a^2}{v} - \frac{v_a^4}{v^3}\right)f_a(v_a) \\ &+ 2\int_v^\infty dv_a v_a f_a(v_a)\Bigg],\end{aligned} \tag{92}$$

where $\kappa = (4\pi G m_a)^2\ln\Lambda$, and $f_a(v_a)$ and m_a are the velocity DF and mass of the field stars, and where we assume that the velocity DF is isotropic. In that case, the velocity DF can be written in terms of the orbital energies of the field stars. Using

$$f_a(v)vdv = -f(E)dE, \tag{93}$$

we obtain

$$\langle\Delta E\rangle = \kappa\left[\frac{m}{m_a}\int_E^\phi (v_a/v)f_a(E_a)dE_a - \int_{-\infty}^E f_a(E_a)dE_a\right], \tag{94}$$

$$\langle\Delta E^2\rangle = \frac{2}{3}\kappa v^2\left[\int_E^\phi dE_a(v_a/v)^3 f_a(E_a) + \int_{-\infty}^E dE_a f_a(E_a)\right], \tag{95}$$

$$\langle \Delta J\rangle = \kappa\left\{-\frac{J}{v^2}\left(\frac{m}{m_a}+1\right)\int_E^{\phi} dE_a\,(v_a/v)f_a\,(E_a) + \frac{r^2}{6J}\int_E^{\phi} dE' f_a\,(E_a)[3(v_a/v)-(v_a/v)^3] + \frac{r^2}{3J}\int_{-\infty}^{E} dE_a f_a\,(E_a)\right\}, \quad (96)$$

$$\langle \Delta J^2\rangle = \frac{\kappa}{v^2}\left\{J^2\int_E^{\phi}(v_a/v)^3 dE_a f_a\,(E_a) - J^2\int_E^{\phi} dE_a\,(v_a/v)f_a\,(E_a) + \frac{r^2v^2}{3}\int_E^{\phi} dE_a\,(3(v_a/v)-(v_a/v)^3)f_a\,(E_a) + \frac{2}{3}r^2v^2\int_{-\infty}^{E} dE_a f_a\,(E_a)\right\}, \quad (97)$$

and

$$\langle \Delta E\Delta J\rangle = -J\frac{2}{3}\kappa\left[\int_E^{\phi} dE_a\,(v_a/v)^3 f_a\,(E_a) + \int_{-\infty}^{E} dE_a f_a\,(E_a)\right]. \quad (98)$$

The corresponding orbit-averaged DCs are given by

$$D_E = 2P^{-1}\int_{r_a}^{r_p}\langle \Delta E\rangle dr/v_r, \quad (99)$$

$$D_{EE} = 2P^{-1}\int_{r_a}^{r_p}\langle (\Delta E)^2\rangle dr/v_r, \quad (100)$$

$$D_J = 2P^{-1}\int_{r_a}^{r_p}\langle \Delta J\rangle dr/v_r, \quad (101)$$

$$D_{JJ} = 2P^{-1}\int_{r_a}^{r_p}\langle (\Delta J)^2\rangle dr/v_r, \quad (102)$$

$$D_{EJ} = 2P^{-1}\int_{r_a}^{r_p}\langle \Delta E\Delta J\rangle dr/v_r. \quad (103)$$

where P is the orbital period.

Assuming that the potential is Keplerian, $\phi = M_\bullet/r$, the energy and angular momentum are

$$E = \frac{M_\bullet}{r} - \frac{1}{2}v_r^2 - \frac{1}{2}\frac{J^2}{r^2} = \frac{GM_\bullet}{2a}, \quad (104)$$

$$J = rv_t, \quad (105)$$

where a is the sma. Using $x = (r/a-1)e^{-1}$, the orbital average is

$$\langle D\rangle_{\circlearrowleft} = \frac{1}{\pi}\int_{-1}^{1} D\frac{1+xe}{\sqrt{1-x^2}}dx, \quad (106)$$

where $e = \sqrt{1-J^2/J_c^2}$ is the eccentricity of the orbits and $J_c = \sqrt{GM_\bullet a}$ is the maximal (circular) angular momentum. The orbital-averaged DCs are therefore

$$D_E/E = -\Gamma_0 + \frac{m}{m_a}\Gamma_{110}, \quad (107)$$

$$D_{EE}/E^2 = \frac{4}{3}\Gamma_{13-1} + \frac{4}{3}\Gamma_0, \quad (108)$$

$$D_J/J_c = \left[\frac{5-3j^2}{12}\Gamma_0 - j^2\frac{m+m_a}{2m_a}\Gamma_{111} + \Gamma_{310} - \frac{1}{3}\Gamma_{330}\right]/j, \quad (109)$$

