

# Self-forced inspirals with spin-orbit precession

Philip Lynch,<sup>1,2</sup> Maarten van de Meent,<sup>1,3</sup> and Niels Warburton<sup>2</sup>

<sup>1</sup>*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, Potsdam 14476, Germany*

<sup>2</sup>*School of Mathematics and Statistics, University College Dublin, Belfield, Dublin 4*

<sup>3</sup>*Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark*

(Dated: May 19, 2023)

We develop the first model for extreme mass-ratio inspirals (EMRIs) with misaligned angular momentum and primary spin, and zero eccentricity —also known as quasi-spherical inspirals— evolving under the influence of the first-order in mass ratio gravitational self-force. The forcing terms are provided by an efficient spectral interpolation of the first-order gravitational self-force in the outgoing radiation gauge. In order to speed up the calculation of the inspiral we apply a near-identity (averaging) transformation to eliminate all dependence of the orbital phases from the equations of motion while maintaining all secular effects of the first-order gravitational self-force at post-adiabatic order. The resulting solutions are defined with respect to ‘Mino time’ so we perform a second averaging transformation so the inspiral is parametrized in terms of Boyer-Lindquist time, which is more convenient of LISA data analysis. We also perform a similar analysis using the two-timescale expansion and find that using either approach yields self-forced inspirals that can be evolved to sub-radian accuracy in less than a second. The dominant contribution to the inspiral phase comes from the adiabatic contributions and so we further refine our self-force model using information from gravitational wave flux calculations. The significant dephasing we observe between the lower and higher accuracy models highlights the importance of accurately capturing adiabatic contributions to the phase evolution.

## I. INTRODUCTION

With the advent of spaced-based gravitational wave (GW) detectors such as Laser Interferometer Space Antenna (LISA) [1], TianQin [2], and Taiji [3] in the 2030s, comes the need to model sources in the millihertz frequency range. A prime source for these future detectors are extreme mass ratio inspirals (EMRIs). These systems consist of a massive black hole (MBH) primary with mass  $M$  and a stellar-mass compact object secondary with mass  $\mu$  (e.g., a stellar-mass black hole or neutron star). These space based detectors will be to be sensitive to EMRIs with mass ratios  $\epsilon = \mu/M \sim 10^{-4} - 10^{-7}$ . Gravitational waves from these systems are expected to stay in the LISA band for months to years with the secondary completing  $\sim 1/\epsilon = 10^4 - 10^7$  orbits spiralling around the primary before merging under the influence gravitational radiation reaction. Accurately modelling these systems through such a large number of orbital cycles to sub-radian accuracy with methods efficient enough for parameter estimation remains an open challenge, but one that will grant precise parameter estimation for MBHs along with some of the most sensitive probes for new physics beyond general relativity [4, 5].

In this work, we look at the subclass of these systems which have zero eccentricity and misaligned orbital angular momentum and primary spin. In the strong field regime this will cause the orbital plane to rapidly precess around the primary’s spin, ergodically filling (part of) a sphere, if we were to ignore orbital evolution due the emission of gravitational waves. Consequently, such systems are also known as quasi-spherical inspirals. One probable EMRI formation channel is the tidal disruption of a comparable mass black hole binary by a nearby MBH

which will result in a circular and arbitrarily inclined inspiral by the time it enters the LISA band [6–8]. While the primary EMRI formation channel of direct capture is expected to produce LISA band inspirals with eccentricity and inclination, radiation reaction will cause an EMRI to circularize (with the exception of right before the last stable orbit) [9, 10] and cause its inclination to slowly increase [11]. Thus an EMRI with any amount of inclination but low eccentricity is likely to evolve into a quasi-spherical inspiral before plunge.

The leading order behaviour of quasi-spherical EMRIs have been modelled by calculating the asymptotic fluxes at infinity and through the horizon of the primary and relating these to the averaged rate of change of energy, angular momentum, and Carter constant of the system [12–14]. This same approach has recently been generalized to inspirals with inclination and eccentricity [14]. While such inspiral calculations are numerically efficient and maybe accurate enough for detection of loud EMRI signals, the resulting adiabatic inspirals do not reach the sub-radian accuracy requirement needed to reach all of LISA’s science goals [15].

For that, one must understand the sub-leading order in mass ratio behaviour and compute post-adiabatic inspirals. This will require knowledge of the local force on the secondary due to the presence of its own gravitational field, i.e., its gravitational self-force (GSF). This is computed by expanding the metric of the binary as the Kerr metric of the primary plus perturbations which are of expressed as a power series of the small mass ratio  $\epsilon$  [16–18]. Using a two-timescale analysis, it has been determined that to reach sub-radian accuracy across the inspiral, one will need to know all first-order in the mass ratio effects of the GSF as well as the orbit averaged dissipative contri-

bution from the second-order in the mass ratio GSF [19]. While second-order results for Schwarzschild spacetime are now emerging [20–23], calculations for inspirals in Kerr spacetime are likely still a few year away. As such, we do not include them in this work but they can be incorporated into our computations once they are known. However, we do have a code that computes first-order GSF in Kerr spacetime for both eccentric equatorial orbits [24] and generic orbits [25] in the outgoing radiation gauge. We have adapted our code to compute the first ever GSF results for spherical orbits.

In order to drive inspirals, we require a model for the GSF can be rapidly evaluated for any point in the parameter space. This is typically done by calculating the GSF at many points and then fitting or interpolating the resulting data [26, 27]. We opt to interpolate our self-force data using Chebyshev polynomials, a technique which has proven to be very effective for EMRI calculations [28, 29]. We grid the parameter space on a  $18 \times 9$  Chebyshev nodes for prograde orbits where the primary spin is  $0.9M$  and for inclination angles up to  $45^\circ$  with respect to the equatorial plane, and interpolate the data using Chebyshev polynomials to sub-percent accuracy.

With an interpolated model in hand, we can then compute the inspiral trajectory using the method of osculating geodesics (OG) [30, 31]. We note that the OG equations for generic Kerr inspirals derived in either Ref. [31] or Ref. [28] become singular in the spherical case. This is not a physical singularity and in this work we derive OG equations in this limit that are finite (see Appendix A).

However, using this approach can take minutes to hours just to compute a single inspiral which is prohibitively slow for LISA data analysis. This is due to the OG equations’ dependence on the rapidly oscillating orbital phases, which forces a numerical solver to take small time steps to resolve all  $\sim 10^4 - 10^7$  orbital cycles of a typical EMRI. To overcome this, we employ the technique of near identity (averaging) transformations (NITs) to find equations of motion that average out the dependence on the orbital phases while faithfully capturing the long term behaviour of the binary. This framework was first sketched out for generic EMRI systems and applied to eccentric Schwarzschild inspirals in Ref. [32], hereafter paper I, and then extended to equatorial, eccentric Kerr inspirals in Ref. [28], hereafter paper II. In both cases, the resulting inspirals retained sub-radian accuracy when compared to the OG equations and could be computed in less than a second. However, in both of these works, the resulting inspirals were not parametrized in terms of Boyer-Lindquist coordinate time  $t$  (the time of an asymptotic observer), which is much more convenient for data analysis as this is directly related to the time at the detector. Overcoming this required an additional interpolation step which increases the overall computation time. Following a procedure presented in Ref. [18], one can parametrize the averaged equations of motion in terms of  $t$  by performing an additional averaging transformation. We apply this improved NIT to the case of

quasi-spherical Kerr inspirals and find that the inspiral calculation retains sub-radian accuracy and sub-second computation time, whilst being parametrized in a form that is more convenient for practical waveform production.

Much of the work done to model EMRIs to post-adiabatic order makes use of the two-timescale framework [18, 19, 33] which has been long known to operate in an almost equivalent way to NITs [34] when applied to the equations of motion. Inspired by Ref. [18], we perform a two-timescale expansion (TTE) on the NIT equations of motion and show that the TTE equations of motion also produces fast inspirals that are accurate to post-adiabatic order but with the drawback of doubling the number of equations to numerically solve. On the other hand these equations are independent of the mass ratio and so, after being solved once (in under a second), inspirals for any mass ratio can be generated in a fraction of a second. This could provide extra efficiency for an appropriately designed search or parameter estimation algorithm. The TTE equations of motion can also be used in conjunction with the two timescale expansion of the field equations to produce complete post-adiabatic waveforms [23].

Unlike the OG equations, both the NIT and TTE equations of motion cleanly separate the adiabatic and post-adiabatic contributions. As such, it is straightforward to increase the overall accuracy of the inspiral by calculating the adiabatic contributions from computationally cheaper, and generally more accurate, gravitational wave flux calculations and interpolating them to higher precision. Since this is the leading order contribution to the inspiral evolution, it will need to be known with a fractional accuracy of  $\sim \epsilon$  [27, 35]. We achieve this using a  $18 \times 19$  Chebyshev grid which covers all inclinations and reaches a fractional accuracy in the adiabatic pieces of  $\lesssim 10^{-5}$ . Using these higher precision interpolants for the adiabatic pieces along with the GSF model for the post-adiabatic pieces results in an improvement to the phase ranging from  $10 - 10^4$  radians, highlighting just how important high precision flux calculations are to both adiabatic and post-adiabatic inspirals [14, 36].

With our more accurate quasi-spherical inspiral model, we examine the effects of the self-force on quasi-spherical Kerr inspirals. We find that the evolution is consistent with what was known at adiabatic order and that the post-adiabatic contributions of the first-order GSF become more important as the inclination of the orbit increases.

This paper is organized as follows. In Section II, we introduce spherical geodesic orbits before giving an overview of the method of osculating geodesics. In Section III, we briefly review the gravitational self-force approach, discuss the modifications made to the Kerr GSF code of Ref. [25] for spherical orbits and outline our interpolation scheme for both the gravitational wave fluxes and the GSF. In Section IV, we review the technique of NITs for generic EMRI systems before describing an

additional averaging transformation used to obtain averaged equations of motion parametrized by  $t$ . We then perform a two timescale expansion on these averaged equations of motion before applying these frameworks to the case of quasi-spherical inspirals. In Section V, we outline our numerical implementation for calculating and using these averaged equations of motion. We demonstrate the results of this implementation in Section V where we start by comparing inspirals and waveforms calculated using either OG, NIT or TTE equations of motion. We then show the effect of using higher precision adiabatic pieces from flux calculations before finally exploring the effects of the self-force across a subsection of the spherical Kerr parameter space. We conclude with some discussion in Section VII. We detail how one can derive OG equations in the quasi-spherical limit in Appendix A and how the phases used in the  $t$  parametrized equations of motion relate to the waveform phases in Appendix B.

Throughout this paper we work in geometrical units where  $G = c = 1$ .

## II. QUASI-SPHERICAL INSPIRALS AROUND A ROTATING BLACK HOLE

In this section we describe the motion of a non-spinning compact object of mass  $\mu$  on a spherical orbit of Boyer-Lindquist radius  $r_p$ , with arbitrary inclination  $z_p = \cos \theta_p$  in the Kerr spacetime under the influence of some arbitrary force. We model the secondary as a point particle with a position given in modified Boyer-Lindquist coordinates by  $x_p^\alpha = \{t_p, r_p, z_p, \phi_p\}$ . Note that hereafter, we use a subscript ‘ $p$ ’ to denote a quantity evaluated at the particle’s location. Later in this work, we will take this perturbing force to be the self-force experienced by the secondary due to its interaction with its own metric perturbation. We denote the mass of the primary by  $M$  and parametrize its spin by  $a = |J|/M$  where  $J$  is the angular momentum of the black hole. The Kerr metric can then be written as

$$ds^2 = - \left(1 - \frac{2Mr}{\Sigma}\right) dt^2 + \frac{\Sigma}{\Delta} dr^2 + \frac{\Sigma}{1 - z^2} dz^2 - \frac{4Mar(1 - z^2)}{\Sigma} dt d\phi + \frac{1 - z^2}{\Sigma} (2a^2 r(1 - z^2) + \Sigma \varpi^2) d\phi^2 \quad (1)$$

where  $\Delta := r^2 + a^2 - 2Mr$ ,  $\Sigma := r^2 + a^2 z^2$ , and  $\varpi := \sqrt{r^2 + a^2}$ . If a force acts upon the secondary, its motion can be described by the forced geodesic equation

$$u^\beta \nabla_\beta u^\alpha = a^\alpha \quad (2)$$

where  $u^\alpha = dx^\alpha/d\tau$  is the secondary’s four-velocity,  $\nabla_\beta$  is the covariant derivative with respect to the Kerr metric, and  $a^\alpha$  is the secondary’s four-acceleration. We seek to recast Eq. (2) into a form useful for applying the near-identity transformations. Before considering the forced equation it is useful to first examine the geodesic limit.

### A. Geodesic motion and orbital parametrization

In the absence of any perturbing force, the secondary will follow a geodesic, i.e.,

$$u^\beta \nabla_\beta u^\alpha = 0, \quad (3)$$

The symmetries of the Kerr spacetime allow for the identification of integrals of motion given by

$$\mathcal{E} = -u_t, \quad \mathcal{L}_z = u_\phi, \quad (4a-b)$$

$$\mathcal{Q} = u_z^2 + z_p^2 \left( \mathcal{A} + \frac{\mathcal{L}_z^2}{1 - z_p^2} - u_z^2 \right), \quad (4c)$$

where  $\mathcal{E}$  is the orbital energy per unit rest mass  $\mu$ ,  $\mathcal{L}_z$  is the z-component of the angular momentum divided by  $\mu$ ,  $\mathcal{Q}$  is the Carter constant divided by  $\mu^2$  [37], and  $\mathcal{A} = a^2(1 - \mathcal{E}^2)$ . This definition of the Carter constant is related to another common definition of the Carter Constant,  $\mathcal{K}$ , by

$$\mathcal{K} = \mathcal{Q} + (\mathcal{L}_z - a\mathcal{E})^2. \quad (5)$$

The geodesic equation can be written explicitly in terms of these integrals of motion [37, 38]:

$$\left(\frac{dr_p}{d\lambda}\right)^2 := V_r = \mathcal{B}_p^2 - \Delta_p (r_p^2 + \mathcal{K}) = 0 \quad (6a)$$

$$\begin{aligned} \left(\frac{dz_p}{d\lambda}\right)^2 &:= V_z = \mathcal{Q} - z_p^2 (\mathcal{A}(1 - z_p^2) + \mathcal{L}_z^2 + \mathcal{Q}) \\ &= (z_p^2 - z_-^2) (\mathcal{A}z_p^2 - z_+^2) \end{aligned} \quad (6b)$$

$$\frac{dt_p}{d\lambda} := f_t^{(0)} = \frac{\varpi_p^2}{\Delta_p} \mathcal{B}_p - a^2 \mathcal{E}(1 - z_p^2) + a\mathcal{L}_z \quad (6c)$$

$$\frac{d\phi_p}{d\lambda} := f_\phi^{(0)} = \frac{a}{\Delta_p} \mathcal{B}_p + \frac{\mathcal{L}_z}{1 - z_p^2} - a\mathcal{E}. \quad (6d)$$

where  $\mathcal{B} = \mathcal{E}\varpi^2 - a\mathcal{L}_z$ , the radial potential  $V_r = 0$  for circular orbits, and  $z_+ > z_-$  are the roots of the polar potential  $V_z$ . We make use of (Carter-)Mino time  $\lambda$  which is related to the secondary’s proper time  $\tau$ , by  $d\tau/d\lambda = \Sigma_p$ , as this decouples the radial and polar equations [77]. While this is not strictly necessary in the spherical case as there is no radial motion, using  $\lambda$  as our time parameter allows us to take advantage of the analytic solutions for generic Kerr geodesics derived in Ref. [39], and allows for a more uniform treatment with paper II and the generic case in the future.

Rather than parametrize an orbit using the integrals of motion  $\vec{\mathcal{J}} = \{\mathcal{E}, \mathcal{L}_z, \mathcal{K}\}$ , we find it more convenient to work with the constants  $\vec{P} = \{r_p, e = 0, x\}$ . Here  $x$  is a measure of the orbital inclination given by  $x = \cos \theta_{\text{inc}}$ . The inclination angle  $\theta_{\text{inc}}$  is related to  $\theta_{\text{min}}$  (the minimum value of  $\theta_p$  measured with respect to the primary’s spin axis) by  $\theta_{\text{inc}} = \pi/2 - \text{sgn}(\mathcal{L}_z)\theta_{\text{min}}$ . Explicit relationships between the integrals of motion  $\vec{\mathcal{J}}$  in terms of  $\vec{P}$  for spherical orbits are too lengthy to be expressed here,

but were first derived in Ref. [12] and can be found in the `KerrGeodesics` package as part of the Black Hole Perturbation Toolkit [40].

There are other common choices for inclination in the literature such as the inclination angle  $\iota$  [11–13, 41] given by  $\cos \iota = \mathcal{L}_z / \sqrt{\mathcal{L}_z^2 + \mathcal{Q}}$ . Both inclination angles smoothly parametrize all inclinations between prograde equatorial motion where  $x = 1 = \cos \iota$  to retrograde equatorial motion where  $x = -1 = \cos \iota$ . However, we opt to use  $x$  as it has a simple relation to the roots of the polar potential  $V_z$  via:

$$z_- = \sqrt{1 - x^2} \quad (7)$$

$$z_+ = \sqrt{\mathcal{A} + \mathcal{L}_z^2/x^2}. \quad (8)$$

It is worth noting that not all values of  $\{r_p, x\}$  correspond to bound geodesics and we denote the position of the innermost stable spherical orbit (ISSO) as  $r_{\text{ISSO}}(a, x)$  [9, 42].

In order to later apply the near-identity (averaging) transformations, it will be useful to employ action-angle formulation to parametrize the geodesic motion [39, 43, 44]. In this description the orbital phase  $\vec{q} = \{q_z\}$  is such that the geodesic equations can be written in the form

$$\frac{dP_j}{d\lambda} = 0 \quad \text{and} \quad \frac{dq_z}{d\lambda} = \Upsilon_z^{(0)}(\vec{P}), \quad (9a-b)$$

where  $\Upsilon_z$  is the Mino time fundamental polar frequency. This is known analytically [39] and is given by

$$\Upsilon_z^{(0)} = \frac{\pi z_+}{2K(k_z)} \quad (10)$$

where  $k_z = \mathcal{A}z_-^2/z_+^2$ ,  $K$  is complete elliptic integral of the first kind with  $K(m) = F(\pi/2|m)$  and  $F$  is the incomplete elliptic integral of the first kind given by

$$F(\phi|m) = \int_0^\phi \frac{d\theta}{\sqrt{1 - m \sin^2 \theta}}. \quad (11)$$

One can also analytically express the polar coordinate  $z$  in terms of  $r, x$  and  $q_z$  via

$$z_p(q_z) = z_- \text{sn} \left( K(k_z) \frac{2(q_z + \frac{\pi}{2})}{\pi} \middle| k_z \right), \quad (12)$$

where  $\text{sn}$  is Jacobi elliptic sine function given by  $\text{sn}(u|m) = \sin(\text{am}(u|m))$  and the Jacobi amplitude  $\text{am}$  is the inverse function of  $F$ .

At this point we note that it is also common in the literature [45, 46] to express this polar coordinate in terms of a quasi-Keplerian angle  $\chi_z$ , via

$$z_p(\chi_z) = z_- \cos(\chi_z). \quad (13)$$

However, the rate of change of  $\chi_z$  depends on  $\chi_z$  itself, which will be inconvenient in later sections as we look to derive averaged equations of motion that have no dependence on the orbital phases.