$$D_{JJ}/J_c^2 = \frac{5-3j^2}{6}\Gamma_0 + \frac{1}{2}j^2\Gamma_{131} - \frac{1}{2}j^2\Gamma_{111} + 2\Gamma_{310} - \frac{2}{3}\Gamma_{330}, \quad (110)$$

$$D_{EJ}/(EJ_c) = -\frac{2}{3}j(\Gamma_0 + \Gamma_{130}), \quad (111)$$

where

$$\Gamma_0 = \kappa\int_{-\infty}^{1} ds f_a\,(sE), \quad (112)$$

and

$$\Gamma_{ijk}\,(E, J) = 2^{1+k-i}\frac{\kappa}{\pi}\int_{-1}^{1} dx\int_{1}^{2/(1+ex)} ds \times f_a\,(sE)\frac{1}{\sqrt{1-x^2}}\frac{(r/a)^i}{(v^2/E)^k}(v_a/v)^j. \quad (113)$$

We can write the Γ_{ijk} functions in terms of the Cohn & Kulsrud (1978) F_i functions (see the definitions after Equation (24) there)

$$F_0 = E\Gamma_0, \quad (114)$$

$$F_1 = E\Gamma_{110}, \quad (115)$$

$$F_2 = E\Gamma_{111}, \quad (116)$$

$$F_3 = E\Gamma_{310}, \quad (117)$$

$$F_4 = E\Gamma_{13-1}, \quad (118)$$

$$F_5 = E\Gamma_{130}, \quad (119)$$

$$F_6 = E\Gamma_{131}, \quad (120)$$

$$F_7 = E\Gamma_{330}. \quad (121)$$

It is sometimes useful to consider diffusion in the dimensionless normalized angular momentum, j. For a general coordinates transformation $\boldsymbol{x}' = \boldsymbol{x}'(\boldsymbol{x})$, the new DCs are given by (e.g., Risken 1989)

$$D_l' = \frac{\partial x_l'}{\partial x_k}D_k + \frac{1}{2}\frac{\partial^2 x_l'}{\partial x_r\partial x_k}D_{rk}, \quad (122)$$

$$D_{lm}' = \frac{\partial x_l'}{\partial x_r}\frac{\partial x_m'}{\partial x_k}D_{rk}. \quad (123)$$

Thus, in the E, j coordinates, the j-related DCs are

$$D_j = D_J/J_c + \frac{1}{2}jD_E/E + \frac{1}{2}D_{EJ}/(J_cE) - \frac{1}{8}jD_{EE}/E^2, \quad (124)$$

$$D_{jj} = D_{JJ}/J_c^2 + \frac{1}{4}j^2D_{EE}/E^2 + jD_{EJ}/(J_cE), \quad (125)$$

$$D_{Ej}/E = D_{EJ}/(J_cE) + \frac{1}{2}jD_{EE}/E^2. \quad (126)$$

Similarly, for the E, $R = j^2$ coordinates, the R-related DCs are (see also Cohn & Kulsrud 1978; Cohn 1979)

$$D_R = 2jD_j + D_{jj} = 2jD_J/J_c + D_{JJ}/J_c^2 + j^2\frac{D_E}{E} + 2\frac{D_{EJ}}{J_cE}j, \quad (127)$$

$$D_{\rm RR} = 4j^2D_{jj} = 4j^2D_{JJ}/J_c^2 + j^4D_{EE}/E^2 + 4j^3\frac{D_{EJ}}{J_cE}, \quad (128)$$

$$D_{ER} = 2jD_{Ej} = 2j\frac{D_{EJ}}{J_cE} + j^2D_{EE}/E^2. \quad (129)$$

APPENDIX D
THE STATISTICAL PROPERTIES OF THE RESONANT TORQUES

We describe and measure here the residual torques acting on a test star due to the near-Keplerian orbits of the background stars. The torque vector τ depends on the test star's angular momentum $\boldsymbol{J}$ and argument of periapsis ω. We discuss the symmetries of these torques and the scaling with the test star's orbital parameters, and empirically measure the torques by static random background simulations.

D.1. Geometrical Description

Consider the angular momentum vector in some fixed reference frame:

$$\boldsymbol{J} = J\begin{pmatrix}\ell_x\\ \ell_y\\ \ell_z\end{pmatrix} = \begin{pmatrix}\sqrt{J^2 - J_z^2}\cos\phi\\ \sqrt{J^2 - J_z^2}\sin\phi\\ J_z\end{pmatrix},$$

where $\hat{\boldsymbol{\ell}} = (\ell_x, \ell_y, \ell_z)$ is the unit vector in the direction of $\boldsymbol{J}$ in a fixed Cartesian reference system (x, y, z). The torque is derived from the Hamiltonian by $\boldsymbol{\tau} = \dot{\boldsymbol{J}} = \{J, H\}$, where $\{...\}$ denotes the Poisson brackets. It is more convenient to represent the torques in an orthonormal spherical coordinate system (J, ϕ, u), where $u \equiv \ell_z = \cos\theta$, with the associated unit vectors $\hat{\boldsymbol{e}}_i = (\partial\boldsymbol{J}/\partial i)/|\partial\boldsymbol{J}/\partial i|$ for $i \in \{J, \phi, u\}$. The change in the angular momentum's magnitude is then $\dot{J} = \tau_J$, and the changes in its direction are described by the angular torques $\dot{\phi} = \tau_\phi/(J\sqrt{1-u^2})$ and $\dot{u} = \tau_u\sqrt{1-u^2}/J$, where the torque vector is given by

$$\boldsymbol{\tau} = \tau_J\hat{\boldsymbol{e}}_J + \tau_\phi\hat{\boldsymbol{e}}_\phi + \tau_u\hat{\boldsymbol{e}}_u. \quad (130)$$

In the reference frame of the orbit, we can define the orbital torques in the direction of the semimajor axis, τ_a, and the semiminor axis, τ_b,

$$\tau_a = \boldsymbol{\tau}\cdot\hat{\boldsymbol{a}} = \sin\omega\tau_\phi - \cos\omega\tau_u, \quad (131)$$

$$\tau_b = \boldsymbol{\tau}\cdot\hat{\boldsymbol{b}} = \cos\omega\tau_\phi + \sin\omega\tau_u, \quad (132)$$

where $\hat{a}$ and $\hat{b}$ are the direction semimajor and semiminor axes

$$\hat{\boldsymbol{a}} = \hat{\boldsymbol{e}}_\phi\sin\omega - \hat{\boldsymbol{e}}_u\cos\omega, \quad (133)$$

$$\hat{\boldsymbol{b}} = \hat{\boldsymbol{e}}_\phi\cos\omega + \hat{\boldsymbol{e}}_u\sin\omega. \quad (134)$$

D.2. Statistical Properties

Define the typical (Poisson) torque (Gürkan & Hopman 2007) as

$$\tilde{\tau}_N \equiv \frac{\sqrt{N(2a)}}{a}GM_\star = J_c\nu_r\frac{\sqrt{N(2a)}}{Q}. \quad (135)$$

where ν_r is the Keplerian radial (orbital) frequency. The mean-squared values of the torques are given by

$$\langle\tau_J^2\rangle/\tilde{\tau}_N^2 = T_\parallel(a, J), \quad (136)$$

$$\langle\tau_\phi^2\rangle/\tilde{\tau}_N^2 = \frac{1}{2}[T_+(a, J) + T_-(a, J)\cos 2\omega], \quad (137)$$

$$\langle\tau_u^2\rangle/\tilde{\tau}_N^2 = \frac{1}{2}[T_+(a, J) - T_-(a, J)\cos 2\omega], \quad (138)$$

and the cross terms are

$$\langle\tau_J\tau_\phi\rangle = \langle\tau_J\tau_u\rangle = 0, \quad (139)$$

$$\langle\tau_u\tau_\phi\rangle/\tilde{\tau}_N^2 = T_-\sin(2\omega)/2, \quad (140)$$

where we defined

$$T_\parallel \equiv \langle\tau_J^2\rangle/\tilde{\tau}_N^2, \quad (141)$$

$$T_+ \equiv (\langle\tau_a^2\rangle + \langle\tau_b^2\rangle)/\tilde{\tau}_N^2, \quad (142)$$

$$T_- \equiv (\langle\tau_b^2\rangle - \langle\tau_a^2\rangle)/\tilde{\tau}_N^2. \quad (143)$$

Note that since the torques $\langle\tau_J^2\rangle$, $\langle\tau_a^2\rangle$, and $\langle\tau_b^2\rangle$ are measured in the orbital plane, they have no angular dependencies, and neither do $T_\parallel$, T_+, and T_-.

The symmetry of the orbits is such that for a circular orbit (i.e., $J = J_c$), there is no preferred direction in the orbital plane and therefore $\langle\tau_a^2\rangle = \langle\tau_b^2\rangle$ and $T_\parallel(a, J_c) = T_-(a, J_c) = 0$. For radial orbits ($J = 0$), τ_a vanish and $\langle\tau_b^2\rangle = \langle\tau_\parallel^2\rangle$, therefore, $T_\parallel(a, 0) = T_+(a, 0) = T_-(a, 0)$.