## B. Osculating Geodesics for quasi-spherical inspirals

To go beyond the geodesic orbits and describe quasi-spherical inspirals, we make use of the method of osculating orbital elements (or osculating geodesics) [30, 31] to recast the equations of motion of a body under the influence of a force obeying Eq. (2).

We chose a set of parameters that uniquely specify a geodesic orbit, such as the integrals of motion  $\vec{P}$  along with the initial values of the orbital phases of the geodesic orbit  $\vec{q}_0$ , and designate them as a set of ‘‘orbital elements’’  $\vec{I} = \{\vec{P}, \vec{q}_0\}$ .

For accelerated orbits, these orbital elements are promoted from constants to functions of time. We note that now the orbital elements  $\vec{q}_0(\lambda)$  are different quantities from the values of  $\vec{q}$  evaluated at  $\lambda = 0$ , i.e.,  $\vec{q}(0)$ . To produce waveforms, we will also require the evolution of certain ‘‘extrinsic quantities’’  $\vec{S} = \{t_p, \phi_p\}$  which are the secondary’s Boyer-Lindquist time and azimuthal coordinates respectively.

We assume that the worldline and four-velocity of the secondary at each instant can be described as the worldline and four-velocity of a test body on a tangent geodesic, i.e.,  $x^\alpha(\lambda) = x_G^\alpha(I^A(\lambda), \lambda)$  and  $u^\alpha(\lambda) = u_G^\alpha(I^A(\lambda), \lambda)$ . Using these assumptions, one can derive the osculating geodesic equations [30, 31]:

$$\frac{\partial x_G^\alpha}{\partial I^A} \frac{dI^A}{d\lambda} = 0 \quad \text{and} \quad \frac{\partial u_G^\alpha}{\partial I^A} \frac{dI^A}{d\lambda} = \Sigma a_\alpha. \quad (14a-b)$$

From these, one can derive explicit evolution equations for each of the orbital elements that take the following form:

$$\frac{dP_j}{d\lambda} = F_j(\vec{P}, \vec{q}), \quad (15a)$$

$$\frac{dq_i}{d\lambda} = \Upsilon_i^{(0)}(\vec{P}) + f_i(\vec{P}, \vec{q}), \quad (15b)$$

$$\frac{dS_k}{d\lambda} = f_k^{(0)}(\vec{P}, \vec{q}) \quad (15c)$$

where  $f_i = dq_{0,i}/d\lambda$ .

For generic Kerr inspirals, this was implemented in Ref. [31] using quasi-Keplerian angles (e.g.,  $\chi$ ) as the orbital phases and the integrals of motion  $\vec{P} = \vec{\mathcal{J}}$ . In order to derive averaged equations of motion, we implemented an alternative version in Paper II [28] which makes use of geodesic action angles  $\vec{q}$  for the orbital phases and the integrals of motion  $\vec{P} = \{p, e, x\}$ . However, the evolution equations for the semilatus rectum  $p$  and eccentricity  $e$  become numerically singular in the quasi-spherical limit where  $p \rightarrow r_p/M$  and  $e \rightarrow 0$ . These are not physical singularities, as one can derive alternative expressions which are regular in this limit by considering the evolution of the integrals of motion  $\vec{\mathcal{J}}$  and then relating this to the evolution of  $r_p$  (and thus  $p$ ). We have done this explicitly in Appendix A. As for eccentricity, it has been well established that under gravitational radiation reaction we

would expect quasi-spherical inspirals with  $e = 0$  to remain quasi-spherical and so  $\dot{e} = F_e = 0$  [10, 11, 47].

Combing these modified evolution equations with the rest of the evolution equations derived in Paper II [28] gives us equations of motion for quasi-spherical inspirals under the influence of the gravitational self-force with the form

$$\frac{dr_p}{d\lambda} = \epsilon F_r^{(1)}(r_p, x, q_z) + \epsilon^2 F_r^{(2)}(r_p, x, q_z) + \mathcal{O}(\epsilon^3), \quad (16a)$$

$$\frac{dx}{d\lambda} = \epsilon F_x^{(1)}(r_p, x, q_z) + \epsilon^2 F_x^{(2)}(r_p, x, q_z) + \mathcal{O}(\epsilon^3), \quad (16b)$$

$$\frac{dq_z}{d\lambda} = \Upsilon_z^{(0)}(r_p, x) + \epsilon f_z^{(1)}(r_p, x, q_z) + \mathcal{O}(\epsilon^2), \quad (16c)$$

$$\frac{dt}{d\lambda} = f_t^{(0)}(r_p, x, q_z), \quad (16d)$$

$$\frac{d\phi}{d\lambda} = f_\phi^{(0)}(r_p, x, q_z). \quad (16e)$$

### III. GRAVITATIONAL SELF-FORCE FOR QUASI-SPHERICAL KERR INSPIRALS

The gravitational self-force approach consists of expanding the metric of the binary around the metric of the primary, i.e.,  $g_{\mu\nu} = \bar{g}_{\mu\nu} + \epsilon h_{\mu\nu}^{(1)} + \epsilon^2 h_{\mu\nu}^{(2)} + \dots$  where  $\bar{g}_{\mu\nu}$  is the Kerr metric and the  $h_{\mu\nu}^{(n)}$  are the  $n$ -th order perturbations to the spacetime due to the presence of the secondary. Using matched asymptotic expansions, one can then derive how the interaction between these metric perturbations their source effects the motion of the secondary [17].

#### A. Gravitational self-force for spherical orbits

This work represents the first calculation of the gravitational self-force for spherical orbits in Kerr spacetime. While the gravitational self-force had been previously calculated for generic (eccentric and inclined) orbits in Kerr in [25], building on previous developments in [24, 48–50], and the scalar self-force for spherical orbits had been calculated in [51], the case of the gravitational self-force for spherical orbits has not been touched before now.

To calculate the gravitational self-force for spherical orbits, we follow the same method as was used in the generic orbit case in [25], which can be adapted without major adjustments. We summarize our method here. We start by solving the Teukolsky equation for the Weyl scalar  $\psi_4$  in the frequency domain using a numerical implementation [52, 53] of the semi-analytical Mano-Suzuki-Takasugi method [54, 55]. From  $\psi_4$ , we obtain the Hertz potential by algebraically inverting the fourth order differential equation relating it to  $\psi_4$  [50, 56] mode-by-mode. From the modes of the Hertz potential we can reconstruct the modes of the local metric perturbation in the outgoing radiation gauge [57–59],

from which in turn we obtain the modes of the gravitational self-force. Initially these modes are expressed in spin-weighted spheroidal harmonics and their derivatives. To facilitate calculating the regular part of the GSF modes using mode-sum regularization [60–63], the spheroidal harmonics are first projected to spin-weighted spherical harmonics using the method introduced in [12], which in turn are projected onto scalar spherical harmonics [24, 25, 50]. The mode-sum is accelerated by a tail fitting procedure leading to an exponential convergence of the mode-sum [50].

The above metric reconstruction procedure captures most of the metric perturbation, except for a perturbation of the background Kerr spacetime’s mass and spin, and a purely gauge contribution. The mass and spin completions are directly related to the energy and angular momentum of the orbit [64, 65]. The gauge completion is more involved. The version of the outgoing radiation gauge used for the GSF calculation, the no-string gauge [49], has a distributional singularity on sphere of the orbit. Recently, [66] has proposed a generalization of our method that produces the metric perturbation in a gauge that is sufficiently smooth to allow it to be used as a source for a second order calculation. We here strive for a less lofty goal and will add the minimal gauge completion needed to ensure orbit averaged observables like the frequency-shifts are invariant. This involves fixing a 2-parameter gauge freedom representing a rescaling of time and time dependent coordinate rotations inside the orbit. In [66] a nice derivation of the needed coefficients was given. However, for this work we used an implementation [67] of a generalization of the method used for the mass and spin completion in Ref. [64]. The gauge vector representing this freedom is given by,

$$\xi^\mu = \{\alpha t, 0, 0, \beta t\}, \quad (17)$$

with the gauge fixing parameters  $\alpha$  and  $\beta$  given by

$$\alpha = \frac{1}{2\pi\Upsilon_t} \int_0^{2\pi} \frac{\mathcal{I}_\alpha(z_p)}{3\Delta_p^2 \Sigma_p^2 (r_p^2 + a^2)(1 - z_p^2)^2} dq_z, \quad \text{and} \quad (18)$$

$$\beta = \frac{1}{2\pi\Upsilon_t} \int_0^{2\pi} \frac{\mathcal{I}_\beta(z_p)}{3ar_p \Delta_p^2 \Sigma_p^2 (r_p^2 + a^2)(1 - z_p^2)^2} dq_z, \quad (19)$$

where the numerators of the integrands,  $\mathcal{I}_\alpha$  and  $\mathcal{I}_\beta$  are given in Appendix C, and  $\Delta_p$  and  $\Sigma_p$  denote  $\Delta$  and  $\Sigma$  evaluated at the particle location.

It is useful to split the self-force into dissipative (time anti-symmetric) and conservative (time-symmetric) contributions [63]. The dissipative pieces cause the orbit to shrink until the secondary plunges into the primary. They also have a small effect on the orbital inclination, causing the orbit to become more inclined over time [11–13, 41]. To produce adiabatic waveforms, we only require knowledge of the orbit averaged dissipative pieces of the first-order self-force. These can be related, via balance laws, to the fluxes of GWs to infinity and down the event horizon. Since calculating fluxes avoids regularization of

the metric perturbation, adiabatic inspirals are typically calculated via flux balance laws [9, 14, 36, 68, 69]. The conservative pieces have more subtle effects on the inspiral, such as altering the rate of periastris advance, the rate of nodal precession, and the location of the innermost stable circular orbit [70–76].

Computing post-adiabatic inspirals requires knowledge of both the dissipative and conservative pieces of the first-order self-force and the orbit average piece of the second-order self-force [19]. There are as yet no calculations of the latter in Kerr spacetime, so we will make do with only the first-order gravitational wave fluxes and the first-order self-force in this work, though this means that the resulting “post-adiabatic” inspirals and waveforms will be gauge dependent until the missing second-order contributions are added [28].

In order to drive the OG equations of motion to calculate the inspiral, we require a model for the GSF that can be rapidly evaluated at each time step during the inspiral. This is typically done by tiling the parameter space with GSF data which is then either fitted to a model or interpolated. While this has been done in a variety of ways for eccentric orbits in Schwarzschild [26, 27] and Kerr [28], we will now describe the first interpolated GSF model for quasi-spherical Kerr inspirals.

### B. Interpolated gravitational wave fluxes for quasi-spherical Kerr inspirals

To introduce our interpolation procedure, we will first interpolate the energy and angular momentum fluxes for spherical orbits. Since flux calculations, which can be obtained directly from  $\psi_4$ , are significantly cheaper than calculating the GSF, it is much more feasible to densely tile a large section of the parameter space with flux data. This in turn will result in more accurate interpolation of the leading-order, adiabatic effects. It will also allow us to carry out consistency checks on our GSF model by comparing the fluxes with the orbit averaged GSF.

We start by fixing the value of the spin parameter of the primary, which we choose to be  $a = 0.9M$ . This reduces our parameter space to two parameters; the orbital radius  $r_p$  and the inclination  $x$ . We then define a parameter  $y$  using  $r_p$  and the position of the innermost stable spherical orbit  $r_{\text{ISSO}}$ . We chose  $y$  to be

$$y = \sqrt{\frac{r_{\text{ISSO}}(a, x)}{r_p}}. \quad (20)$$

Tiling the parameter space with  $y$  instead of  $r_p$  will concentrate more points near the ISSO where the fluxes and the GSF experience the most variation. We let  $y$  range from 0 to 1 and  $x$  range from  $-1$  to  $1$  thus covering all inclinations for both prograde and retrograde orbits.

This parametrization is convenient when using Chebyshev polynomials of the first kind, where the order  $n$  polynomial is defined by  $T_n(\cos \vartheta) := \cos(n\vartheta)$ . The Cheby-

shev nodes are the roots these polynomials, and the location of the  $k$ th root of  $n$ th polynomial is given by

$$N(n, k) = \cos\left(\frac{2k-1}{2n}\pi\right) \quad (21)$$

We then calculate the fluxes on a  $18 \times 19$  grid of Chebyshev nodes, with the  $y$  values given by the roots of the 18th order Chebyshev polynomial and the  $x$  values given by the roots of the 19th order Chebyshev polynomial. The size of this grid was picked empirically until the interpolant reached the desired accuracy across the parameter space.

At each point on this grid, we calculate the energy and angular momentum flux both at infinity and at the event horizon. From these, one can calculate the leading order orbit averaged rate of change of the energy and angular momentum of the secondary via the following balance laws [77–81]:

$$\left\langle \frac{d\mathcal{E}}{dt} \right\rangle = -\epsilon(\mathcal{F}_{\mathcal{E}}^\infty + \mathcal{F}_{\mathcal{E}}^{\mathcal{H}}) + \mathcal{O}(\epsilon^2), \quad (22a)$$

$$\left\langle \frac{d\mathcal{L}_z}{dt} \right\rangle = -\epsilon(\mathcal{F}_{\mathcal{L}_z}^\infty + \mathcal{F}_{\mathcal{L}_z}^{\mathcal{H}}) + \mathcal{O}(\epsilon^2), \quad (22b)$$

where the orbit average of a quantity  $A$  is given by

$$\langle A \rangle(\vec{P}) = \frac{1}{2\pi} \int_0^{2\pi} A(\vec{P}, q_z) dq_z. \quad (23)$$

We could interpolate the rates of change of energy and angular momentum, but we find it more convenient to work with the orbital elements  $r_p$  and  $x$ . As such, we find their rates of change via the chain rule:

$$\left\langle \frac{dr_p}{dt} \right\rangle = \frac{\partial r_p}{\partial \mathcal{E}} \left\langle \frac{d\mathcal{E}}{dt} \right\rangle + \frac{\partial r_p}{\partial \mathcal{L}_z} \left\langle \frac{d\mathcal{L}_z}{dt} \right\rangle + \mathcal{O}(\epsilon^2) \quad (24a)$$

$$= \epsilon \Gamma_r^{(1)} + \mathcal{O}(\epsilon^2)$$

$$\left\langle \frac{dx}{dt} \right\rangle = \frac{\partial x}{\partial \mathcal{E}} \left\langle \frac{d\mathcal{E}}{dt} \right\rangle + \frac{\partial x}{\partial \mathcal{L}_z} \left\langle \frac{d\mathcal{L}_z}{dt} \right\rangle + \mathcal{O}(\epsilon^2) \quad (24b)$$

$$= \epsilon \Gamma_x^{(1)} + \mathcal{O}(\epsilon^2).$$

The partial derivatives can be found using the analytic expressions for  $\mathcal{E}(r_p, x)$  and  $\mathcal{L}_z(r_p, x)$  from Ref. [12] to construct the Jacobian

$$J = \begin{bmatrix} \frac{d\mathcal{E}}{dr_p} & \frac{d\mathcal{E}}{dx} \\ \frac{d\mathcal{L}_z}{dr_p} & \frac{d\mathcal{L}_z}{dx} \end{bmatrix}. \quad (25)$$

This can then be inverted to give

$$J^{-1} = \begin{bmatrix} \frac{dr_p}{d\mathcal{E}} & \frac{dr_p}{d\mathcal{L}_z} \\ \frac{dx}{d\mathcal{E}} & \frac{dx}{d\mathcal{L}_z} \end{bmatrix} = \frac{1}{\det J} \begin{bmatrix} \frac{d\mathcal{L}_z}{dr_p} & -\frac{d\mathcal{E}}{dx} \\ -\frac{d\mathcal{L}_z}{dr_p} & \frac{d\mathcal{E}}{dx} \end{bmatrix}. \quad (26)$$

To improve the accuracy of our interpolation, we rescale the data for  $\Gamma_r^{(1)}$  and  $\Gamma_x^{(1)}$  by a factor of  $r_p^3(1-y)$  and  $r_p^{11/2}(1-x^2)$  respectively. This scaling comes from

the leading order PN terms for  $\Gamma_r^{(1)}$  and  $\Gamma_x^{(1)}$ , times a term that is zero for the limiting cases of either the separatrix or the equatorial plane respectively. Finally, we take a discrete cosine transformation of the data on our Chebyshev grid to obtain the Chebyshev polynomial coefficients  $C_{r_p/x}^{ij}$ . Summing these coefficients together with Chebyshev polynomials gives us the following interpolants for  $\Gamma_r^{(1)}$  and  $\Gamma_x^{(1)}$ :

$$\Gamma_r^{(1)} = \frac{1}{r_p^3(1-y)} \sum_{i=0}^{17} \sum_{j=0}^{18} C_p^{ij} T_i(2y-1) T_j(x), \quad (27a)$$

$$\Gamma_x^{(1)} = \frac{1}{r_p^{11/2}(1-x^2)} \sum_{i=0}^{17} \sum_{j=0}^{18} C_x^{ij} T_i(2y-1) T_j(x). \quad (27b)$$

Using the largest coefficient for  $i = 17$  and  $j = 18$  to estimate the relative error, we infer that these interpolants should have a relative error of  $\sim 10^{-6}$ .

To test the accuracy of this interpolation, we compare the interpolants to data on a grid which was not used for the interpolation while using the relations  $\mathcal{F}_\mathcal{E}^\infty + \mathcal{F}_\mathcal{E}^\mathcal{H} = -(\partial\mathcal{E}/\partial p)\Gamma_p^{(1)} - (\partial\mathcal{E}/\partial x)\Gamma_x^{(1)}$  and  $\mathcal{F}_{\mathcal{L}_z}^\infty + \mathcal{F}_{\mathcal{L}_z}^\mathcal{H} = -(\partial\mathcal{L}_z/\partial p)\Gamma_p^{(1)} - (\partial\mathcal{L}_z/\partial x)\Gamma_x^{(1)}$ . We see that the interpolants match the energy and angular momentum fluxes have a relative error of  $\lesssim 10^{-5}$  with the exception of points very close to the ISSO. In principle this model interpolates out to  $r_p = \infty$  and the leading PN scaling should give it the right behaviour in the weak field. We would expect the model to retain the comparable level of accuracy at large  $r_p$ , but it has not been tested beyond  $r_p \sim 30$  as in this work we are more interested in the strong field dynamics.

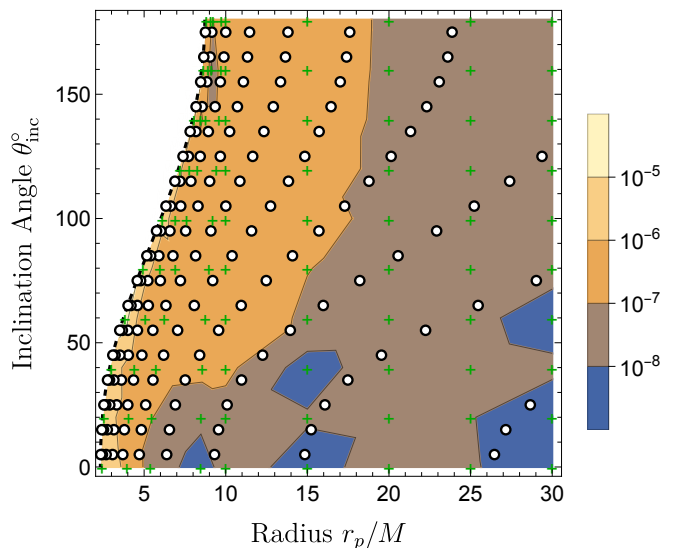
### C. Interpolated gravitational self-force model for quasi-spherical Kerr inspirals

We now use a similar interpolation scheme to create a model for the gravitational self-force that is continuous throughout the parameter space and fast to evaluate which can be used with the osculating geodesic equations to describe quasi-spherical Kerr inspirals.

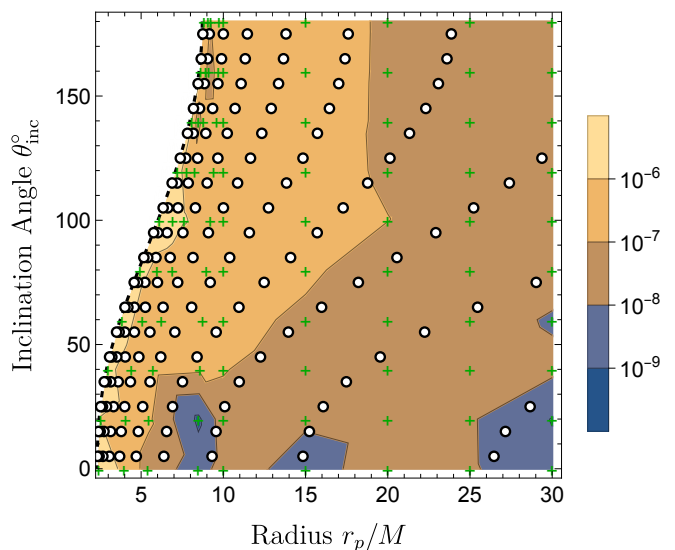
However, given the cost of our GSF code—the computation of the GSF on a single geodesic can take hundreds of CPU hours—we restrict ourselves to a 2D slice of the EMRI parameter space using only a few hundred points. Once again, we restrict  $a = 0.9M$  and let  $y$  range from  $y_{\min} = 0.1$  to  $y_{\max} = 1$ , but instead of  $x$ , we opt to tile in  $z_-^2 = 1 - x^2$  and let  $z_{-, \min}^2 = 0$  to  $z_{-, \max}^2 = 0.5$ . This allows us to cover inclination angles ranging from  $0^\circ \leq \theta_{\text{inc}} \leq 45^\circ$ , in a way that is easy to rescale for Chebyshev interpolation.

We define parameters  $u$  and  $v$  which cover this parameter space as they range from  $(-1, 1)$

$$u := \frac{y - (y_{\min} + y_{\max})/2}{(y_{\min} - y_{\max})/2} \text{ and,} \quad (28a)$$



(a) The relative error of the energy flux.



(b) The relative error of the angular momentum flux.

FIG. 1. The relative error in the rate of energy (a) and angular momentum (b) loss between the interpolated fluxes and the fluxes calculated with a grid of verification points that were not used for the interpolation. The white dots indicate the locations of the data points that were interpolated to produce the model and the green crosses indicate the positions of verification data. The relative error is always  $< 10^{-5}$  with the exception of orbits that are very close to the ISSO.

$$v := \frac{z_-^2 - (z_{-, \min}^2 + z_{-, \max}^2)/2}{(z_{-, \min}^2 - z_{-, \max}^2)/2}. \quad (28b)$$

We then calculate the GSF on a  $18 \times 9$  grid of Chebyshev nodes, with the  $u$  values given by the roots of the 18th order polynomial and the  $v$  values given by the roots of the 9th order polynomial. At each point on our grid, we Fourier decompose each component of the force with respect to the polar action angle  $q_z$ . As inclination in-

creases more and more Fourier modes of the GSF become relevant. However, the number of relevant modes stays finite even for polar orbits. To build a Chebyshev interpolant we need to resolve all Fourier modes at all grid points. Resolving the higher order harmonics at low inclination grid points is numerically challenging and therefore computationally expensive. This is the main reason for limiting the range of inclinations to  $0^\circ \leq \theta_{\text{inc}} \leq 45^\circ$  as this limits the number of Fourier modes that need to be resolved.

To smooth the behaviour of the force near the separatrix and improves the accuracy of our interpolation, we then multiply the data for each Fourier coefficient by a factor of  $(1-y)^2$ . Next, we use Chebyshev polynomials to interpolate each Fourier coefficient across the  $(u, v)$  grid. We then sum the modes to reconstruct our interpolated gravitational self-force model:

$$a_\mu = \sum_{\kappa=0}^{24} \frac{A_\mu^\kappa(y, z_-^2) \cos(\kappa q_z) + B_\mu^\kappa(y, z_-^2) \sin(\kappa q_z)}{(1-y)^2}, \quad (29a)$$

$$A_\mu^\kappa(y, e) = \sum_{i=0}^{17} \sum_{j=0}^8 A_\mu^{\kappa ij} T_i(u) T_j(v), \quad (29b)$$

$$B_\mu^\kappa(y, e) = \sum_{i=0}^{17} \sum_{j=0}^8 B_\mu^{\kappa ij} T_i(u) T_j(v). \quad (29c)$$

We note that this choice of rescaling forces each component to become singular at the ISSO, and while the components of the GSF change rapidly as one approaches the ISSO, we still expect them to be finite at the ISSO. A greater understanding of the analytic structure of the GSF in this region would greatly improve this and any future interpolated GSF models.

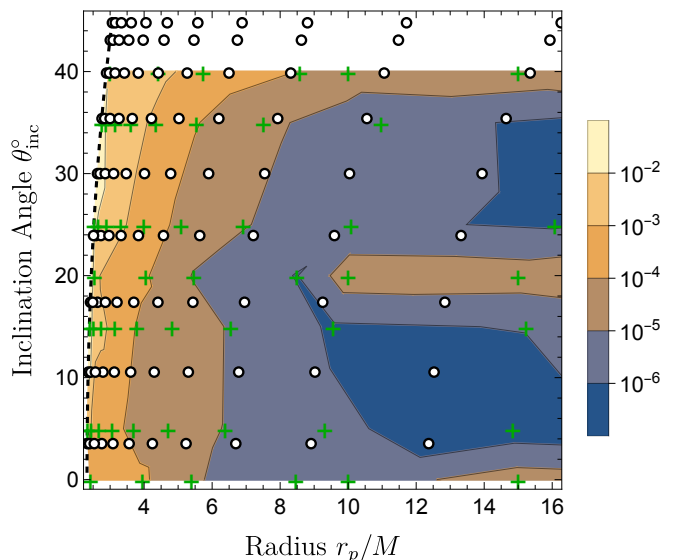
We also note that the GSF should satisfy the orthogonality condition with the geodesic four-velocity, i.e.,  $a_\mu u^\mu = 0$ . Interpolation will bring with it a certain amount of error which can cause this condition to be violated. Since the OG equations are derived assuming this condition to be true [28, 31], we project the force so that this condition is always satisfied, i.e.,

$$a_\mu^\perp = a_\mu + a_\nu u^\nu u_\nu. \quad (30)$$

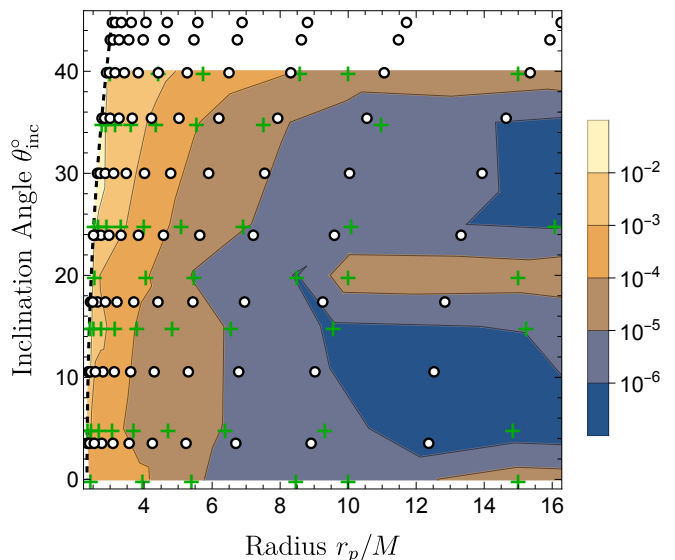
To verify the accuracy of our interpolated model, we employ the flux balance laws to compare the local energy and angular momentum lost by the secondary against the energy and angular momentum fluxes radiated at infinity and down the horizon:

$$\frac{\Upsilon_\tau^{(0)}}{\Upsilon_t^{(0)}} \langle a_t^\perp \rangle = (\mathcal{F}_{\mathcal{E}, \infty} + \mathcal{F}_{\mathcal{E}, H}), \quad (31a)$$

$$\frac{\Upsilon_\tau^{(0)}}{\Upsilon_t^{(0)}} \langle a_\phi^\perp \rangle = -(\mathcal{F}_{\mathcal{L}_z, \infty} + \mathcal{F}_{\mathcal{L}_z, H}), \quad (31b)$$



(a) The relative error for rate of change of energy.



(b) The relative error for rate of change of angular momentum.

FIG. 2. The relative error in the rate of energy (a) and angular momentum (b) loss between the asymptotic fluxes and the interpolated GSF model. The white dots indicate the locations of the data points the model is interpolating. The green crosses indicate the locations of the flux calculations and thus where the comparisons are made. The relative error only rises to  $\sim 10^{-2}$  when close to the ISSO, and is otherwise typically of the order  $\sim 10^{-3} - 10^{-5}$ .

where  $\Upsilon_t^{(0)} = \langle f_t^{(0)} \rangle$  and  $\Upsilon_\tau^{(0)} = \langle \Sigma_p \rangle$  are the fundamental Mino time frequencies for the secondary's Boyer-Lindquist time coordinate and proper time with analytic expressions found in Ref. [39] and appendix C of Ref. [44] respectively.

As we can see from Figure 2, the relative error in the fluxes across the parameter space. The white circles indicate the points used to generate the interpolated GSF



model, whereas the green crosses indicate the points in the parameter space where the fluxes were calculated and thus where the comparisons were made. We find that the relative error in the fluxes only rises to  $\sim 10^{-2}$  when close to the ISSO, and is typically of the order  $\sim 10^{-3} - 10^{-5}$ . This is comparable to previous methods which required tens of thousands of points to achieve the same level of accuracy [26, 27]. While this would not be sufficiently accurate for LISA data analysis which would require fluxes accurate to  $\sim \epsilon$ , such an interpolating error may be permissible for the post-adiabatic contributions of the first order GSF [27, 35]. We will go on to improve this model by incorporating the higher precision flux interpolants in Sec. VIC.

We can now use this model in conjunction with the OG equations of motion (16) to calculate inspiral trajectories. However, we find these trajectories take minutes to hours to compute, due to the need to resolve hundreds of thousands of orbital cycles. We will now look to leverage averaging transformations which can remove the dependence of the orbital phases from the equations of motion while retaining the accuracy required to produce post-adiabatic EMRI waveforms.

#### IV. AVERAGING TECHNIQUES FOR EMRI EQUATIONS OF MOTION

Near-identity (averaging) transformations (NITs) are a well known technique in applied mathematics and celestial mechanics [34]. This approach involves making small transformations to the equations of motion, such that the short timescale physics is averaged over, while retaining information about the long term evolution of a system. In paper I [32], these transformations were derived for a generic EMRI system which we review in Section IV A below.

As we have seen above, for Kerr geodesics it is helpful to parametrize the motion by Mino time,  $\lambda$ . However, for data analysis one requires the final waveforms to be parametrized by the time at the detector. Naively, transforming back to Boyer-Lindquist time,  $t$ , requires either numerically interpolating  $\lambda(t)$ , or adding orbital timescale oscillations back into our NIT equations of motion. The former additional step is unwelcome in the pursuit of fast waveform generation, and the latter would remove all the benefit of the averaging transformation. Fortunately, it was noted in Ref. [18] that these oscillations can be averaged out if one performs an additional averaging transformation. We outline the details of this procedure in Section IV B.

NITs are not the only averaging procedure that can be used to accelerate the inspiral calculation. In Section IV C we present a two-timescale expansion of the equations of motion. One advantage of this approach is that the adiabatic and post-adiabatic inspirals can be computed offline. Given a particular mass ratio, the true inspiral can then be computed from interpolating func-

tions as a very fast online step. It may be advantageous to design data analysis pipelines to leverage this additional speed up.

We conclude this section by presenting averaged equations of motion for the case of quasi-spherical Kerr inspirals in Section IV D.

##### A. A review of near-identity transformations for generic EMRI systems

The NIT variables,  $\tilde{P}_j$ ,  $\tilde{q}_i$  and  $\tilde{S}_k$ , are related to the OG variables  $P_j$ ,  $q_i$  and  $S_k$  via

$$\tilde{P}_j = P_j + \epsilon Y_j^{(1)}(\vec{P}, \vec{q}) + \epsilon^2 Y_j^{(2)}(\vec{P}, \vec{q}) + \mathcal{O}(\epsilon^3), \quad (32a)$$

$$\tilde{q}_i = q_i + \epsilon X_i^{(1)}(\vec{P}, \vec{q}) + \epsilon^2 X_i^{(2)}(\vec{P}, \vec{q}) + \mathcal{O}(\epsilon^3), \quad (32b)$$

$$\tilde{S}_k = S_k + Z_k^{(0)}(\vec{P}, \vec{q}) + \epsilon Z_k^{(1)}(\vec{P}, \vec{q}) + \mathcal{O}(\epsilon^2). \quad (32c)$$

Here, the transformation functions  $Y_j^{(n)}$ ,  $X_i^{(n)}$ , and  $Z_k^{(n)}$  are required to be smooth, periodic functions of the orbital phases  $\vec{q}$ . At leading order, Eqs. (32) are identity transformations for  $P_k$  and  $q_i$  but not for  $S_k$  due to the presence of a zeroth order transformation term  $Z_k^{(0)}$ .

The inverse transformations can be found for  $P_k$  and  $q_i$  by requiring that their composition with the transformations in Eqs. (32) must give the identity transformation. To first order in  $\epsilon$ , this gives us

$$P_j = \tilde{P}_j - \epsilon Y_j^{(1)}(\vec{P}, \vec{q}) + \mathcal{O}(\epsilon^2), \quad (33a)$$

$$q_i = \tilde{q}_i - \epsilon X_i^{(1)}(\vec{P}, \vec{q}) + \mathcal{O}(\epsilon^2). \quad (33b)$$

It will be useful to decompose various functions into Fourier series where we use the convention:

$$A(\vec{P}, \vec{q}) = \sum_{\vec{\kappa} \in \mathbb{Z}^N} A_{\vec{\kappa}}(\vec{P}) e^{i\vec{\kappa} \cdot \vec{q}}, \quad (34)$$

where  $N$  is the number of orbital phases. Based on this, we can split the function into an averaged piece  $\langle A \rangle(\vec{P})$  and an oscillating piece

$$\check{A}(\vec{P}, \vec{q}) = A(\vec{P}, \vec{q}) - \langle A \rangle(\vec{P}) = \sum_{\vec{\kappa} \neq \vec{0}} A_{\vec{\kappa}}(\vec{P}) e^{i\vec{\kappa} \cdot \vec{q}}. \quad (35)$$

Using the above transformations along with the equations of motion, and working order by order in  $\epsilon$ , we can chose values for the transformation functions  $Y_j^{(1)}$ ,  $Y_j^{(2)}$ ,  $X_i^{(1)}$ ,  $X_i^{(2)}$ ,  $Z_k^{(0)}$  and  $Z_k^{(1)}$  such that the resulting equations of motion for  $\tilde{P}_j$ ,  $\tilde{q}_i$  and  $\tilde{S}_k$  take the following form

$$\frac{d\tilde{P}_j}{d\lambda} = \epsilon \tilde{F}_j^{(1)}(\vec{P}) + \epsilon^2 \tilde{F}_j^{(2)}(\vec{P}) + \mathcal{O}(\epsilon^3), \quad (36a)$$

$$\frac{d\tilde{q}_i}{d\lambda} = \Upsilon_i^{(0)}(\vec{P}) + \epsilon \Upsilon_i^{(1)}(\vec{P}) + \mathcal{O}(\epsilon^2), \quad (36b)$$

$$\frac{d\tilde{S}_k}{d\lambda} = \Upsilon_k^{(0)}(\vec{P}) + \epsilon \Upsilon_k^{(1)}(\vec{P}) + \mathcal{O}(\epsilon^2). \quad (36c)$$

Crucially, these equations of motion are now independent of the orbital phases  $\vec{q}$ . Deriving the relationship between the transformed forcing functions,  $\tilde{F}_j^{(1)}$ ,  $\tilde{F}_j^{(2)}$ ,  $\Upsilon_\alpha^{(0)}$  and  $\Upsilon_\alpha^{(1)}$ , to the original forcing functions is quite an involved process with several freedoms and choices, each with their own merits and drawbacks. This is discussed at length in paper I, so for brevity we will summarize the results and the particular choices we have made in this work. The transformed forcing functions are related to the original functions by

$$\tilde{F}_j^{(1)} = \langle F_j^{(1)} \rangle, \quad \Upsilon_i^{(1)} = \langle f_i^{(1)} \rangle, \quad \Upsilon_k^{(0)} = \langle f_k^{(0)} \rangle, \quad (37a-c)$$

$$\tilde{F}_j^{(2)} = \langle F_j^{(2)} \rangle + \left\langle \frac{\partial \check{Y}_j^{(1)}}{\partial \tilde{q}_i} \check{f}_i^{(1)} \right\rangle + \left\langle \frac{\partial \check{Y}_j^{(1)}}{\partial \tilde{P}_k} \check{F}_k^{(1)} \right\rangle, \quad (37d)$$

$$\Upsilon_k^{(1)} = - \left\langle \frac{\partial \check{f}_k^{(0)}}{\partial \tilde{P}_j} \check{Y}_j^{(1)} \right\rangle - \left\langle \frac{\partial \check{f}_k^{(0)}}{\partial \tilde{q}_i} \check{X}_i^{(1)} \right\rangle. \quad (37e)$$

Note that since we currently lack second order in mass ratio contributions, we will set  $\langle F_j^{(2)} \rangle = 0$ . In deriving these equations of motion, we have constrained the oscillating pieces of the NIT transformation functions to be

$$\check{Y}_j^{(1)} \equiv \sum_{\vec{k} \neq \vec{0}} \frac{i}{\vec{k} \cdot \check{\Upsilon}^{(0)}} F_{j,\vec{k}}^{(1)} e^{i\vec{k} \cdot \vec{q}}, \quad (38)$$

$$\check{X}_i^{(1)} \equiv \sum_{\vec{k} \neq \vec{0}} \left( \frac{i}{\vec{k} \cdot \check{\Upsilon}^{(0)}} f_{i,\vec{k}}^{(1)} + \frac{1}{(\vec{k} \cdot \check{\Upsilon}^{(0)})^2} \frac{\partial \Upsilon_i^{(0)}}{\partial P_j} F_{j,\vec{k}}^{(1)} \right) e^{i\vec{k} \cdot \vec{q}}, \quad (39)$$

and  $\check{Z}_k^{(0)}$  is found by solving

$$\check{f}_k^{(0)} + \frac{\partial \check{Z}_k^{(0)}}{\partial \tilde{q}_i} \Upsilon_i^{(0)} = 0. \quad (40)$$

This equation is satisfied by the oscillating pieces for the analytic solutions for the geodesic motion of  $t$  and  $\phi$ ,

$$\check{Z}_k^{(0)} = -\check{S}_{k,r}(q_r) - \check{S}_{k,z}(q_z). \quad (41)$$

We have chosen the averaged pieces  $\langle Y_j^{(1)} \rangle = \langle Y_j^{(2)} \rangle = \langle X_i^{(1)} \rangle = \langle X_i^{(2)} \rangle = \langle Z_k^{(0)} \rangle = \langle Z_k^{(1)} \rangle = 0$  for simplicity, though one could make other choices like in Ref. [18]. In order to generate waveforms, one only needs to know the transformations in Eq. (32) to zeroth order in the mass ratio, i.e.,

$$P_j = \tilde{P}_j + \mathcal{O}(\epsilon), \quad (42a)$$

$$q_i = \tilde{q}_i + \mathcal{O}(\epsilon), \quad (42b)$$

$$S_k = \tilde{S}_k - Z_k^{(0)}(\tilde{P}, \tilde{q}) + \mathcal{O}(\epsilon). \quad (42c)$$

Furthermore, to be able to directly compare between OG and NIT inspirals, we will need to match their initial conditions to sufficient accuracy. To maintain an overall phase difference of  $\mathcal{O}(\epsilon)$  throughout an inspiral, this requires the transformation of the  $P_j$ 's (32a) to linear order in  $\epsilon$ , while it is sufficient to know the rest of Eq. (32) to zeroth order.