D.3. Measuring the Torques

We used static wires simulations (e.g., Gürkan & Hopman 2007) to measure the j dependence of the residual torque ($j = \sqrt{1-e^2}$) on a test orbit with sma a and eccentricity e. This was carried out by simulating in the background as many fixed Keplerian wire orbits, and measuring the three components of the orbital torques τ_a, τ_b, and τ_J, for many independent random realizations of the background, integrating over the orbit of the test star and the orbits of the field stars with the efficient Touma et al. (2009) algorithm. We decomposed the measured orbital torques to $T_\parallel$, T_+, and T_- and fitted them to a second-order polynomial in j. The best-fit results are (Figure 20)

$$T_\parallel \approx 0.08(1-j)(1-j/8), \quad (144)$$

$$T_- \approx 0.08(1-j)(1+j/4), \quad (145)$$

$$T_+ \approx 0.08(1-5j/8). \quad (146)$$

The residual torque in the $\boldsymbol{J}$ direction is therefore

$$\sqrt{\langle\tau_J^2\rangle}/\tilde{\tau}_N \approx 0.28\sqrt{(1-j)(1-j/8)} \approx 0.28\sqrt{1-j}. \quad (147)$$

This analytic fit is consistent with the Gürkan & Hopman (2007) results (fitted to $\sqrt{\langle\tau_J^2\rangle}/\tilde{\tau}_N \approx 0.25\sqrt{1-j^2}$; see Equation (13) there) but is better, since it matches their data

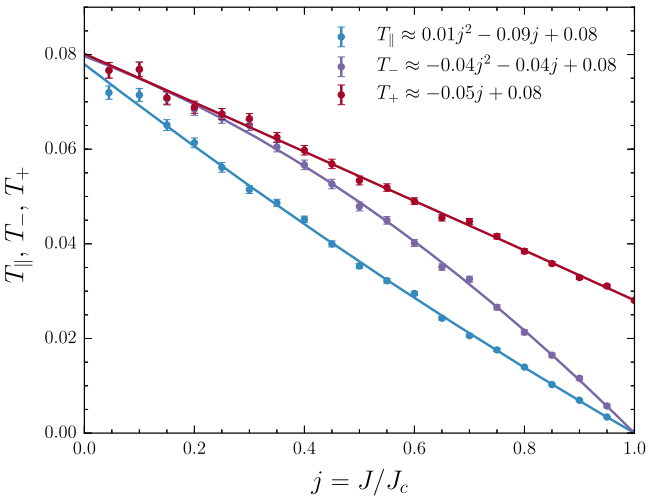

Figure 20. Mean-square of the torques (normalized to the typical (Poisson) torque $\tilde{\tau}$) as a function of the angular momentum. The torques measured from static wires simulations (circles) are approximated by polynomial fits (solid lines). As expected from considerations of symmetry, for a circular orbit ($J = J_c$), there is no torque in the $\boldsymbol{J}$ direction and both perpendicular torques (τ_b, τ_b) are equal, i.e., $T_\parallel = T_- = 0$. As $J \to 0$, the orbit's geometry approaches a rod, and the torque in the $\hat{a}$ direction vanishes because its lever arm goes to zero, while the torque in the $\boldsymbol{J}$ and $\hat{b}$ directions become equal, i.e., $T_\parallel = T_+ = T_-$.

over the entire j range. For the out-of-plane torques, we obtain

$$\sqrt{\tau_\perp^2}\Big/\tilde{\tau}_N \equiv \sqrt{\tau_a^2 + \tau_b^2} \approx 0.28\sqrt{1 - 5j/8}\,. \tag{148}$$

This is different from the Gürkan & Hopman (2007) results (fitted to $\sqrt{\langle\tau_\perp^2\rangle}\Big/\tilde{\tau}_N \approx 0.28(3\Big/2 - j^2)$ see Equation (14) there). However, the discrepancy can be traced to an error in their randomization procedure (Equations ((9)–(11)) there), which is not truly isotropic.

Using N-body simulations, Eilon et al. (2009) measured the isotropic averaged residual torques. They defined and measured the j-averaged quantities

$$\beta_s^2 = \frac{Q^2}{N_{\rm tot}}\langle j^2 P^2/J_c^2\rangle = 4\pi^2\,\langle\tau_\parallel^2\Big/\tilde{\tau}^2\rangle, \tag{149}$$

and

$$\beta_v^2 = \frac{Q^2}{N_{\rm tot}}\sqrt{\langle|\dot{\boldsymbol{J}}|^2 P^2/J_c^2\rangle} = 4\pi^2\,\langle(\tau_\parallel^2 + \tau_\perp^2)\Big/\tilde{\tau}^2\rangle. \tag{150}$$

Averaging the results here over j, we obtain $\beta_s \approx 1.0$ and $\beta_v \approx 1.7$ for $\gamma = 7/4$, in agreement with Eilon et al. (2009).[14]

APPENDIX E
THE MAXIMAL ENTROPY PRINCIPLE

The maximal entropy principle (MEP) has been shown to be a powerful tool in determining the steady-state (or quasi-steady state) of dynamical systems. In particular, it was studied extensively in the context of collisionless self-gravitating systems (e.g., Lynden-Bell 1967). Here, we examine the more restricted problem of a near-Keplerian system. The MEP is relevant for stellar systems only when the interactions conserve energy. We prove here that the energy-conserving RR DCs used in this study comply with the MEP, and so do the J-only NR DCs.

In a system where a central object of mass $M_\bullet$ dominates the potential (e.g., planetary systems, nuclear clusters), stars move on nearly Keplerian orbits. That is, the potential is almost regular and the orbital elements are almost constant. In particular, since the potential varies on much longer timescales than the orbital time, the potential can be orbit-averaged, and therefore orbital energy is conserved. Furthermore, the conserved orbital energy equals the Keplerian energy to first order in $M_{\rm enc}/M_\bullet \ll 1$, where $M_{\rm enc}$ is the enclosed stellar mass within a given orbit.

The entropy of the system is given by

$$S = -\int d^3\boldsymbol{r}d^3\boldsymbol{v} f(\boldsymbol{r},\boldsymbol{v})\log f(\boldsymbol{r},\boldsymbol{v}), \tag{151}$$

where f is the stellar DF. Since we assume that the Keplerian energy E of each star relative to the MBH is conserved, the density $n(E)$ is conserved, where

$$n(E;f) = \int d^3\boldsymbol{r}d^3\boldsymbol{v} f(\boldsymbol{r},\boldsymbol{v})\delta\left(E - \frac{GM_\bullet}{r} + \frac{1}{2}v^2\right). \tag{152}$$

This implies the conservation of the total Keplerian energy $E_\bullet[f] = \int En(E;f)dE$ and of the total number of stars $N_{\rm tot}[f] = \int n(E;f)dE$. Although the Keplerian energy of each star, and therefore its sma, are fixed, the orbital orientations and eccentricities are perturbed by the interaction with other stars. Conservation of the total energy $E_{\rm tot} = E_\bullet[f] + E_\star[f]$ then implies the conservation of the total potential energy due to the self-interactions of the stellar background:

$$E_\star[f] = \frac{1}{2}\int d^3\boldsymbol{r}\int d^3\boldsymbol{v} f(\boldsymbol{r},\boldsymbol{v})\psi(\boldsymbol{r};f), \tag{153}$$

where

$$\psi(\boldsymbol{r};f) = GM_\star\int d^3\boldsymbol{r}'\int d^3\boldsymbol{v}'\frac{f(\boldsymbol{r}',\boldsymbol{v}')}{|\boldsymbol{r} - \boldsymbol{r}'|}, \tag{154}$$

is the star–star potential. The total angular momentum is also conserved,

$$\boldsymbol{L}_{\rm tot}[f] = \int d^3\boldsymbol{r}d^3\boldsymbol{v} f(\boldsymbol{r},\boldsymbol{v})\boldsymbol{L}, \tag{155}$$

where $\boldsymbol{L} = \boldsymbol{r}\times\boldsymbol{v}$. Using the Lagrange multipliers β, $\boldsymbol{b}$, and $\lambda(E)$, we write the target function

$$\begin{aligned}\mathcal{S} = S &+ \beta\left(\int d^3\boldsymbol{r}d^3\boldsymbol{v} f(\boldsymbol{r}',\boldsymbol{v}')\psi(\boldsymbol{r};f) - 2E_\star\right)\\ &+ \boldsymbol{b}\cdot\left(\int d^3\boldsymbol{r}d^3\boldsymbol{v} f(\boldsymbol{r},\boldsymbol{v})\boldsymbol{L} - \boldsymbol{L}_{\rm tot}\right)\\ &+ \int d^3\boldsymbol{r}d^3\boldsymbol{v} f(\boldsymbol{r},\boldsymbol{v})\lambda(E)\left(\int d^3\boldsymbol{r}'d^3\boldsymbol{v}' f(\boldsymbol{r}',\boldsymbol{v}')\right.\\ &\left.\times\,\delta\left(E - \frac{GM_\bullet}{r'} + \frac{1}{2}v'^2\right) - n_0(E)\right),\end{aligned} \tag{156}$$

[14] We correct two issues in the comparison to the Gürkan & Hopman (2007) results made by Eilon et al. (2009, Section 4.3). First, since β_s and β_v are the rms values, the average values of Gürkan & Hopman (2007) should be estimated by $\langle\beta_s\rangle = \sqrt{\langle\beta_s^2(e)\rangle}$ and $\langle\beta_v\rangle = \sqrt{\langle\beta_v^2\rangle}$. Second, $\beta_v^2 = \tau_\parallel^2 + \tau_\perp^2$, defined by Eilon et al. (2009), should be compared to the sum $\tau_\parallel^2 + \tau_\perp^2$ of the quantities defined in Gürkan & Hopman (2007).