## B. Averaging transformations for motion parametrized by Boyer-Lindquist coordinate time

Solving the above equations results in solutions for  $\tilde{P}$ ,  $\tilde{q}$  and  $\tilde{S}$  as functions of Mino-time  $\lambda$ . While this would include  $t(\lambda)$ , the transformation to  $\lambda(t)$  is non-trivial, and in practice is done via interpolation which can be costly for long inspirals. It would be significantly more convenient for the solutions to be functions of  $t$  from the start so that one can produce waveforms for data analysis without this post-processing step. This can be accomplished for the OG equations by simply using the chain rule:

$$\frac{dP_j}{dt} = \frac{1}{f_t^{(0)}(\tilde{P}, \tilde{q})} \left( \epsilon F_j^{(1)}(\tilde{P}) \right), \quad (43a)$$

$$\frac{dq_i}{dt} = \frac{1}{f_t^{(0)}(\tilde{P}, \tilde{q})} \left( \Upsilon_i(\tilde{P}) + \epsilon f_i(\tilde{P}, \tilde{q}) \right), \quad (43b)$$

$$\frac{d\phi}{dt} = \frac{1}{f_t^{(0)}(\tilde{P}, \tilde{q})} \left( f_\phi^{(0)}(\tilde{P}, \tilde{q}) \right). \quad (43c)$$

Notice that we have one less equation of motion to solve. However, using the same approach to the NIT equations of motion results in

$$\frac{d\tilde{P}_j}{dt} = \frac{1}{f_t^{(0)}(\tilde{P}, \tilde{q})} \left( \epsilon \tilde{F}_j^{(1)}(\tilde{P}) + \epsilon^2 \tilde{F}_j^{(2)}(\tilde{P}) \right), \quad (44a)$$

$$\frac{d\tilde{q}_i}{dt} = \frac{1}{f_t^{(0)}(\tilde{P}, \tilde{q})} \left( \Upsilon_i(\tilde{P}) + \epsilon \Upsilon_i^{(1)}(\tilde{P}) \right), \quad (44b)$$

$$\frac{d\tilde{\phi}}{dt} = \frac{1}{f_t^{(0)}(\tilde{P}, \tilde{q})} \left( \Upsilon_\phi(\tilde{P}) + \epsilon \Upsilon_\phi^{(1)}(\tilde{P}) \right). \quad (44c)$$

As we can see, we have now re-introduced a dependence on the orbital phases  $\vec{q}$ , defeating the purpose of our original NIT. Thankfully, as outlined in Ref. [18], these oscillations can also be averaged out by performing another transformation:

$$P_j = \tilde{P}_j + \epsilon \Pi_j^{(1)}(\tilde{P}, \tilde{q}) + \epsilon^2 \Pi_j^{(2)}(\tilde{P}, \tilde{q}) + \mathcal{O}(\epsilon^3), \quad (45a)$$

$$\varphi_\alpha = \tilde{Q}_\alpha + \Delta\varphi_\alpha + \epsilon \Phi_\alpha^{(1)}(\tilde{P}, \tilde{q}) + \mathcal{O}(\epsilon^2), \quad (45b)$$

where  $\tilde{Q} = \{\tilde{q}, \tilde{\phi}\}$ ,  $\Delta\varphi_\alpha = \Omega_\alpha^{(0)}(\tilde{P})\Delta t^{(0)}$  and  $\Omega_\alpha^{(0)}$  is the Boyer-Lindquist fundamental frequency of the tangent geodesic.

To obtain the equations of motion for  $\mathcal{P}_j$  and  $\varphi_i$ , we take the time derivative of Eq. (45), substitute in the expression for the NIT equations of motion, and then use

the inverse transformation of Eq. (45) to ensure that all functions are expressed in terms of  $\vec{\mathcal{P}}$  and  $\vec{q}$  and expanding order by order in  $\epsilon$ . We then chose the oscillatory functions  $\Delta t$ ,  $\Phi_i^{(1)}$ ,  $\Pi_j^{(1)}$  and  $\Pi_j^{(2)}$  to cancel out any oscillatory terms that appear at each order in  $\epsilon$ .

This results in averaged equations of motion that take the following form:

$$\frac{d\mathcal{P}_j}{dt} = \epsilon \Gamma_j^{(1)}(\vec{\mathcal{P}}) + \epsilon^2 \Gamma_j^{(2)}(\vec{\mathcal{P}}) + \mathcal{O}(\epsilon^3), \quad (46a)$$

$$\frac{d\varphi_\alpha}{dt} = \Omega_\alpha^{(0)}(\vec{\mathcal{P}}) + \epsilon \Omega_\alpha^{(1)}(\vec{\mathcal{P}}) + \mathcal{O}(\epsilon^2). \quad (46b)$$

These equations of motion are related to the Mino time averaged equations of motion Eqs. (36) where the adiabatic terms are given by

$$\Gamma_j^{(1)} = \frac{\tilde{F}_j^{(0)}}{\Upsilon_t^{(0)}}, \quad \Omega_\alpha^{(0)} = \frac{\Upsilon_\alpha^{(0)}}{\Upsilon_t^{(0)}}, \quad (47a-b)$$

and the post-adiabatic terms are given by

$$\begin{aligned} \Gamma_j^{(2)} &= \frac{1}{\Upsilon_t} \left( \tilde{F}_j^{(2)} + \tilde{F}_j^{(1)} \frac{\partial}{\partial \mathcal{P}_j} \langle \Pi_j^{(1)} \rangle \right. \\ &\quad \left. - \langle f_t^{(0)} \Pi_k^{(1)} \rangle \frac{\partial \Gamma_j^{(1)}}{\partial \mathcal{P}_k} - \Upsilon_t^{(1)} \Gamma_j^{(1)} \right), \end{aligned} \quad (48a)$$

$$\begin{aligned} \Omega_\alpha^{(1)} &= \frac{1}{\Upsilon_t^{(0)}} \left( \Upsilon_\alpha^{(1)} + \tilde{F}_j^{(1)} \left\langle \frac{\partial \Delta \varphi_\alpha}{\partial \mathcal{P}_j} \right\rangle \right. \\ &\quad \left. - \langle f_t^{(0)} \Pi_k^{(1)} \rangle \frac{\partial \Omega_\alpha^{(0)}}{\partial \mathcal{P}_k} - \Upsilon_t^{(1)} \Omega_\alpha^{(0)} \right). \end{aligned} \quad (48b)$$

This constrains the oscillating pieces of our transformation to be

$$\Delta t = \sum_{\kappa \neq 0} \frac{f_{t,\vec{\kappa}}^{(0)}}{-i\vec{\kappa} \cdot \vec{\Upsilon}^{(0)}} = -\check{Z}_t, \quad (49a)$$

$$\check{\Pi}_j^{(1)} = \sum_{\kappa \neq 0} \frac{f_{t,\vec{\kappa}}^{(0)}}{-i\vec{\kappa} \cdot \vec{\Upsilon}^{(0)}} \Gamma_j^{(1)} = -\check{Z}_t \Gamma_j^{(1)}, \text{ and} \quad (49b)$$

$$\begin{aligned} \Phi_{\alpha,\vec{\kappa}}^{(1)} &= \frac{i}{\vec{\kappa} \cdot \vec{\Upsilon}^{(0)}} \left( \frac{\partial \Delta \varphi_{\alpha,\vec{\kappa}}}{\partial \mathcal{P}_j} \tilde{F}_j^{(1)} - \frac{f_{t,\vec{\kappa}}^{(0)}}{\Upsilon_t^{(0)}} \Upsilon_t^{(1)} \Omega_\alpha^{(0)} \right. \\ &\quad \left. + \sum_{\vec{\kappa}' \neq \vec{0}} \left[ \left( i\vec{\kappa}' \cdot \vec{X}_{\vec{\kappa}-\vec{\kappa}'}^{(1)} f_{t,\vec{\kappa}'}^{(0)} + Y_{j,\vec{\kappa}-\vec{\kappa}'}^{(1)} \frac{\partial f_{t,\vec{\kappa}'}^{(0)}}{\partial \mathcal{P}_j} \right) \Omega_\alpha^{(0)} \right. \right. \\ &\quad \left. \left. - \Pi_{j,\vec{\kappa}-\vec{\kappa}'}^{(1)} f_{t,\vec{\kappa}'}^{(0)} \frac{\partial \Omega_\alpha^{(0)}}{\partial \mathcal{P}_j} \right] \right). \end{aligned} \quad (49c)$$

We are free to chose the averaged pieces of  $\Pi_j^{(1)}$  and we make the simplification that  $\langle \Pi_j^{(1)} \rangle = 0$ . With this and the identity  $\langle f_t^{(0)} (\int f_t^{(0)} d\vec{q}) \rangle = 0$ , we get the further

simplification  $\langle f_t^{(0)} \Pi_j^{(1)} \rangle = 0$ . Thus the expressions for  $\Gamma_j^{(2)}$  and  $\Omega_\alpha^{(1)}$  simplify to

$$\Gamma_j^{(2)} = \frac{1}{\Upsilon_t^{(0)}} \left( \tilde{F}_j^{(2)} - \Upsilon_t^{(1)} \Gamma_j^{(1)} \right), \quad (50a)$$

$$\Omega_\alpha^{(1)} = \frac{1}{\Upsilon_t^{(0)}} \left( \Upsilon_\alpha^{(1)} - \Upsilon_t^{(1)} \Omega_\alpha^{(0)} \right). \quad (50b)$$

What is most useful about these equations of motion is that their solutions  $\vec{\mathcal{P}}(t)$  and  $\vec{\varphi}(t)$  is exactly what is required to feed into waveform generating schemes. We show the equivalence between  $\vec{\varphi}(t)$  and the relationship derived in Ref. [82] between the solutions to the original NIT equations of motion and the waveform phases  $\Phi_{mn}$  in Appendix B.

### C. Two-timescale expansion

There is a related way of obtaining the above solutions via the two-timescale expansion (TTE). We exploit the difference between the timescales of the system by defining  $\mathcal{T} := \epsilon t$  as the slow time which governs the long term, secular behaviour of the system and defining  $t$  as the fast time of the system that governs the short term, orbital dynamics and treating these two times independent variables.

As such, we expand the transformed variables as

$$\mathcal{P}_j(\mathcal{T}, \epsilon) = \mathcal{P}_j^{(0)}(\mathcal{T}) + \epsilon \mathcal{P}_j^{(1)}(\mathcal{T}) + \mathcal{O}(\epsilon^2), \quad (51a)$$

$$\varphi_\alpha(\mathcal{T}, \epsilon) = \frac{1}{\epsilon} \left[ \varphi_\alpha^{(0)}(\mathcal{T}) + \epsilon \varphi_\alpha^{(1)}(\mathcal{T}) \right] + \mathcal{O}(\epsilon^1). \quad (51b)$$

Applying this expansion to the  $t$  parametrized NIT equations of motion, one finds that the equations of motion for the two-timescale expanded variables takes the form:

$$\frac{d\mathcal{P}_j^{(0)}}{d\mathcal{T}} = \Gamma_j^{(0)}(\vec{\mathcal{P}}^{(0)}), \quad (52a)$$

$$\frac{d\varphi_\alpha^{(0)}}{d\mathcal{T}} = \Omega_\alpha^{(0)}(\vec{\mathcal{P}}^{(0)}), \quad (52b)$$

$$\frac{d\mathcal{P}_j^{(1)}}{d\mathcal{T}} = \Gamma_j^{(2)}(\vec{\mathcal{P}}^{(0)}) + \mathcal{P}_k^{(1)} \left( \frac{\partial \Gamma_j^{(1)}}{\partial \mathcal{P}_k}(\vec{\mathcal{P}}^{(0)}) \right), \quad (52c)$$

$$\frac{d\varphi_\alpha^{(1)}}{d\mathcal{T}} = \Omega_\alpha^{(1)}(\vec{\mathcal{P}}^{(0)}) + \mathcal{P}_k^{(1)} \left( \frac{\partial \Omega_\alpha^{(0)}}{\partial \mathcal{P}_k}(\vec{\mathcal{P}}^{(0)}) \right). \quad (52d)$$

There is a trade-off for solving these equations of motion. We now have to solve a system of coupled differential equations that is twice the size and thus is more expensive to solve numerically, but the solutions are independent of  $\epsilon$  and one can construct a solution for any given value of  $\epsilon$  using Eqs. 51. Thus, if one wants to compute multiple inspirals with varying mass ratios, the TTE can be more efficient overall. However, there is also an

issue where the inspiral will stop earlier, as the variables  $\mathcal{P}_j^{(0)}$  typically reach values which correspond to the ISSO before  $\mathcal{P}_j$  do. In this regime, one should instead employ a transition to plunge as this is where the adiabaticity assumptions of the OG equations and the two-timescale expansion is expected to break down. As such we will use both the NIT and the TTE equations of motion to produce waveforms and compare them to waveforms generated using the OG equations to assess which is the more practical framework for producing post-adiabatic EMRI waveforms.

#### D. Averaged equations of motion for quasi-spherical Kerr inspirals

We now specialize the above results to the case of quasi-spherical inspirals into a Kerr black hole. The NIT equations of motion parameterized by Mino time take the form:

$$\frac{d\tilde{r}_p}{d\lambda} = \epsilon\tilde{F}_r^{(1)}(a, \tilde{r}_p, \tilde{x}) + \epsilon^2\tilde{F}_r^{(2)}(a, \tilde{r}_p, \tilde{x}), \quad (53a)$$

$$\frac{d\tilde{x}}{d\lambda} = \epsilon\tilde{F}_x^{(1)}(a, \tilde{r}_p, \tilde{x}) + \epsilon^2\tilde{F}_x^{(2)}(a, \tilde{r}_p, \tilde{x}), \quad (53b)$$

$$\frac{d\tilde{q}_z}{d\lambda} = \Upsilon_z^{(0)}(a, \tilde{r}_p, \tilde{x}) + \epsilon\Upsilon_z^{(1)}(a, \tilde{r}_p, \tilde{x}), \quad (53c)$$

$$\frac{d\tilde{t}}{d\lambda} = \Upsilon_t^{(0)}(a, \tilde{r}_p, \tilde{x}) + \epsilon\Upsilon_t^{(1)}(a, \tilde{r}_p, \tilde{x}), \quad (53d)$$

$$\frac{d\tilde{\phi}}{d\lambda} = \Upsilon_\phi^{(0)}(a, \tilde{r}_p, \tilde{x}) + \epsilon\Upsilon_\phi^{(1)}(a, \tilde{r}_p, \tilde{x}). \quad (53e)$$

The leading order terms in each equation of motion are simply the original functions, derived in Appendix A and paper II, averaged over a single geodesic orbit, i.e.,

$$\tilde{F}_r^{(1)} = \langle F_r^{(1)} \rangle, \quad \tilde{F}_x^{(1)} = \langle F_x^{(1)} \rangle, \quad \Upsilon_z^{(1)} = \langle f_z^{(1)} \rangle, \quad (54a-c)$$

$$\Upsilon_t^{(0)} = \langle f_t^{(0)} \rangle, \quad \Upsilon_\phi^{(0)} = \langle f_\phi^{(0)} \rangle, \quad (54d-e)$$

where  $\Upsilon_t^{(0)}$  and  $\Upsilon_\phi^{(0)}$  are the Mino-time  $t$  and  $\phi$  fundamental frequencies which are known analytically [39]. The remaining terms are more complicated and require Fourier decomposing the original functions and their derivatives with respect to the orbital elements  $(r_p, x)$ . To express the result, for any function

$$A(a, r_p, x, q_z) = \sum_{\kappa} A_{\kappa}(a, r_p, x) e^{i\kappa q_z}, \quad (55)$$

we define the operator

$$\begin{aligned} \mathcal{N}(A) = & \sum_{\kappa \neq 0} \frac{-1}{\Upsilon_z^{(0)}} \left[ A_{\kappa} f_{z, -\kappa} - \frac{i}{\kappa} \left( \frac{\partial A_{\kappa}}{\partial \tilde{r}_p} F_{r, -\kappa} + \frac{\partial A_{\kappa}}{\partial \tilde{x}} F_{x, -\kappa} \right. \right. \\ & \left. \left. - \frac{A_{\kappa}}{\Upsilon_z^{(0)}} \left( \frac{\partial \Upsilon_z^{(0)}}{\partial \tilde{r}_p} F_{r, -\kappa} + \frac{\partial \Upsilon_z^{(0)}}{\partial \tilde{x}} F_{x, -\kappa} \right) \right) \right]. \end{aligned} \quad (56)$$

With this in hand, the remaining terms in the equations of motion are found to be

$$\begin{aligned} \tilde{F}_r^{(2)} &= \mathcal{N}(F_r^{(1)}), & \tilde{F}_x^{(2)} &= \mathcal{N}(F_x^{(1)}), \\ \Upsilon_t^{(1)} &= \mathcal{N}(f_t^{(0)}), & \Upsilon_\phi^{(1)} &= \mathcal{N}(f_\phi^{(0)}). \end{aligned} \quad (56a-d)$$

Combining these results with Eqs. (46), one can find the NIT equations of motion parameterized by Boyer-Lindquist time  $t$  for the phases  $\vec{\varphi} = \{\varphi_z, \varphi_\phi\}$  and orbital elements  $\vec{\mathcal{P}} = \{r_\varphi, x_\varphi\}$  in the form

$$\frac{dr_\varphi}{dt} = \epsilon\Gamma_r^{(1)}(a, r_\varphi, x_\varphi) + \epsilon^2\Gamma_r^{(2)}(a, r_\varphi, x_\varphi), \quad (57a)$$

$$\frac{dx_\varphi}{dt} = \epsilon\Gamma_x^{(1)}(a, r_\varphi, x_\varphi) + \epsilon^2\Gamma_x^{(2)}(a, r_\varphi, x_\varphi), \quad (57b)$$

$$\frac{d\varphi_z}{dt} = \Omega_z^{(0)}(a, r_\varphi, x_\varphi) + \epsilon\Omega_z^{(1)}(a, r_\varphi, x_\varphi), \quad (57c)$$

$$\frac{d\varphi_\phi}{dt} = \Omega_\phi^{(0)}(a, r_\varphi, x_\varphi) + \epsilon\Omega_\phi^{(1)}(a, r_\varphi, x_\varphi). \quad (57d)$$

The leading order terms in these equations are given by

$$\Gamma_r^{(1)} = \tilde{F}_r^{(1)}/\Upsilon_t^{(0)}, \quad \Gamma_x^{(1)} = \tilde{F}_x^{(1)}/\Upsilon_t^{(0)}, \quad (58a-b)$$

$$\Omega_z^{(0)} = \Upsilon_z^{(0)}/\Upsilon_t^{(0)}, \quad \Omega_\phi^{(0)} = \Upsilon_\phi^{(0)}/\Upsilon_t^{(0)}. \quad (58c-d)$$

The sub-leading terms are given by

$$\Gamma_r^{(2)} = \frac{1}{\Upsilon_t^{(0)}} \left( \tilde{F}_r^{(2)} - \Upsilon_t^{(1)}\Gamma_r^{(1)} \right), \quad (59a)$$

$$\Gamma_x^{(2)} = \frac{1}{\Upsilon_t^{(0)}} \left( \tilde{F}_x^{(2)} - \Upsilon_t^{(1)}\Gamma_x^{(1)} \right), \quad (59b)$$

$$\Omega_z^{(1)} = \frac{1}{\Upsilon_t^{(0)}} \left( \Upsilon_z^{(1)} - \Upsilon_t^{(1)}\Omega_z^{(0)} \right), \quad (59c)$$

$$\Omega_\phi^{(1)} = \frac{1}{\Upsilon_t^{(0)}} \left( \Upsilon_\phi^{(1)} - \Upsilon_t^{(1)}\Omega_\phi^{(0)} \right). \quad (59d)$$

Finally, using the two timescale expansion the adiabatic equations of motion are given by

$$\frac{dr_\varphi^{(0)}}{d\mathcal{T}} = \Gamma_r^{(1)}, \quad \frac{dx_\varphi^{(0)}}{d\mathcal{T}} = \Gamma_x^{(1)}, \quad (60a-b)$$

$$\frac{d\varphi_z^{(0)}}{d\mathcal{T}} = \Omega_z^{(0)}, \quad \frac{d\varphi_\phi^{(0)}}{d\mathcal{T}} = \Omega_\phi^{(0)}. \quad (60c-d)$$

The post-adiabatic contributions to the equations of motion are given by

$$\frac{dr_\varphi^{(1)}}{d\mathcal{T}} = \Gamma_r^{(2)} + r_\varphi^{(1)} \frac{\partial \Gamma_r^{(1)}}{\partial r_\varphi^{(0)}} + x_\varphi^{(1)} \frac{\partial \Gamma_r^{(1)}}{\partial x_\varphi^{(0)}}, \quad (61a)$$

$$\frac{dx_\varphi^{(1)}}{d\mathcal{T}} = \Gamma_x^{(2)} + r_\varphi^{(1)} \frac{\partial \Gamma_x^{(1)}}{\partial r_\varphi^{(0)}} + x_\varphi^{(1)} \frac{\partial \Gamma_x^{(1)}}{\partial x_\varphi^{(0)}}, \quad (61b)$$

$$\frac{d\varphi_z^{(1)}}{d\mathcal{T}} = \Omega_z^{(1)} + r_\varphi^{(1)} \frac{\partial \Omega_z^{(0)}}{\partial r_\varphi^{(0)}} + x_\varphi^{(1)} \frac{\partial \Omega_z^{(0)}}{\partial x_\varphi^{(0)}}, \quad (61c)$$

$$\frac{d\varphi_\phi^{(1)}}{d\mathcal{T}} = \Omega_\phi^{(1)} + r_\varphi^{(1)} \frac{\partial \Omega_\phi^{(0)}}{\partial r_\varphi^{(0)}} + x_\varphi^{(1)} \frac{\partial \Omega_\phi^{(0)}}{\partial x_\varphi^{(0)}}, \quad (61d)$$

Solutions for post-adiabatic inspirals can be obtained by solving Eqs. (60) and (61) simultaneously and using Eqs. (51) along with a value for the mass ratio  $\epsilon$  to recover  $r_\varphi(t)$ ,  $x_\varphi(t)$ ,  $\varphi_z(t)$  and  $\varphi_z(t)$ .

With the averaged equations for quasi-spherical Kerr inspirals in hand, we will now outline our numerical implementation for rapidly computing quasi-spherical self-forced inspirals.

## V. IMPLEMENTATION

Combining the interpolated GSF model along with our action angle formulation of the OG equations provides us with all the information required to calculate the NIT and TTE equations of motion. We first evaluate and interpolate the various terms in the NIT and TTE equations of motion across the parameter space. While this offline process can be expensive, it only need to be completed once. Once completed, the online process of calculating self-forced inspirals can be completed in less than a second. This procedure is very similar to Papers I & II, though now we also export interpolating functions for the partial derivatives needed for the TTE equations of motion.

### A. Offline Steps

The offline calculation consists of the following steps.

1. We start by selecting a grid which covers the parameter space. We chose  $y$  values between 0.099 and 0.999 in 451 equally spaced steps and  $z_-^2$  values from 0.002 to 0.5 in 250 equally spaced steps (giving a total of 112,750 points).<sup>1</sup>
2. For each point in the parameter space  $(a, y, z_-^2)$  we evaluate the functions  $F_{r\backslash x}^{(1)}$ ,  $f_z^{(1)}$  and  $f_{t\backslash\phi}^{(0)}$  along with their derivatives with respect to  $r_p$  and  $x$  for 49 equally spaced values of  $q_z$  from 0 to  $2\pi$ .
3. We then perform a fast Fourier transform on the output data to obtain the Fourier coefficients of the forcing functions and their analytical derivatives with respect to  $r_p$  and  $x$ .
4. With these, we then use Eqs. (54a-c), (54d-e), (56) and (56a-d) to construct  $\tilde{F}_{r\backslash x}^{(1\backslash 2)}$ ,  $\partial\tilde{F}_{r\backslash x}^{(1)}/\partial r_p$ ,

---

<sup>1</sup> Evaluating the NIT functions is computationally cheap so using a dense grid does not significantly increase the computational burden. Using an equally spaced grid also allows us to use Mathematica's default Hermite polynomial interpolation method for convenience of implementation. The grid spacing is chosen to be sufficiently dense that interpolation error is a negligible source of error for our comparisons between the OG, NIT and TTE inspirals, though a less dense grid may also achieve this.

$\partial\tilde{F}_{r\backslash x}^{(1)}/\partial x$ , and  $\Upsilon_{t\backslash z\backslash\phi}^{(1)}$  and that point in the parameter space. All other terms needed for the NIT and TTE can be derived from these terms or are already known analytically.

5. We also use Eqs. (38) to construct the Fourier coefficients of the first order transformation functions  $Y_{r\backslash x}^{(1)}$ . These are needed when comparing NIT and OG/TTE inspirals, to ensure that the initial conditions are comparable. Otherwise this step can be skipped.
6. We then repeat this procedure across the parameter space for each point in our grid.
7. Finally we interpolate the values for  $\tilde{F}_{r\backslash x}^{(1\backslash 2)}$ ,  $\partial\tilde{F}_{r\backslash x}^{(1)}/\partial r_p$ ,  $\partial\tilde{F}_{r\backslash x}^{(1)}/\partial p$ , and  $\Upsilon_{t\backslash z\backslash\phi}^{(1)}$  along with the coefficients of  $Y_{r\backslash x}^{(1)}$  across this grid using Hermite interpolation and store the interpolants for future use.

We implemented the above algorithm in Mathematica 12.2 and find the calculation takes about 5 hours to complete when parallelized across 20 CPU cores. Since these offline steps need only be completed once, this is a comparatively small price to pay.

### B. Online Steps

By contrast, the online steps are required for every inspiral calculation, but are computationally inexpensive. The online steps for computing a NIT or TTE inspiral are as follows.

1. We load in the interpolants for  $\tilde{F}_{r\backslash x}^{(1\backslash 2)}$ , and  $\Upsilon_{t\backslash z\backslash\phi}^{(1)}$ , and define the NIT equations of motion. For TTE inspirals, one also needs to load in the interpolated derivatives  $\partial\tilde{F}_{r\backslash x}^{(1)}/\partial r_p$  and  $\partial\tilde{F}_{r\backslash x}^{(1)}/\partial x$  then define the TTE equations of motion.
2. In order to facilitate comparisons between OG, NIT, and TTE inspirals, we also load interpolants of the Fourier coefficients of  $\check{Y}_{r/x}^{(1)}$  and Eqs. (32), (45) and (49b) to construct first order near-identity transformations.
3. We state the initial conditions of the inspiral  $(r_p(0), x(0), q_z(0))$  and use the NIT to leading order in the mass ratio to transform these into initial conditions for the NIT/TTE equations of motion, i.e.,  $(r_\varphi(0), x_\varphi(0), \varphi_z(0))$ .
4. We then evolve the NIT or TTE equations of motion using an ODE solver (in our case Mathematica's `NDSolve`).

As with the offline steps we implement the online steps in Mathematica. Note that steps (ii) and (iii) are only

necessary because we want to make direct comparisons between NIT and OG inspirals with the same initial conditions. In general, the difference between the NIT and OG variables will always be  $\mathcal{O}(\epsilon)$ , and so performing the NIT transformation or inverse transformation to greater than zeroth order in mass ratio will not be necessary when producing waveforms to post adiabatic order, i.e. with phases accurate to  $\mathcal{O}(\epsilon)$ .

### C. Waveform Generation

With a trajectory in hand, can now generate the gravitational waveform. Ideally, this would be done using interpolated Teukolsky fluxes as implemented by, e.g., the `FastEMRIWaveform` package for eccentric Schwarzschild inspirals [83, 84]. Unfortunately, such a model is not currently available for quasi-spherical Kerr inspirals, and so we generate all of our waveforms using the “semi-relativistic” approximation used by the numerical kludge models [85]. In this approach the Boyer-Lindquist coordinates of the particle are mapped to flat spacetime coordinates which are fed into the quadrupole formula which is used to produce the final waveform strain  $h = h_+ + ih_\times$ . This approximation fares surprisingly well compared to Teukolsky snapshot waveforms, even in the strong field regime. What is most important for our purposes is that all waveforms are calculated using the same kludge formula so that any differences in the waveforms is due to differences in the calculations of the inspiral trajectory. For our implementation, we sample every  $t_{\text{step}} = 2M$  and calculate the waveform strain at each time-step to produce our numerical waveforms.

In order to calculate the fractional overlap  $\mathcal{O}(h_1, h_2)$  between two time-domain waveform strains  $h_1$  and  $h_2$ , we first define the noise weighted inner product of these two waveforms as

$$\langle h_1 | h_2 \rangle = 2 \int_0^\infty \frac{\tilde{h}_1^*(f) \tilde{h}_2(f) + \tilde{h}_1(f) \tilde{h}_2^*(f)}{S_n(f)} df, \quad (62)$$

where  $\tilde{h}(f)$  denotes the Fourier transform of the time domain waveform  $h(t)$  and  $h^*$  is the complex conjugate of  $h$ . We take the power spectral density (PSD) of the detector noise  $S_n(f)$  to be a flat noise curve.

From this, one can calculate the fractional waveform overlap  $\mathcal{O}$  via

$$\mathcal{O} = \frac{\langle h_1 | h_2 \rangle}{\sqrt{\langle h_1 | h_1 \rangle \langle h_2 | h_2 \rangle}}. \quad (63)$$

This overlap ranges from 1, when the two waveforms are identical, to 0, when the waveforms are perfectly orthogonal. When dealing with waveform overlaps close to 1, it is often useful to talk in terms of the fractional waveform mismatch  $\mathcal{M} = 1 - \mathcal{O}$ . In practice, we make use of the `WaveformMatch` function from the `SimulationTools` package to calculate waveform overlaps [86].

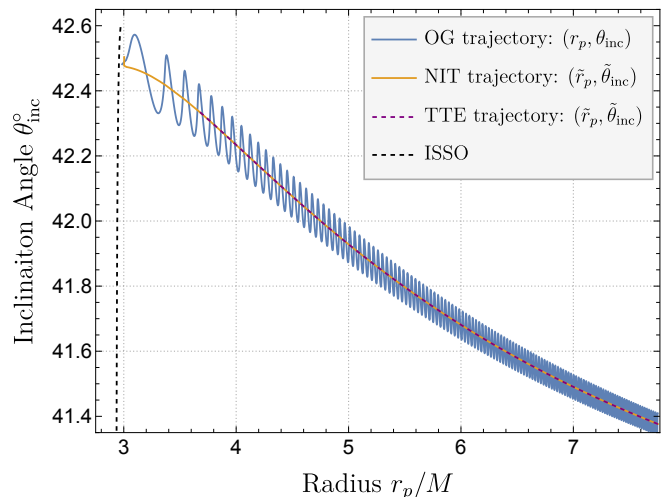


FIG. 3. The trajectory through  $(r_p, \theta_{\text{inc}})$  space for an inspiral with  $\epsilon = 10^{-2}$ ,  $a = 0.9M$ , and initial conditions  $r_p(0) = 7.75M$ ,  $x(0) = 0.75$ . We use such a large mass ratio to highlight the orbital timescale oscillations one encounters when using the OG equations. Using the NIT or TTE equations of motion averages out these oscillations and results in almost identical inspirals. We also see that the TTE equations of motion break down further from the ISSO than the OG or NIT equations of motion.

## VI. RESULTS

### A. OG vs NIT and TTE inspirals

To test the accuracy of our NIT and TTE implementations, we compare inspirals calculated using the OG equations of motion against inspirals calculated using the NIT or TTE equations of motion. We use a case study of a typical system EMRI with a primary of mass  $M = 10^6 M_\odot$ . We chose the initial conditions for this inspiral to have an inclination of  $x = 0.75$  and radius  $r_p = 7.75M$ . This highly inclined, strong field inspiral provides a good test of our numerical implementations. We see similar results for other initial conditions. We also set the initial phases  $q_z(0) = \phi(0) = 0$  for simplicity.

Figure 3 demonstrates the difference in the trajectories produced by each method. We have purposely chosen an unusually large mass ratio for an EMRI of  $\epsilon = 10^{-2}$  so that the orbital timescale oscillations are clearly visible on the plot. We see that for all trajectories the orbital separation decreases substantially, while the orbit becomes slightly more inclined. This is consistent with previous results for adiabatic quasi-spherical inspirals [11–13]. The OG trajectory oscillates on the orbital timescale, while the NIT and TTE trajectories faithfully capture the average evolution of the trajectory. While the NIT and TTE trajectories may appear at first glance to be identical, it is important to note that the TTE equations of motion break down sooner than the NIT equations of motion. This is due to the evolution of  $r_\varphi^{(0)}$

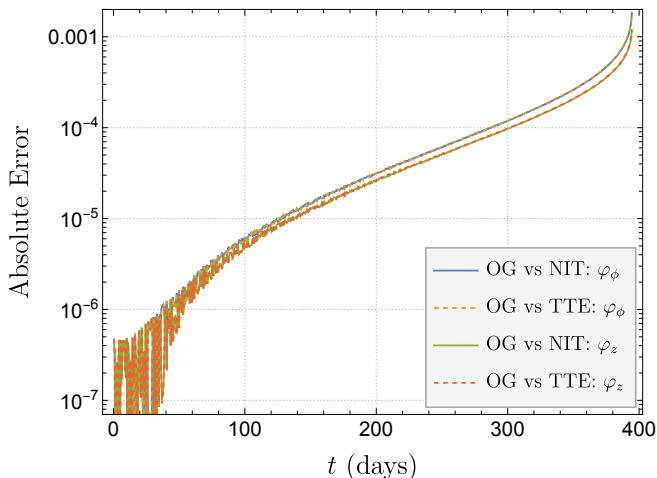


FIG. 4. The difference in the orbital phases for a quasi-spherical Kerr inspiral with  $M = 10^6 M_\odot$ ,  $\epsilon = 10^{-5}$ , and  $a = 0.9M$  when using the OG equations verse the NIT or TTE equations of motion. In both cases, the difference stays small throughout the inspiral, only becoming large as the secondary approaches the ISSO when the adiabatic assumption breaks down.

and  $x_\varphi^{(0)}$  reaching the location of the ISSO sooner than  $r_\varphi$  and  $x_\varphi$ . This is not as important of an issue as one might expect, as this close to the ISSO one should swap over to a “transition to plunge” scheme [87–90].

Figure 4 shows the difference in the phases between the OG trajectory and the NIT and TTE trajectories for a year long inspiral with a primary of mass  $M = 10^6 M_\odot$  and mass ratio of  $\epsilon = 10^{-5}$ . In both cases, we find that the difference in the phases stays below  $\sim 10^{-3}$  throughout the inspiral, spiking only when the inspiral approaches the ISSO where the adiabatic assumption implicit in the OG, NIT and TTE equations of motion starts to break down. Even then, the difference in the phases is substantially lower than subradian accuracy requirement needed for LISA data analysis. We have also found that the growth in the error over time is most closely correlated with the interpolation error for the terms in the NIT and TTE equations of motion, and so interpolating on a denser grid should reduce this error even more.

We note that formally both the NIT and the TTE should induce an error in both the phases and the orbital elements that scales linearly with the mass ratio. To ensure that our implementation is converging correctly, we evolve inspirals from initial conditions  $r_p(0) = 5M$  and  $x(0) = 0.75$  until they reach  $r_p = 4M$  with different values of the mass ratio. Since the OG quantities will have oscillations of  $\mathcal{O}(\epsilon)$ , sampling the error at only the final time-step would make it difficult to determine the convergence with the mass ratio. As such we calculate the error at several time-steps in the last three orbital cycles and report the largest error. The results of this test can be seen in Figure 5. We see that for larger mass ratios the errors converge linearly, as expected. However, for mass

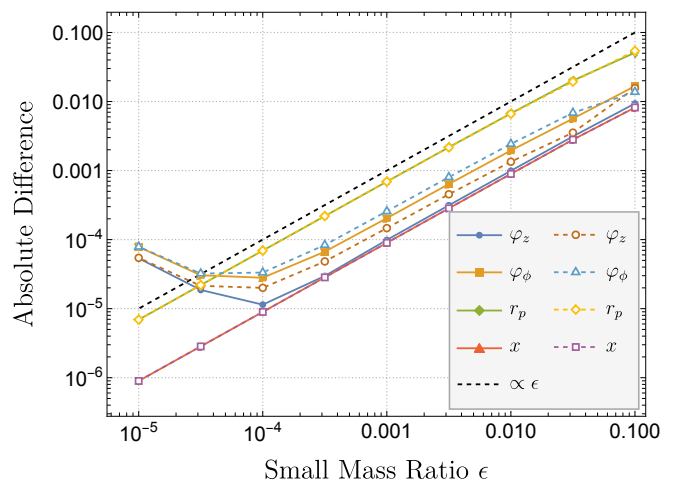


FIG. 5. The absolute difference in the quantities of a prograde inspiral with  $a = 0.9M$  and initial inclination of  $x_0 = 0.75$  after evolving from  $r_p = 5M$  to  $r_p = 4M$  when using different evolution equations. The solid lines are the difference between OG and NIT equations of motion and the dashed lines are the difference between using the OG and TTE equations of motion. We see that the differences generally follow the black  $\epsilon$  curve and so converge linearly with the mass ratio until we reach mass ratios  $\leq 10^{-3}$  where the error in the phases becomes dominated by interpolation and numerical error.