which is minimized by requiring

$$\frac{\delta \mathcal{S}}{\delta f} = -\log f - 1 + \beta\psi(\boldsymbol{r}; f) + \boldsymbol{b}\cdot\boldsymbol{L} + 2\lambda(E)n(E; f) = 0. \quad (157)$$

Therefore, the DF that maximizes the entropy is

$$f(\boldsymbol{r}, \boldsymbol{v}) = A(E)e^{\beta\psi + \boldsymbol{b}\cdot\boldsymbol{L}}, \quad (158)$$

where $A(E) = \exp[2\lambda(E)n(E)]$, and β and $\boldsymbol{b}$ are constants determined by the constraints on $n(E)$, $E_\star$, and $\boldsymbol{L}_{\rm tot}$. Note that an isotropic system must have $b = 0$ to ensure that the DF does not depend on $\boldsymbol{L}$, and must also have $\beta = 0$ since the star–star potential depends on L even in an isotropic system.[15]

E.1. Fluctuation Dissipation Relation for a Spherically Symmetric System

Since stars are assumed to move on Keplerian orbits, it is convenient to work in action-angle coordinates. Then, the Hamiltonian is given by (e.g., Sridhar et al. 1999; Touma & Sridhar 2012)

$$H = H_{\rm K} + \bar{\psi} \quad (159)$$

where $H_{\rm K}$ is the Keplerian Hamiltonian, which is constant, and $\bar{\psi}$ is the orbital-averaged star–star interaction potential. Choosing the z coordinate in the direction of the total angular momentum, the general steady-state density is

$$n(E, J, J_z) = n(E)\frac{e^{\beta\bar{\psi} + bL_z}}{\int\int e^{\beta\bar{\psi} + bL_z} dJ_z dJ}, \quad (160)$$

which is expressed for convenience in terms of the conserved energy E instead of the related conserved action $I = \sqrt{GM_\bullet a} = GM_\bullet/\sqrt{2|E|}$.

For a spherical symmetric system with $\boldsymbol{L}_{\rm tot} = 0$, the steady-state DF is given by an implicit integral equation:

$$n(E, J) = n(E)\frac{2Je^{\beta\bar{\psi}(E,J,n)}}{\int e^{\beta\bar{\psi}(E,J,n)} dJ^2}. \quad (161)$$

MEP considerations do not require the additional assumptions that go into the FP equation (e.g., diffusion described by a Markovian process), and the MEP solution is independent of the path that the system took to reach it from its initial conditions. Therefore, the MEP solution must also be satisfied by the FP equation in steady state. This enforces a connection between the DCs, which is known as the fluctuation–dissipation (F–D) relation.

For the symmetries and conserved quantities of the system studied here (described by Equation (161)), the functional form of the F–D relation is derived from the steady-state zero-flux FP equation (e.g., Section 7.4.3 Binney & Tremaine 2008)

$$D_J n(E, J) = \frac{1}{2}\frac{\partial}{\partial J}(D_{JJ} n(E, J)), \quad (162)$$

which yields the F–D relation

$$2JD_J e^{\beta\psi} = \frac{\partial}{\partial J}[JD_{JJ} e^{\beta\psi}]. \quad (163)$$

[15] The L dependence of ψ arises from the eccentricity dependence of the enclosed mass seen by the star. This is manifested dynamically by the retrograde evolution of the argument of periapsis—mass precession.

From this point on, we will restrict ourselves to isotropic systems which reach an MEP solution with $\beta = 0$. This is a solution of the form $n(E, J) = 2n(E)J/J_c^2$. This means that DCs derived under the assumption of an isotropic background (and fixed Keplerian energy) must satisfy the F–D relation $2JD_J = \partial(JD_{JJ})/\partial J$. In particular, the RR DCs that were used in this study obey this relation, as required (Bar-Or & Alexander 2014).

E.2. Fluctuation–Dissipation Relation for J-only Two-body Relaxation

For NR, the coherence time is shorter than the orbital period (Section 2), and therefore orbital energy is not conserved. However, in Section 3, we argue that the flow pattern in phase-space justifies the approximate treatment of the fast J diffusion as separate from the slower E diffusion. We now use this property to show that in that case, the NR J-only DCs also satisfy the fluctuation dissipation relation, which is a partial test of the validity of the general NR DCs and is used in Section 3.