$\epsilon$	OG	NIT	Speedup	TTE	Speedup
$10^{-2}$	24.7s	0.48s	$\sim 51$	22ms	$\sim 1.1 \times 10^6$
$10^{-3}$	4m 20s	0.17s	$\sim 1530$	39ms	$\sim 6.7 \times 10^6$
$10^{-4}$	43m 19s	0.43s	$\sim 6044$	27ms	$\sim 9.6 \times 10^7$
$10^{-5}$	7hrs 42m	0.49s	$\sim 54380$	22ms	$\sim 1.2 \times 10^9$

TABLE I. Computational time required to evolve an inspiral from its initial conditions of  $r_p(0) = 7.75M$  and  $x(0) = 0.75$  to the last stable orbit for different values of the mass ratio, as calculated in Mathematica 13 on an Intel Core i7 @ 2.2GHz. The computational time for the OG inspiral scales inversely with the mass ratio, whereas the computational time for NIT inspirals is independent of the mass ratio. The computation for the TTE inspiral was 0.53s, which is slightly longer than any of the NIT inspirals. However, if we consider the TTE calculation as an offline step, one can then immediately recover the solution for any value of  $\epsilon$  in a matter of milliseconds.

ratios  $\lesssim 10^{-4}$ , the formal error in the NIT and TTE is no longer the dominant source of error for the evolution of the phases. Instead the error is dominated by either interpolation error or the error in the ODE solver when solving the OG equations as we have found empirically that this error is reduced by increasing the number of grid points used for interpolation and decreasing the relative tolerance of the integrator. While both of these sources can be further suppressed at the cost of more expensive offline and online steps respectively, we show in Sec. VI B that it is not necessary for producing trajectories accurate enough for LISA data science.

Now that we have established our averaging procedure does not change the accuracy of our model, we demonstrate the speed up that one enjoys from using either the NIT or TTE equations of motion instead of using the OG equations of motion in Table. I. For each of these calculations, the initial conditions are set to  $r_p(0) = 7.75M$  and  $x(0) = 0.75$  the inspirals are evolved to just before the ISSO, when  $y(r_p, x) = 0.998$ . All calculations were done using machine precision arithmetic and an accuracy goal of 7 digits for Mathematica's `NDSolve` function. We find that using the OG inspiral calculation takes longer as the mass ratio gets smaller since the solver will have to resolve many more orbital cycles before the inspiral reaches the ISSO. However, the NIT inspirals take roughly the same amount of time regardless of the mass ratio. The TTE inspiral takes slightly longer than the NIT inspiral, as one has to solve for twice as many equations. However once this step is complete, inspirals with the same initial conditions but any mass ratio can be computed extremely rapidly by evaluating the resulting interpolating functions.

### B. Waveform mismatches

To generate waveforms, we use a semi-relativistic approximation to produce quadrupole waveforms which can be seen in Figs. 6a and 6b [85]. These figures show the first and last 3.5 hours of the waveforms produced by the OG, NIT and TTE trajectories respectively, sampled once every 10s. From the plots it is clear that the two waveforms overlap significantly. The waveform mismatch between the OG waveforms and both the NIT and TTE waveforms is  $\sim 2.5 \times 10^{-8}$ . This means that they meet the indistinguishability criteria [91–93] from the OG waveforms for signal-to-noise ratios (SNRs) of up to at least 4500.

Intuitively, one would expect the error in our averaging procedure to scale with the size of the orbital timescale oscillations which themselves scale with the mass ratio and the orbital inclination. As such, we wish determine the section of the parameter space where the difference in the waveforms between the OG and the NIT/TTE inspirals would be small enough not to matter for LISA data science. To do this, we first fix the mass of the primary to be  $10^6 M_\odot$  and created a function which uses adiabatic inspirals and root-finding to numerically compute the initial orbital separation for an inspiral which will take one year to reach the ISSO for a given inclination and mass ratio. We then create a grid of mass ratios  $\epsilon = \{1, 10^{-1/2}, 10^{-1}, 10^{-3/2}, 10^{-2}, 10^{-5/2}, 10^{-3}, 10^{-7/2}, 10^{-4}\}$  and inclinations  $x = \{0.74, 0.79, 0.84, 0.89, 0.94, 0.99\}$ . For each point on this grid, we calculate a year long inspiral using the OG equations of motion, the NIT equations of motion, and the TTE equations of motion and then generate a waveform from each.

The waveform mismatch between the OG and NIT or TTE waveforms is displayed in Figure 7. From these

plots we see that there does not seem to be a substantial difference in terms of accuracy from using either the NIT or TTE equations of motion. While there may be some correlation between mismatch and inclination, this does not appear to be a strong effect for  $\theta_{\text{inc}} \leq 45^\circ$ . The strongest effect on the waveform mismatch comes from the mass ratio. It is worth noting that  $3 \times 10^{-2}$  is a commonly chosen maximum mismatch for a waveform template bank that corresponds to a 90 %-ideal observed event rate [94]. Our results suggests that one could, in principle, produce such a template bank of quasi-spherical inspirals using NIT or TTE waveforms even for mass ratios as large as  $\epsilon \approx 0.5$ . In practice, we are still missing important post-adiabatic contributions such as second-order effects and contributions from the spin of the secondary, and so such a template bank would have substantial systematic biases. Also note that, formally, when all post-adiabatic corrections are included, the OG, NIT, and TTE inspirals have the same accuracy in terms of powers of the mass-ratio. Consequently, we cannot say a priori which will be more faithful to Nature.

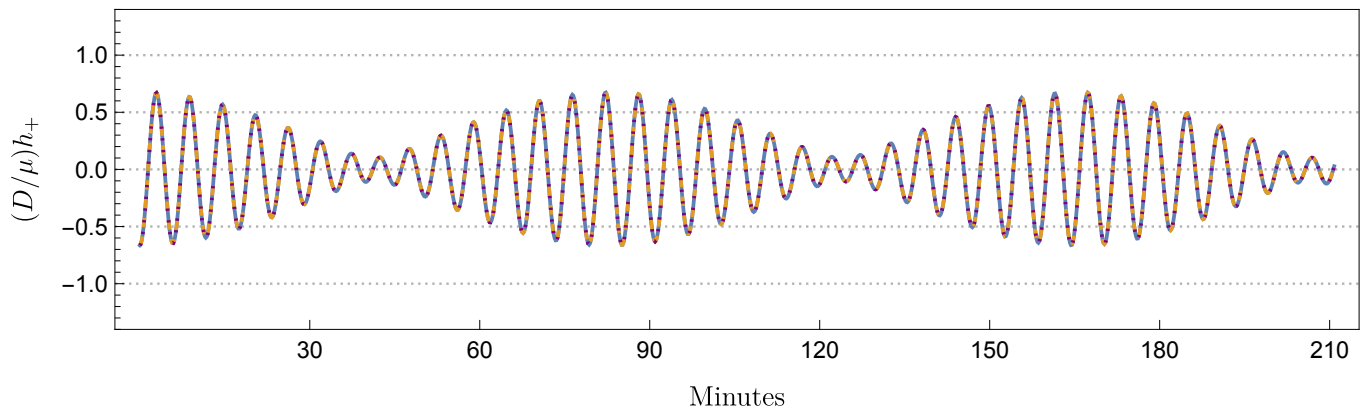
### C. Using Higher Precision Fluxes

Now that we can compute fast and accurate inspirals, it is worth recalling that the relative accuracy of our interpolated GSF model is currently too low for production level waveforms. This is primarily due to the harsh relative accuracy requirements for the adiabatic pieces of  $\lesssim \epsilon$  for subradian accuracy in the phases. This can be improved by incorporating information from the asymptotic fluxes, which can be interpolated to a much higher accuracy across the parameter space due to the much cheaper cost of flux calculations compared to GSF calculations. This is why our interpolated flux model is accurate to  $\sim 10^{-6}$  while our interpolated GSF model is only accurate to  $\sim 10^{-3}$ . Such information can be incorporated into the GSF model itself to improve the accuracy of the OG inspirals as well as the resulting NIT and TTE inspirals [27]. But since the NIT and TTE equations of motion are naturally split into adiabatic and post-adiabatic pieces, one can calculate  $\Gamma_j^{(1)}$  directly from the fluxes via equations (22) and (24), interpolate them to higher precision than using a GSF model, and then substitute these improved adiabatic terms into the averaged equations of motion.

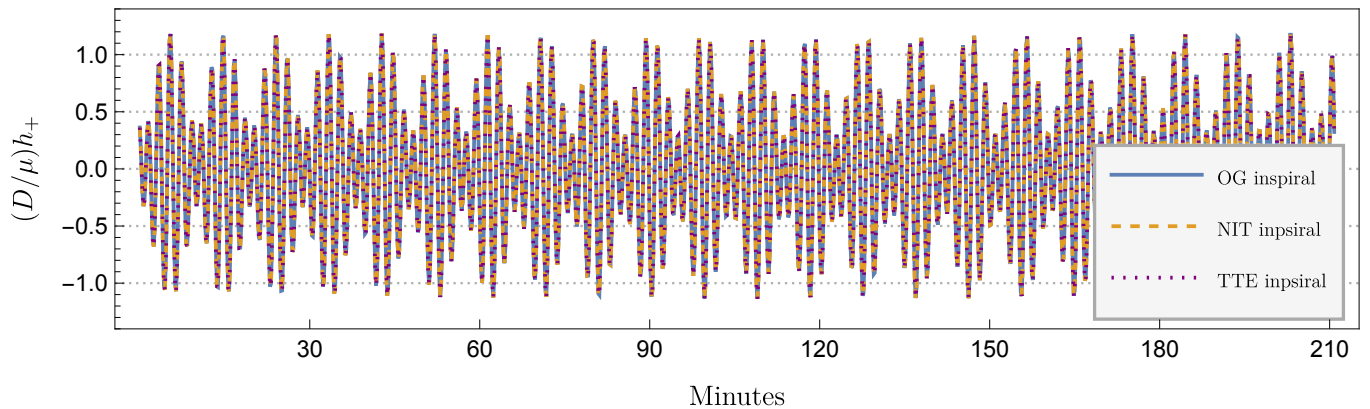
To test the difference this would make to the overall accuracy of our post-adiabatic inspirals, we looked at the final error in the  $\varphi_\phi$  phase when evolved from one point in the parameter space to the ISSO when using the flux+GSF model verses using only the GSF model for inspirals with  $\epsilon = 10^{-5}$ . From Figure 8, we see that improving the adiabatic pieces results in a phase difference that can range for from tens to tens of thousands of radians for multi-year long inspirals and this difference gets larger as one moves away from the ISSO.

As such, using high precision fluxes for the adiabatic





(a) The first 3.5 hours of the waveform.



(b) The last 3.5 hours of the waveform.

FIG. 6. The “semi-relativistic” quadrupole waveform strains for an EMRI with  $M = 10^6 M_\odot$ ,  $\epsilon = 10^{-5}$ ,  $r_p(0) = 7.75M$ , and  $x(0) = 0.75$  as viewed from edge on sampled once every 10s. Panel (a) shows the first 3.5 hours and panel (b) shows the last 3.5 hours of this year long waveform. The blue curve is the waveform generated from the OG inspiral, the dashed yellow curve is the waveform generated from the NIT inspiral, and the purple dotted curve is the waveform generated from the TTE inspiral.

contributions to the equations of motion is vital for obtaining accurate adiabatic and post-adiabatic inspirals.

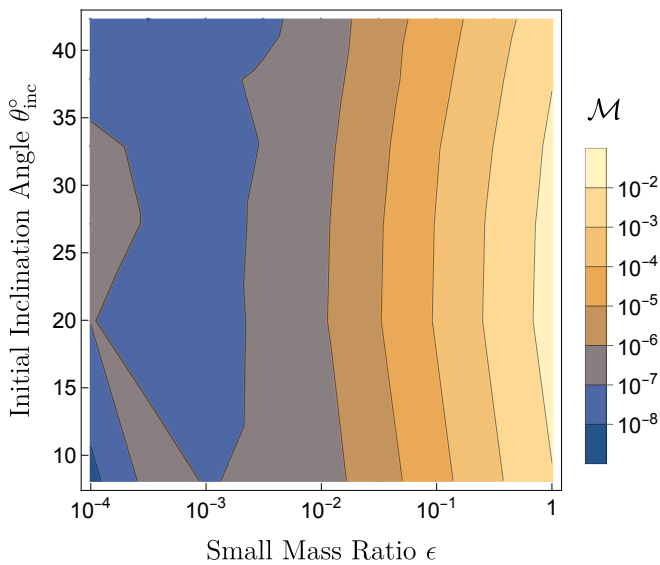
#### D. Impact of the self-force on spherical inspirals

Since this is the first time that the full first order self-force has been computed for quasi-spherical Kerr inspirals, we examine the impact that the adiabatic and post-adiabatic contributions have on the inspiral in Figure 9. The blue curves show typical adiabatic trajectories through  $\{r_p, \theta_{\text{inc}}\}$  space. From this we see that the self-force causes the orbital radius to decrease over time, but also cause the inclination angle to increase slightly throughout the inspiral. This is consistent with previous work on adiabatic quasi-spherical inspirals [11, 13]. The post-adiabatic contributions to the solutions for  $r_p$  and  $\theta_{\text{inc}}$  do not change this trend in any significant way since their contribution to the orbital elements is  $\mathcal{O}(\epsilon)$ . However, the post-adiabatic contribution to the orbital phases is of order  $\mathcal{O}(\epsilon^0)$  and is thus very significant. This

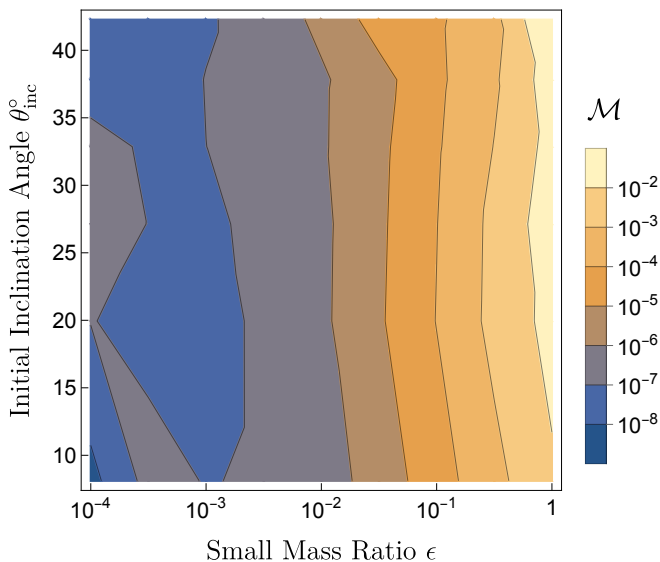
is demonstrated by the dashed contours on Figure 9, which indicate the final value of the post-adiabatic piece of the azimuthal phase  $\varphi_\phi^{(1)}$ , when evolved from a given point in the parameter space to the ISSO. Since it is the azimuthal phase that most strongly impacts the final waveform phase, this gives a good estimate of how many radians one can expect the waveform to be out of phase in the absence of post-adiabatic contributions. It is worth restating that this does not include the second-order GSF contributions to  $\varphi_\phi^{(1)}$  as these are not currently known.

## VII. DISCUSSION

This paper presents the first calculations of the first order gravitational self-force in the radiation gauge for spherical orbits in Kerr spacetime by utilizing a modified version of the code found in Refs. [24, 25, 50]. The main new element in this calculation is the inclusion of the gauge completion piece that puts the result in the correct asymptotically flat gauge.



(a) Mismatch between OG and NIT waveforms.



(b) Mismatch between OG and TTE waveforms.

FIG. 7. The mismatch between year long OG and NIT/TTE waveforms as a function of orbital inclination and mass ratio. We do not see a substantial difference in accuracy between using either NIT or TTE waveforms. We also see that the mismatch remains smaller than 0.03 for mass ratios as large as  $\epsilon \approx 0.5$ .

To produce a continuous model this data is interpolated using Fourier decomposition and Chebyshev interpolation. Using only 162 points we obtain a model with sub-percent accuracy for inclinations up to  $0^\circ \leq \theta_{\text{inc}} \leq 45^\circ$ . This same method also allows interpolation of the orbit averaged rate of change of energy and angular momentum from the asymptotic fluxes for all inclinations (both prograde and retrograde) with an accuracy of  $\sim 10^{-6}$  using only 342 points. This is a significant improvement over other interpolation methods found in

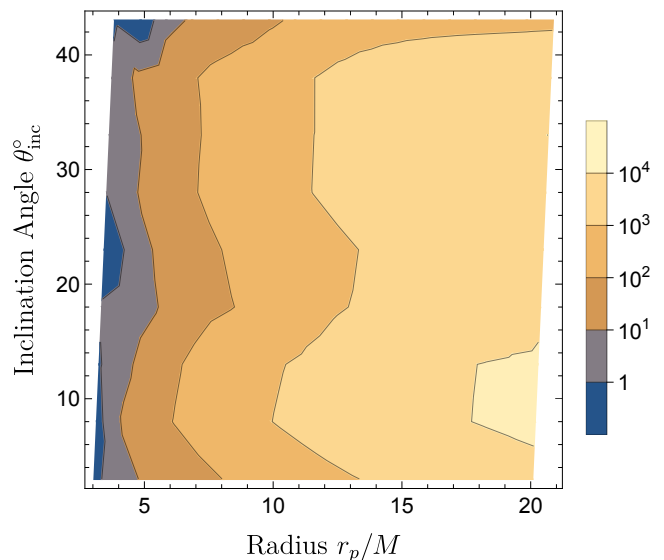


FIG. 8. The absolute difference in  $\varphi_\phi$  for a NIT inspiral evolved from a point in the parameter space to the ISSO with a typical EMRI mass ratio of  $\epsilon = 10^{-5}$ , depending on whether one incorporates high precision fluxes or not. Using the high precision fluxes results in a more faithful inspiral, with phase differences with respect to the lower accuracy model ranging from  $\sim 10 - 10^4$  radians. As such, interpolating the adiabatic contributions to the averaged equations of motion to high precision is vitally important for accurate adiabatic and post-adiabatic EMRI waveforms.

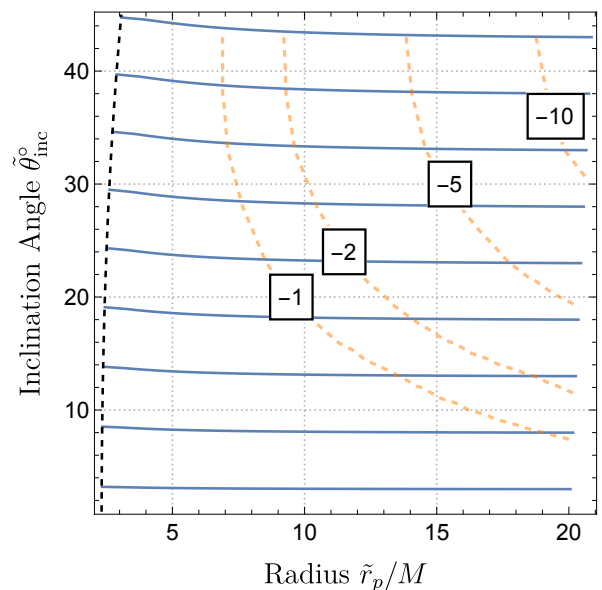


FIG. 9. The blue curves show the adiabatic trajectories through  $\{r_p, \theta_{\text{inc}}\}$  space, while the orange contours denote the final value of  $\varphi_\phi^{(1)}$  when evolved from that point in the parameter space to the ISSO.

the literature which require at least an order of magnitude more points to achieve a comparable level of accuracy [26, 27]. This could be further improved with a better choice when rescaling the data before interpolation, ideally choosing a function informed by the leading order PN contributions in the weak fields and/or the analytic structure of the GSF near the ISSO. So far this model is only valid for orbits where the primary’s spin is  $a = 0.9M$  and so interpolating over the other values of spin is left for future work. However this work, along with paper II and Ref. [29], shows that the Chebyshev interpolation methods are a promising approach for interpolating information from expensive GSF and flux codes across the vast 4-dimensional parameter space of generic Kerr inspirals.

Using our interpolated GSF model with the osculating geodesics (OG) formulation outlined in Ref. [31] and paper II along with the spherical limit derived in Appendix A, allows us to calculate the first ever quasi-spherical self-forced inspirals around a Kerr black hole. Our Mathematica implementation of this will be made publicly available on the Black Hole Perturbation Toolkit [40]. However, for a binary with a small mass ratio  $\epsilon$ , numerically evolving these inspirals can take minutes to hours due to the need to resolve the  $\sim 1/\epsilon$  orbital oscillations.

We overcome this by employing the technique of near-identity averaging transformations (NITs), as outlined in papers I and II, which produce equations of motion that capture the correct long-term secular evolution of the binary but can also be rapidly solved numerically. Following the methodology of Ref. [18], we improve upon this formulation by employing a second averaging transformation such that the solutions to our equations of motion are parametrized by Boyer-linquist coordinate time instead of Mino time. Since Boyer-linquist time can be related to the time at the detector, this improved NIT procedure is much more convenient for generating waveforms and for data analysis. We also employ a two-timescale expansion (TTE) of the NIT equations of motion which factors out the dependence of the mass ratio, at the cost of doubling the number of equations to be solved.

The quantities calculated from either the NIT or TTE equations of motion remain close to the OG evolution variables throughout the inspiral to the expected order in the mass-ratio, scaling similarly as the error expected from omitted higher order post-adiabatic corrections. These averaging schemes work particularly well for quasi-spherical inspirals as we find the mismatch between waveforms calculated from NIT or TTE inspirals and waveforms calculated from OG inspirals is  $\leq 10^{-3}$ , even for binaries with  $\epsilon \sim 10^{-1}$ .

With our efficient inspiral trajectories, we investigate the effect of first order self-force on inspirals across the spherical Kerr parameter space. We find that the orbital separation decreases with time while the orbit becomes more inclined over time, which is consistent with the findings of adiabatic evolutions of quasi-spherical Kerr inspi-

ral. We also find that neglecting the conservative effects can result in an  $\mathcal{O}(\epsilon^0)$  radian de-phasing in the azimuthal phase, with this effect becoming larger with the inclination angle of the orbit.

We note that both NIT and TTE equations of motion allow us to further improve the accuracy of our waveforms by replacing the adiabatic terms with higher accuracy interpolants calculated from the asymptotic energy and angular momentum fluxes. Making this improvement can result in phase differences ranging from tens to tens of thousands of radians for multi-year long EMRIs. This highlights the necessity of efficient gravitational wave flux codes for both adiabatic and post-adiabatic EMRI waveforms [14].

This work only incorporates first order GSF results. Post-adiabatic waveforms will require second order results not only to reach  $\mathcal{O}(\epsilon^0)$  accuracy in the phases [19], but also, as was noted in paper II, to produce gauge invariant waveforms [28]. As such, the inspirals and waveforms produced here are only representative of the outgoing radiation gauge. While second order results are not currently available for Kerr orbits, our framework can accommodate them when they become available.

We are also missing any post-adiabatic contributions due to the spin of the secondary [95–101]. Incorporating the linear in spin contribution to the energy and angular momentum fluxes into our averaged equations should be straightforward. However, incorporating the conservative effects from the Mathisson-Papapetrou-Dixon equations will require careful consideration and will be the subject of future work.

We plan to extend this framework to the case of generic Kerr inspirals, but there are two challenges standing in our way. The first of which is the computational cost of generic Kerr GSF code coupled with the 4-dimensional parameter space makes interpolating a GSF model impractical for now. The second challenge is the presence of transient self-force resonances were our NIT equations are formally singular which will force us to apply an alternate averaging procedure in their vicinity [102]. There has been a lot of work covering the effects of transient resonances on EMRI trajectories due to the self-force [103–109] or an external third body [110, 111]. However, evolving through self-force resonances while incorporating all self-force effects and retaining the subradian accuracy requirement remains an open problem.

Finally, we note we have used the semi-relativistic quadrupole formula to generate the waveforms from the OG, NIT and TTE inspirals. This is sufficient for this work as we only wish to compare the difference in the waveforms caused by different inspiral calculations. However LISA data analysis will require fully-relativistic waveform amplitudes such as those currently in the FastEMRIWaveforms (FEW) package for Schwarzschild inspirals [83]. Currently FEW only uses adiabatic inspirals, but this can be improved by employing either our NIT or TTE equations of motion. Once the waveform amplitudes have been interpolated for Kerr inspi-

rals, they can be combined immediately with the implementation presented in this work.

### ACKNOWLEDGMENTS

We thank Scott Hughes for sharing gravitational flux data for spherical orbits with us, and thank Lisa Drummond for a careful review of our osculating orbit evolution code. We also thank Ian Hinder and Barry Wardell for the SimulationTools analysis package. This work makes use of the Black Hole Perturbation Toolkit. PL acknowledges support from the Irish Research Council under grant GOIPG/2018/1978. NW acknowledges sup-

port from a Royal Society–Science Foundation Ireland Research Fellowship. MvdM acknowledges financial support by the VILLUM Foundation (grant no. VIL37766) and the DNRF Chair program (grant no. DNRF162) by the Danish National Research Foundation. This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant numbers 16/RS-URF/3428 and 17/RS-URF-RG/3490. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. The numerical calculation of the spherical orbit GSF were obtained using the IRIDIS High Performance Computing Facility at the University of Southampton.

- 
- [1] J. Baker *et al.*, The Laser Interferometer Space Antenna: Unveiling the Millihertz Gravitational Wave Sky, (2019), arXiv:1907.06482 [astro-ph.IM].
- [2] J. Mei *et al.*, The TianQin project: Current progress on science and technology, Progress of Theoretical and Experimental Physics **2021**, 10.1093/ptep/ptaa114 (2020), 05A107, <https://academic.oup.com/ptep/article-pdf/2021/5/05A107/37953035/ptaa114.pdf>.
- [3] Z. Luo, Z. Guo, G. Jin, Y. Wu, and W. Hu, A brief analysis to taiji: Science and technology, Results in Physics **16**, 102918 (2020).
- [4] C. P. L. Berry, S. A. Hughes, C. F. Sopuerta, A. J. K. Chua, A. Heffernan, K. Holley-Bockelmann, D. P. Mihaylov, M. C. Miller, and A. Sesana, The unique potential of extreme mass-ratio inspirals for gravitational-wave astronomy, (2019), arXiv:1903.03686.
- [5] J. R. Gair, M. Vallisneri, S. L. Larson, and J. G. Baker, Testing General Relativity with Low-Frequency, Space-Based Gravitational-Wave Detectors, Living Rev. Rel. **16**, 7 (2013), arXiv:1212.5575 [gr-qc].
- [6] S. Babak, J. R. Gair, and R. H. Cole, Extreme mass ratio inspirals: perspectives for their detection, Fund. Theor. Phys. **179**, 783 (2015), arXiv:1411.5253 [gr-qc].
- [7] C. Hopman, BINARY DYNAMICS NEAR a MASSIVE BLACK HOLE, The Astrophysical Journal **700**, 1933 (2009).
- [8] M. Coleman Miller, M. Freitag, D. P. Hamilton, and V. M. Lauburg, Binary encounters with supermassive black holes: Zero-eccentricity LISA events, Astrophys. J. Lett. **631**, L117 (2005), arXiv:astro-ph/0507133.
- [9] K. Glampedakis and D. Kennefick, Zoom and whirl: Eccentric equatorial orbits around spinning black holes and their evolution under gravitational radiation reaction, Phys. Rev. D **66**, 10.1103/PhysRevD.66.044002 (2002), arXiv:0203086 [gr-qc].
- [10] C. M. Will, Compact binary inspiral: Nature is perfectly happy with a circle, Class. Quant. Grav. **36**, 195013 (2019), arXiv:1906.08064 [gr-qc].
- [11] F. D. Ryan, Effect of gravitational radiation reaction on nonequatorial orbits around a Kerr black hole, Phys. Rev. D **53**, 3064 (1996), arXiv:gr-qc/9511062.
- [12] S. A. Hughes, Evolution of circular, nonequatorial orbits of Kerr black holes due to gravitational-wave emission, Phys. Rev. D - Part. Fields, Gravit. Cosmol. **61**, 084004 (2000), arXiv:9910091 [gr-qc].
- [13] S. A. Hughes, Evolution of circular, nonequatorial orbits of Kerr black holes due to gravitational-wave emission. II. Inspiral trajectories and gravitational waveforms, Phys. Rev. D **64**, 15 (2001), arXiv:0104041 [gr-qc].
- [14] S. A. Hughes, N. Warburton, G. Khanna, A. J. Chua, and M. L. Katz, Adiabatic waveforms for extreme mass-ratio inspirals via multivoice decomposition in time and frequency, Phys. Rev. D **103**, 10.1103/PhysRevD.103.104014 (2021), arXiv:2102.02713.
- [15] P. Amaro-Seoane, B. F. Schutz, and J. Thornburg, Gravitational Waves Notes, Issue #5 : 'The Capra research programme for capture of small compact objects by massive black holes', (2011), arXiv:1102.3647 [astro-ph.CO].
- [16] L. Barack and A. Pound, Self-force and radiation reaction in general relativity, Rept. Prog. Phys. **82**, 016904 (2019), arXiv:1805.10385 [gr-qc].
- [17] E. Poisson, A. Pound, and I. Vega, The Motion of point particles in curved spacetime, Living Rev. Rel. **14**, 7 (2011), arXiv:1102.0529 [gr-qc].
- [18] A. Pound and B. Wardell, Black hole perturbation theory and gravitational self-force, (2021), arXiv:2101.04592.
- [19] T. Hinderer and É. É. Flanagan, Two-timescale analysis of extreme mass ratio inspirals in Kerr spacetime: Orbital motion, Phys. Rev. D - Part. Fields, Gravit. Cosmol. **78**, 10.1103/PhysRevD.78.064028 (2008), arXiv:0805.3337.
- [20] A. Pound, B. Wardell, N. Warburton, and J. Miller, Second-Order Self-Force Calculation of Gravitational Binding Energy in Compact Binaries, Phys. Rev. Lett. **124**, 10.1103/PhysRevLett.124.021101 (2020), arXiv:1908.07419.
- [21] N. Warburton, A. Pound, B. Wardell, J. Miller, and L. Durkan, Gravitational-Wave Energy Flux for Compact Binaries through Second Order in the Mass Ratio, Phys. Rev. Lett. **127**, 151102 (2021), arXiv:2107.01298 [gr-qc].
- [22] L. Durkan and N. Warburton, Slow evolution of the metric perturbation due to a quasicircular inspiral into a Schwarzschild black hole, Phys. Rev. D **106**, 084023

- (2022), arXiv:2206.08179 [gr-qc].
- [23] B. Wardell, A. Pound, N. Warburton, J. Miller, L. Durkan, and A. Le Tiec, Gravitational waveforms for compact binaries from second-order self-force theory, (2021), arXiv:2112.12265 [gr-qc].
- [24] M. van de Meent, Gravitational self-force on eccentric equatorial orbits around a Kerr black hole, *Phys. Rev. D* **94**, 044034 (2016), arXiv:1606.06297 [gr-qc].
- [25] M. van de Meent, Gravitational self-force on generic bound geodesics in Kerr spacetime, *Phys. Rev. D* **97**, 104033 (2018), arXiv:1711.09607 [gr-qc].
- [26] N. Warburton, S. Akcay, L. Barack, J. R. Gair, and N. Sago, Evolution of inspiral orbits around a Schwarzschild black hole, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **85**, 10.1103/PhysRevD.85.061501 (2012), arXiv:1111.6908.
- [27] T. Osburn, N. Warburton, and C. R. Evans, Highly eccentric inspirals into a black hole, *Phys. Rev. D* **93**, 10.1103/PhysRevD.93.064024 (2016), arXiv:1511.01498.
- [28] P. Lynch, M. van de Meent, and N. Warburton, Eccentric self-forced inspirals into a rotating black hole, *Class. Quant. Grav.* **39**, 145004 (2022), arXiv:2112.05651 [gr-qc].
- [29] V. Skoupý and G. Lukes-Gerakopoulos, Adiabatic equatorial inspirals of a spinning body into a Kerr black hole, *Phys. Rev. D* **105**, 084033 (2022), arXiv:2201.07044 [gr-qc].
- [30] A. Pound and E. Poisson, Osculating orbits in Schwarzschild spacetime, with an application to extreme mass-ratio inspirals, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **77**, 10.1103/PhysRevD.77.044013 (2008), arXiv:0708.3033.
- [31] J. R. Gair, É. I. Flanagan, S. Drasco, T. Hinderer, and S. Babak, Forced motion near black holes, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **83**, 10.1103/PhysRevD.83.044037 (2011), arXiv:1012.5111.
- [32] M. van de Meent and N. Warburton, Fast Self-forced Inspirals, *Class. Quant. Grav.* **35**, 144003 (2018), arXiv:1802.05281 [gr-qc].
- [33] J. Miller and A. Pound, Two-timescale evolution of extreme-mass-ratio inspirals: Waveform generation scheme for quasicircular orbits in Schwarzschild spacetime, *Phys. Rev. D* **103**, 10.1103/PhysRevD.103.064048 (2021), arXiv:2006.11263.
- [34] J. Keavorkian, *Perturbation Techniques for Oscillatory Systems With Slowly Varying Coefficients.*, *SIAM Rev.* **29**, 391 (1987).
- [35] O. Burke, *Extreme precision and extreme complexity: source modelling and data analysis development for the laser interferometer space antenna*, Ph.D. thesis, University of Edinburgh (2021).
- [36] S. Isoyama, R. Fujita, A. J. K. Chua, H. Nakano, A. Pound, and N. Sago, Adiabatic Waveforms from Extreme-Mass-Ratio Inspirals: An Analytical Approach, *Phys. Rev. Lett.* **128**, 231101 (2022), arXiv:2111.05288 [gr-qc].
- [37] B. Carter, Global structure of the kerr family of gravitational fields, *Phys. Rev.* **174**, 1559 (1968).
- [38] S. Drasco and S. A. Hughes, Rotating black hole orbit functionals in the frequency domain, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **69**, 10.1103/PhysRevD.69.044015 (2004), arXiv:0308479 [astro-ph].
- [39] R. Fujita and W. Hikida, Analytical solutions of bound timelike geodesic orbits in Kerr spacetime, *Class. Quantum Gravity* **26**, 10.1088/0264-9381/26/13/135002 (2009), arXiv:0906.1420.
- [40] Black Hole Perturbation Toolkit, (bhptoolkit.org).
- [41] F. D. Ryan, Effect of gravitational radiation reaction on circular orbits around a spinning black hole, *Phys. Rev. D* **52**, R3159 (1995), arXiv:gr-qc/9506023.
- [42] L. C. Stein and N. Warburton, Location of the last stable orbit in Kerr spacetime, *Phys. Rev. D* **101**, 10.1103/PhysRevD.101.064007 (2020), arXiv:1912.07609.
- [43] W. Schmidt, *Class. Quantum Gravity*, Tech. Rep. 10 (2002) arXiv:0202090 [gr-qc].
- [44] M. van de Meent, Analytic solutions for parallel transport along generic bound geodesics in Kerr spacetime, *Class. Quant. Grav.* **37**, 145007 (2020), arXiv:1906.05090 [gr-qc].
- [45] S. Drasco and S. A. Hughes, Rotating black hole orbit functionals in the frequency domain, *Phys. Rev. D* **69**, 044015 (2004), arXiv:astro-ph/0308479.
- [46] Z. Nasipak, T. Osburn, and C. R. Evans, Repeated faint quasinormal bursts in extreme-mass-ratio inspiral waveforms: Evidence from frequency-domain scalar self-force calculations on generic Kerr orbits, *Phys. Rev. D* **100**, 064008 (2019), arXiv:1905.13237 [gr-qc].
- [47] D. Kennefick and A. Ori, Radiation reaction induced evolution of circular orbits of particles around Kerr black holes, *Phys. Rev. D* **53**, 4319 (1996), arXiv:gr-qc/9512018.
- [48] A. G. Shah, J. L. Friedman, and T. S. Keidl, EMRI corrections to the angular velocity and redshift factor of a mass in circular orbit about a Kerr black hole, *Phys. Rev. D* **86**, 084059 (2012), arXiv:1207.5595 [gr-qc].
- [49] A. Pound, C. Merlin, and L. Barack, Gravitational self-force from radiation-gauge metric perturbations, *Phys. Rev. D* **89**, 024009 (2014), arXiv:1310.1513 [gr-qc].
- [50] M. van de Meent and A. G. Shah, Metric perturbations produced by eccentric equatorial orbits around a Kerr black hole, *Phys. Rev. D* **92**, 064025 (2015), arXiv:1506.04755 [gr-qc].
- [51] N. Warburton, Self force on a scalar charge in Kerr spacetime: inclined circular orbits, *Phys. Rev. D* **91**, 024045 (2015), arXiv:1408.2885 [gr-qc].
- [52] R. Fujita and H. Tagoshi, New numerical methods to evaluate homogeneous solutions of the Teukolsky equation, *Prog. Theor. Phys.* **112**, 415 (2004), arXiv:gr-qc/0410018.
- [53] R. Fujita, W. Hikida, and H. Tagoshi, An Efficient Numerical Method for Computing Gravitational Waves Induced by a Particle Moving on Eccentric Inclined Orbits around a Kerr Black Hole, *Prog. Theor. Phys.* **121**, 843 (2009), arXiv:0904.3810 [gr-qc].
- [54] S. Mano, H. Suzuki, and E. Takasugi, Analytic solutions of the Teukolsky equation and their low frequency expansions, *Prog. Theor. Phys.* **95**, 1079 (1996), arXiv:gr-qc/9603020.
- [55] S. Mano and E. Takasugi, Analytic solutions of the Teukolsky equation and their properties, *Prog. Theor. Phys.* **97**, 213 (1997), arXiv:gr-qc/9611014.
- [56] A. Ori, Reconstruction of inhomogeneous metric perturbations and electromagnetic four potential in Kerr space-time, *Phys. Rev. D* **67**, 124010 (2003), arXiv:gr-qc/0207045.