Using Equation (79) and setting $\Delta E = 0$, we obtain

$$2v\Delta v_\parallel + \Delta v^2 = 0. \quad (164)$$

Therefore, to first order in $\Delta v^2/v^2$, the local diffusion coefficients are

$$\langle \Delta J\rangle = \frac{1}{2J}\left(\frac{r^2}{2} - \frac{J^2}{v^2}\right)\langle \Delta v_\perp^2\rangle, \quad (165)$$

$$\langle (\Delta J)^2\rangle = \frac{1}{2}\left(r^2 - \frac{J^2}{v^2}\right)\langle (\Delta v_\perp)^2\rangle. \quad (166)$$

The orbit-averaged diffusion coefficients are

$$D_{xy} = \langle\langle \Delta x\Delta y\rangle\rangle_\circlearrowright = 2\int_{r_p}^{r_a}\langle \Delta x\Delta y\rangle\frac{dr}{v_r}, \quad (167)$$

where v_r is the radial velocity. Assuming a spherical potential $\Phi(r)$, the energy E and angular momentum J are

$$E = \frac{1}{2}v_r^2 + \frac{1}{2}(J^2/r^2) + \Phi(r), \quad (168)$$

and

$$J = rv_t, \quad (169)$$

where v_t is the transverse velocity. Therefore,

$$v_r = \sqrt{2E - 2\Phi(r) - J^2/r^2}, \quad (170)$$

and

$$\frac{\partial(1/v_r)}{\partial J} = \frac{1}{v_r^2}\frac{1}{v_r}\frac{J}{r^2} = \frac{1}{(r^2 - J^2/v^2)}\frac{1}{v_r}\frac{J}{v^2}. \quad (171)$$

It then follows that

$$\frac{\partial}{\partial J}\langle JX\rangle_\circlearrowright = \left\langle\frac{\partial}{\partial J}JX\right\rangle_\circlearrowright + J^2\left\langle\frac{X}{(r^2v^2 - J^2)}\right\rangle_\circlearrowright, \quad (172)$$

and in particular

$$\frac{1}{2J}\frac{\partial}{\partial J}JD_{JJ} = \frac{1}{2J}\left\langle\left(\frac{1}{2}r^2 - \frac{J^2}{v^2}\right)\langle (\Delta v_\perp)^2\rangle\right\rangle_\circlearrowright = D_J, \quad (173)$$

which therefore proves that the J-only NR DCs indeed satisfy the F–D relation.

REFERENCES

Alexander, T. 2005, PhR, 419, 65
Alexander, T. 2006, JPhCS, 54, 243
Alexander, T. 2010, in Gravitational-wave and Electromagnetic Signatures of Massive Black Hole Binaries and Extreme Mass-ratio Inspirals, ed. P. Amaro-Seoane, & E. Porter (Paris: Laboratoire Astroparticule et Cosmologie) http://www.aei.mpg.de/~pau/conf_vid4/Alexander.pdf
Alexander, T. 2011, in ASP Conf. Ser. 439, The Galactic Center: a Window to the Nuclear Environment of Disk Galaxies, ed. M. R. Morris, Q. D. Wang, & F. Yuan (San Francisco, CA: ASP), 129
Alexander, T. 2012, in European Physical Journal Web of Conf. 39, Tidal Disruption and AGN Outburst, 5001
Alexander, T. 2015, in Gravitation 100 years after GR, Proc. 50th Rencontres de Moriond, ed. E. Auglé, J. Dumarchez, & J. T. T. Vân (Odense: ARISF), 353
Alexander, T., & Hopman, C. 2003, ApJL, 590, L29
Alexander, T., & Livio, M. 2001, ApJL, 560, L143
Alexander, T., & Morris, M. 2003, ApJL, 590, L25
Alexander, T., & Pfuhl, O. 2014, ApJ, 780, 148
Amaro-Seoane, P., Gair, J. R., Freitag, M., et al. 2007, CQGra, 24, 113
Amaro-Seoane, P., Sopuerta, C. F., & Freitag, M. D. 2013, MNRAS, 429, 3155
Antonini, F. 2014, ApJ, 794, 106
Antonini, F., & Merritt, D. 2013, ApJL, 763, L10
Bahcall, J. N., & Wolf, R. A. 1976, ApJ, 209, 214
Bar-Or, B., & Alexander, T. 2014, CQGra, 31, 244003
Bar-Or, B., Kupi, G., & Alexander, T. 2013, ApJ, 764, 52
Binney, J., & Tremaine, S. 2008, Galactic Dynamics (Princeton, NJ: Princeton Univ. Press)
Brem, P., Amaro-Seoane, P., & Sopuerta, C. F. 2014, MNRAS, 437, 1259
Chandrasekhar, S. 1944, ApJ, 99, 47
Cohn, H. 1979, ApJ, 234, 1036
Cohn, H., & Kulsrud, R. M. 1978, ApJ, 226, 1087
Eilon, E., Kupi, G., & Alexander, T. 2009, ApJ, 698, 641
Ferrarese, L., & Merritt, D. 2000, ApJL, 539, L9
Frank, J., & Rees, M. J. 1976, MNRAS, 176, 633
Freitag, M. 2001, CQGra, 18, 4033
Freitag, M. 2003, ApJL, 583, L21
Gair, J. R., Kennefick, D. J., & Larson, S. L. 2006, ApJ, 639, 999
Gebhardt, K., Bender, R., Bower, G., et al. 2000, ApJL, 539, L13
Ghez, A. M., Duchêne, G., Matthews, K., et al. 2003, ApJL, 586, L127
Gillessen, S., Eisenhauer, F., Trippe, S., et al. 2009a, ApJ, 692, 1075
Gillessen, S., Eisenhauer, F., Trippe, S., et al. 2009b, ApJ, 692, 1075
Goodman, J. 2003, MNRAS, 339, 937
Gürkan, M. A., & Hopman, C. 2007, MNRAS, 379, 1083
Hamers, A. S., Portegies Zwart, S. F., & Merritt, D. 2014, MNRAS, 443, 355
Heggie, D., & Hut, P. 2003, The Gravitational Million-Body Problem: A Multidisciplinary Approach to Star Cluster Dynamics (Cambridge: Cambridge Univ. Press)
Hills, J. G. 1988, Natur, 331, 687
Hils, D., & Bender, P. L. 1995, ApJL, 445, L7
Hopman, C., & Alexander, T. 2005, ApJ, 629, 362
Hopman, C., & Alexander, T. 2006a, ApJ, 645, 1152
Hopman, C., & Alexander, T. 2006b, ApJL, 645, L133
Ivanov, P. B. 2002, MNRAS, 336, 373
Kocsis, B., & Tremaine, S. 2015, MNRAS, 448, 3265
Lightman, A. P., & Shapiro, S. L. 1977, ApJ, 211, 244
Lynden-Bell, D. 1967, MNRAS, 136, 101
Lyne, A. G., Burgay, M., Kramer, M., et al. 2004, Sci, 303, 1153
Madigan, A.-M., Hopman, C., & Levin, Y. 2011, ApJ, 738, 99
Magorrian, J., & Tremaine, S. 1999, MNRAS, 309, 447
Merritt, D. 2015a, ApJ, 804, 52
Merritt, D. 2015b, ApJ, 804, 128
Merritt, D., Alexander, T., Mikkola, S., & Will, C. M. 2010, PhRvD, 81, 062002
Merritt, D., Alexander, T., Mikkola, S., & Will, C. M. 2011, PhRvD, 84, 044024
Miralda-Escudé, J., & Gould, A. 2000, ApJ, 545, 847
Nelson, R. W., & Tremaine, S. 1999, MNRAS, 306, 1
Ostriker, J. P. 1983, ApJ, 273, 99
Perets, H. B., Gualandris, A., Kupi, G., Merritt, D., & Alexander, T. 2009, ApJ, 702, 884
Perets, H. B., Hopman, C., & Alexander, T. 2007, ApJ, 656, 709
Peters, P. C. 1964, PhRv, 136, 1224
Peters, P. C., & Mathews, J. 1963, PhRv, 131, 435
Rauch, K. P., & Tremaine, S. 1996, NewA, 1, 149
Rees, M. J. 1988, Natur, 333, 523
Risken, H. 1989, The Fokker–Planck Equation. Methods of Solution and Applications (Berlin: Springer)
Shapiro, S. L., & Marchant, A. B. 1978, ApJ, 225, 603
Sigurdsson, S. 2003, CQGra, 20, 45
Sigurdsson, S., & Rees, M. J. 1997, MNRAS, 284, 318
Spitzer, L. 1962, Physics of Fully Ionized Gases (New York: Interscience)
Spitzer, L., Jr., & Shapiro, S. L. 1972, ApJ, 173, 529
Sridhar, S., Syer, D., & Touma, J. 1999, in ASP Conf. Ser. 160, Astrophysical Discs—an EC Summer School, ed. J. A. Sellwood, & J. Goodman (San Francisco, CA: ASP), 307
Syer, D., & Ulmer, A. 1999, MNRAS, 306, 35
Touma, J. R., & Sridhar, S. 2012, MNRAS, 423, 2083
Touma, J. R., Tremaine, S., & Kazandjian, M. V. 2009, MNRAS, 394, 1085
Will, C. M. 2006, LRR, 9, 3
Young, P. J., Shields, G. A., & Wheeler, J. C. 1977, ApJ, 212, 367