- [57] P. L. Chrzanowski, Vector Potential and Metric Perturbations of a Rotating Black Hole, *Phys. Rev. D* **11**, 2042 (1975).
- [58] L. S. Kegeles and J. M. Cohen, CONSTRUCTIVE PROCEDURE FOR PERTURBATIONS OF SPACETIMES, *Phys. Rev. D* **19**, 1641 (1979).
- [59] R. M. Wald, Construction of Solutions of Gravitational, Electromagnetic, Or Other Perturbation Equations from Solutions of Decoupled Equations, *Phys. Rev. Lett.* **41**, 203 (1978).
- [60] L. Barack and A. Ori, Mode sum regularization approach for the selfforce in black hole space-time, *Phys. Rev. D* **61**, 061502 (2000), arXiv:gr-qc/9912010.
- [61] L. Barack, Y. Mino, H. Nakano, A. Ori, and M. Sasaki, Calculating the gravitational selfforce in Schwarzschild space-time, *Phys. Rev. Lett.* **88**, 091101 (2002), arXiv:gr-qc/0111001.
- [62] L. Barack and A. Ori, Gravitational selfforce on a particle orbiting a Kerr black hole, *Phys. Rev. Lett.* **90**, 111101 (2003), arXiv:gr-qc/0212103.
- [63] L. Barack, Gravitational self force in extreme mass-ratio inspirals, *Class. Quant. Grav.* **26**, 213001 (2009), arXiv:0908.1664 [gr-qc].
- [64] C. Merlin, A. Ori, L. Barack, A. Pound, and M. van de Meent, Completion of metric reconstruction for a particle orbiting a Kerr black hole, *Phys. Rev. D* **94**, 104066 (2016), arXiv:1609.01227 [gr-qc].
- [65] M. van de Meent, The mass and angular momentum of reconstructed metric perturbations, *Class. Quant. Grav.* **34**, 124003 (2017), arXiv:1702.00969 [gr-qc].
- [66] V. Toomani, P. Zimmerman, A. Spiers, S. Hollands, A. Pound, and S. R. Green, New metric reconstruction scheme for gravitational self-force calculations, *Class. Quant. Grav.* **39**, 015019 (2022), arXiv:2108.04273 [gr-qc].
- [67] M. van de Meent and A. Pound, Gauge completion in radiation gauge metric reconstruction, (unpublished).
- [68] S. A. Hughes, S. Drasco, E. E. Flanagan, and J. Franklin, Gravitational radiation reaction and inspiral waveforms in the adiabatic limit, *Phys. Rev. Lett.* **94**, 10.1103/PhysRevLett.94.221101 (2005).
- [69] R. Fujita and M. Shibata, Extreme mass ratio inspirals on the equatorial plane in the adiabatic order, *Phys. Rev. D* **102**, 064005 (2020), arXiv:2008.13554 [gr-qc].
- [70] L. Barack and N. Sago, Gravitational self-force correction to the innermost stable circular orbit of a schwarzschild black hole, *Phys. Rev. Lett.* **102**, 10.1103/PhysRevLett.102.191101 (2009), arXiv:0902.0573.
- [71] L. Barack, T. Damour, and N. Sago, Precession effect of the gravitational self-force in a Schwarzschild space-time and the effective one-body formalism, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **82**, 10.1103/PhysRevD.82.084036 (2010), arXiv:1008.0935.
- [72] L. Barack and N. Sago, Beyond the geodesic approximation: Conservative effects of the gravitational self-force in eccentric orbits around a Schwarzschild black hole, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **83**, 10.1103/PhysRevD.83.084023 (2011), arXiv:1101.3331.
- [73] N. Warburton and L. Barack, Self-force on a scalar charge in Kerr spacetime: Eccentric equatorial orbits, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **83**, 10.1103/PhysRevD.83.124038 (2011), arXiv:1103.0287.
- [74] M. van de Meent, Self-force corrections to the periapsis advance around a spinning black hole, *Phys. Rev. Lett.* **118**, 011101 (2017), arXiv:1610.03497 [gr-qc].
- [75] J. Vines and E. E. Flanagan, Is motion under the conservative self-force in black hole spacetimes an integrable Hamiltonian system?, *Phys. Rev. D* **92**, 064039 (2015), arXiv:1503.04727 [gr-qc].
- [76] R. Fujita, S. Isoyama, A. Le Tiec, H. Nakano, N. Sago, and T. Tanaka, Hamiltonian Formulation of the Conservative Self-Force Dynamics in the Kerr Geometry, *Class. Quant. Grav.* **34**, 134001 (2017), arXiv:1612.02504 [gr-qc].
- [77] Y. Mino, Perturbative approach to an orbital evolution around a supermassive black hole, *Phys. Rev. D* **67**, 084027 (2003), arXiv:gr-qc/0302075.
- [78] Y. Mino, Self-force in the radiation reaction formula, *Prog. Theor. Phys.* **113**, 733 (2005), arXiv:gr-qc/0506003.
- [79] Y. Mino, From the self-force problem to the radiation reaction formula, *Class. Quant. Grav.* **22**, S717 (2005), arXiv:gr-qc/0506002.
- [80] T. C. Quinn and R. M. Wald, Energy conservation for point particles undergoing radiation reaction, *Phys. Rev. D* **60**, 064009 (1999), arXiv:gr-qc/9903014.
- [81] D. V. Gal'tsov, Radiation reaction in the Kerr gravitational field, *J. Phys. A* **15**, 3737 (1982).
- [82] J. McCart, T. Osburn, and J. Y. J. Burton, Highly eccentric extreme-mass-ratio-inspiral waveforms via fast self-forced inspirals, *Phys. Rev. D* **104**, 084050 (2021), arXiv:2109.00056 [gr-qc].
- [83] A. J. Chua, M. L. Katz, N. Warburton, and S. A. Hughes, Rapid Generation of Fully Relativistic Extreme-Mass-Ratio-Inspiral Waveform Templates for LISA Data Analysis, *Phys. Rev. Lett.* **126**, 10.1103/PhysRevLett.126.051102 (2021), arXiv:2008.06071.
- [84] M. L. Katz, A. J. K. Chua, L. Speri, N. Warburton, and S. A. Hughes, Fast extreme-mass-ratio-inspiral waveforms: New tools for millihertz gravitational-wave data analysis, *Phys. Rev. D* **104**, 064047 (2021), arXiv:2104.04582 [gr-qc].
- [85] S. Babak, H. Fang, J. R. Gair, K. Glampedakis, and S. A. Hughes, "kludge" gravitational waveforms for a test-body orbiting a Kerr black hole, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **75**, 10.1103/PhysRevD.75.024005 (2007), arXiv:0607007 [gr-qc].
- [86] SimulationTools, (simulationtools.org).
- [87] K. S. Thorne, Transition from inspiral to plunge for a compact body in a circular equatorial orbit around a massive, spinning black hole, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **62**, 8 (2000), arXiv:0003032 [gr-qc].
- [88] O. Burke, J. Gair, and J. Simón, Transition from inspiral to plunge: A complete near-extremal trajectory and associated waveform, *Phys. Rev. D* **101**, 10.1103/PhysRevD.101.064026 (2020), arXiv:1909.12846.
- [89] G. Compère and L. Kuchler, Self-Consistent Adiabatic Inspiral and Transition Motion, *Phys. Rev. Lett.* **126**, 10.1103/PhysRevLett.126.241106 (2021), arXiv:2102.12747.
- [90] G. Compère and L. Kuchler, Asymptotically matched quasi-circular inspiral and transition-to-plunge in the small mass ratio expansion, *SciPost Phys.* **13**, 043 (2022), arXiv:2112.02114 [gr-qc].
- [91] S. A. Hughes, Measuring gravitational waves from bi-

- nary black hole coalescences. I. Signal to noise for inspiral, merger, and ringdown, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **57**, 4535 (1998), arXiv:9701039 [gr-qc].
- [92] L. Lindblom, B. J. Owen, and D. A. Brown, Model waveform accuracy standards for gravitational wave data analysis, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **78**, 10.1103/PhysRevD.78.124020 (2008), arXiv:0809.3844.
- [93] S. T. McWilliams, B. J. Kelly, and J. G. Baker, Observing mergers of nonspinning black-hole binaries, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **82**, 10.1103/PhysRevD.82.024014 (2010), arXiv:1004.0961.
- [94] B. J. Owen, Search templates for gravitational waves from inspiraling binaries: Choice of template spacing, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **53**, 6749 (1996), arXiv:9511032 [gr-qc].
- [95] V. Skoupý and G. Lukes-Gerakopoulos, Spinning test body orbiting around a Kerr black hole: Eccentric equatorial orbits and their asymptotic gravitational-wave fluxes, *Phys. Rev. D* **103**, 104045 (2021), arXiv:2102.04819 [gr-qc].
- [96] G. A. Piovano, A. Maselli, and P. Pani, Extreme mass ratio inspirals with spinning secondary: a detailed study of equatorial circular motion, *Phys. Rev. D* **102**, 024041 (2020), arXiv:2004.02654 [gr-qc].
- [97] G. A. Piovano, R. Brito, A. Maselli, and P. Pani, Assessing the detectability of the secondary spin in extreme mass-ratio inspirals with fully relativistic numerical waveforms, *Phys. Rev. D* **104**, 124019 (2021), arXiv:2105.07083 [gr-qc].
- [98] J. Mathews, A. Pound, and B. Wardell, Self-force calculations with a spinning secondary, *Phys. Rev. D* **105**, 084031 (2022), arXiv:2112.13069 [gr-qc].
- [99] L. V. Drummond and S. A. Hughes, Precisely computing bound orbits of spinning bodies around black holes. I. General framework and results for nearly equatorial orbits, *Phys. Rev. D* **105**, 124040 (2022), arXiv:2201.13334 [gr-qc].
- [100] L. V. Drummond and S. A. Hughes, Precisely computing bound orbits of spinning bodies around black holes. II. Generic orbits, *Phys. Rev. D* **105**, 124041 (2022), arXiv:2201.13335 [gr-qc].
- [101] V. Skoupý, G. Lukes-Gerakopoulos, L. V. Drummond, and S. A. Hughes, Asymptotic gravitational-wave fluxes from a spinning test body on generic orbits around a Kerr black hole, (2023), arXiv:2303.16798 [gr-qc].
- [102] G. Lukes-Gerakopoulos and V. Witzany, Nonlinear Effects in EMRI Dynamics and Their Imprints on Gravitational Waves, *Handb. Gravitational Wave Astron.*, 1 (2021), arXiv:2103.06724.
- [103] I. I. Flanagan and T. Hinderer, Transient resonances in the inspirals of point particles into black holes, *Phys. Rev. Lett.* **109**, 10.1103/PhysRevLett.109.071102 (2012).
- [104] É. E. Flanagan, S. A. Hughes, and U. Ruangsri, Resonantly enhanced and diminished strong-field gravitational-wave fluxes, *Phys. Rev. D - Part. Fields, Gravit. Cosmol.* **89**, 10.1103/PhysRevD.89.084028 (2014), arXiv:1208.3906.
- [105] M. van de Meent, Conditions for Sustained Orbital Resonances in Extreme Mass Ratio Inspirals, *Phys. Rev. D* **89**, 084033 (2014), arXiv:1311.4457 [gr-qc].
- [106] C. P. Berry, R. H. Cole, P. Cañizares, and J. R. Gair, Importance of transient resonances in extreme-mass-ratio inspirals, *Phys. Rev. D* **94**, 10.1103/PhysRevD.94.124042 (2016).
- [107] D. P. Mihaylov and J. R. Gair, Transition of EMRIs through resonance: corrections to higher order in the on-resonance flux modification, *J. Math. Phys.* **58**, 112501 (2017), arXiv:1706.06639 [gr-qc].
- [108] L. Speri and J. R. Gair, Assessing the impact of transient orbital resonances, *Phys. Rev. D* **103**, 10.1103/PhysRevD.103.124032 (2021), arXiv:2103.06306.
- [109] Z. Nasipak and C. R. Evans, Resonant self-force effects in extreme-mass-ratio binaries: A scalar model, *Phys. Rev. D* **104**, 084011 (2021), arXiv:2105.15188 [gr-qc].
- [110] P. Gupta, B. Bonga, A. J. K. Chua, and T. Tanaka, Importance of tidal resonances in extreme-mass-ratio inspirals, *Phys. Rev. D* **104**, 044056 (2021), arXiv:2104.03422 [gr-qc].
- [111] P. Gupta, L. Speri, B. Bonga, A. J. K. Chua, and T. Tanaka, Modeling transient resonances in extreme-mass-ratio inspirals, *Phys. Rev. D* **106**, 104001 (2022), arXiv:2205.04808 [gr-qc].
- [112] S. Drasco and S. A. Hughes, Gravitational wave snapshots of generic extreme mass ratio inspirals, *Phys. Rev. D* **73**, 024027 (2006), [Erratum: *Phys. Rev. D* **88**, 109905 (2013), Erratum: *Phys. Rev. D* **90**, 109905 (2014)], arXiv:gr-qc/0509101.

### Appendix A: Osculating geodesic equations in the spherical orbit limit

In this appendix we derive the evolution equation for the orbital radius  $r_p$  in the spherical limit due to the presence of a perturbing acceleration  $\{a_t, a_r, a_z, a_\phi\}$ . Using the chain rule, we obtain

$$\frac{dr_p}{d\lambda} = \frac{\partial r_p}{\partial \mathcal{J}_j} \frac{d\mathcal{J}_j}{d\lambda} \quad (\text{A1})$$

Recalling the expressions for the evolution equations for the integrals of motion  $\vec{\mathcal{J}} = \{\mathcal{E}, \mathcal{L}_z, \mathcal{K}\}$  given in Appendix B of paper II,

$$\frac{d\mathcal{E}}{d\lambda} = -\Sigma_p a_t, \quad \frac{d\mathcal{L}_z}{d\lambda} = \Sigma_p a_\phi, \quad (\text{A2a-b})$$

$$\frac{d\mathcal{K}}{d\lambda} = -\frac{2\Sigma_p}{\Delta_p} (\varpi_p^2 \mathcal{B}_p a_t + a \mathcal{B}_p a_\phi + \Delta_p^2 u_r a_r). \quad (\text{A2c})$$

To find the partial derivatives  $\partial r_p / \partial \mathcal{J}_j$ , make use of the value of radial potential  $V_r$  defined as

$$\begin{aligned} V_r(r) &= \mathcal{B}^2 - \Delta (r^2 + \mathcal{K}) \\ &= -(1 - \mathcal{E}^2)(r - r_1)(r - r_2)(r - r_3)(r - r_4) \quad (\text{A3}) \\ &= -(1 - \mathcal{E}^2)(r - r_p)^2(r - r_3)(r - r_4). \end{aligned}$$

where  $r_1, r_2, r_3, r_4$  are the roots of  $V_r$  and in the spherical case:  $r_1 = r_2 = r_p$ . We then define the derivative of  $V_r$

with respect to  $r$  as

$$\begin{aligned}\kappa &:= \frac{dV_r}{dr} \\ &= 4\mathcal{E}\mathcal{B}r - 2\Delta r - 2(r - M)(r^2 + K) \\ &= -2(1 - \mathcal{E}^2) \left[ (r - r_3)(r - r_4) \right. \\ &\quad \left. + (r - r_p)(r - r_4) + (r - r_p)(r - r_3) \right].\end{aligned}\quad (\text{A4})$$

If we take the derivative of  $\kappa$  with respect to  $r$  and evaluate at  $r = r_p$ , we obtain

$$\frac{d\kappa_p}{dr} = -2(1 - \mathcal{E}^2)(r_p - r_3)(r_p - r_4). \quad (\text{A5})$$

However, if we now take the partial derivatives of  $\kappa$  with respect to  $\vec{\mathcal{J}}$  and evaluate at  $r = r_p$  we are left with

$$\begin{aligned}\frac{\partial \kappa_p}{\partial \mathcal{J}_j} &= 2(1 - \mathcal{E}^2)(r_p - r_3)(r_p - r_4) \frac{\partial r_p}{\partial \mathcal{J}_j} \\ &= -\frac{d\kappa_p}{dr} \frac{\partial r_p}{\partial \mathcal{J}_j}.\end{aligned}\quad (\text{A6})$$

Subbing the above result into Eq. (A1), explicitly calculating  $\partial \kappa / \partial \mathcal{J}_j$  using Eq. (A4), evaluating at  $r = r_p$  and simplifying gives us our final result

$$\begin{aligned}\frac{dr_p}{d\lambda} &= -4r_p \frac{(\mathcal{B}_p + \varpi_p^2 \mathcal{E}) \frac{dE}{d\lambda} - a\mathcal{E} \frac{d\mathcal{L}_z}{d\lambda} + \frac{M - r_p}{2r_p} \frac{dK}{d\lambda}}{\frac{d\kappa_p}{dr}} \\ &:= F_r^{(1)}\end{aligned}\quad (\text{A7})$$

The above relationship is finite for all bound stable spherical orbits and only becomes ill-defined when  $d\kappa_p/dr = 0$ , which, only occurs when  $r_p = r_3$ , i.e., when  $r_p = r_{\text{ISSO}}$ .

### Appendix B: Equivalence between $\vec{\varphi}$ and waveform voices

In this appendix, we wish to establish a relationship between the solutions for the inspiral quantities and the waveform. We start by assigning the origin of our coordinate scheme to be the position of the primary MBH. We then set the Cartesian coordinates of our observer to be  $x = (t_{\text{ret}}, \vec{x})$ . It is useful to express the observer's coordinates in terms of spherical coordinates  $(R, \theta, \Phi)$ , where  $R$  is the distance from the observer to origin given by  $R^2 = \vec{x} \cdot \vec{x}$ ,  $\Theta$  is the observer's latitude and  $\Phi$  is the observer's azimuth. We assign the Cartesian coordinates of the secondary, which we model as a point particle, to be  $x_p = (t_p, \vec{x}_p)$ . As such, we can express the retarded time as measured by the observer in terms of Boyer-Lindquist coordinate time  $t$  via  $t_{\text{ret}} = t - |\vec{x} - \vec{x}_p| \approx t - R$  in the limit where  $R$  is large. The complex waveform strain can be decomposed onto a basis of spin-weighted spherical harmonics  ${}_{-2}Y_{lm}$  given by:

$$h(t_{\text{ret}}) = h_+ - ih_\times = \frac{1}{R} \sum_{lm} H_{lm}(t_{\text{ret}}) {}_{-2}Y_{lm}(\Theta, \Phi). \quad (\text{B1})$$

If we assume the secondary is moving on a geodesic, the waveform modes  $H_{lm}$  exhibit a discrete frequency spectrum and so can be Fourier decomposed into:

$$H_{lm}(t_{\text{ret}}) = \sum_{(n,k)} A_{lmnk}(\vec{P}) e^{-i(m\Omega_\phi^{(0)} + n\Omega_r^{(0)} + k\Omega_z^{(0)})t_{\text{ret}}} \quad (\text{B2})$$

where the complex amplitudes  $A_{lmnk}$  can be related to non-homogeneous Teukolsky amplitudes, which can be precomputed for a given set of orbital elements  $\vec{P}$  [14]. Such waveforms are known as ‘‘snapshot’’ waveforms, as they only capture a small section of the total waveform [112].

For a full EMRI waveform, one needs to account for the fact that the frequencies and the orbital elements will slowly evolve with time, resulting in a continuous frequency spectrum. As such, this Fourier mode decomposition becomes a ‘‘multi-voice’’ decomposition [14]:

$$H_{lm}(t_{\text{ret}}) = \sum_{(n,k)} A_{lmnk}(\vec{P}(t_{\text{ret}})) e^{-i\Phi_{mnk}(t_{\text{ret}})}, \quad (\text{B3})$$

where the waveform ‘‘voices’’  $\Phi_{mnk}$  are given by

$$\Phi_{mnk} = m \int_0^t \Omega_\phi(t) dt + n \int_0^t \Omega_r(t) dt + k \int_0^t \Omega_z(t) dt. \quad (\text{B4})$$

Let us recall the equations of motion for the orbital phases obtains after performing the NIT and the additional transformation such that our solutions are in terms of Boyer-Lindquist time (46). When we express this in integral form, one obtains

$$\begin{aligned}\varphi_\alpha &= \int_0^t \left( \Omega_\alpha^{(0)}(t) + \epsilon \Omega_\alpha^{(1)}(t) + \mathcal{O}(\epsilon^2) \right) dt \\ &= \int_0^t \Omega_\alpha(t) dt + \mathcal{O}(\epsilon)\end{aligned}\quad (\text{B5})$$

As such we can see that to leading order in  $\epsilon$ , these phases are directly related to the waveform voices:

$$\Phi_{mnk} = m\varphi_\phi + n\varphi_r + k\varphi_z + \mathcal{O}(\epsilon) \quad (\text{B6})$$

This is further supported by Ref. [82], where a relationship between the waveform phases and the NIT phases  $\vec{q}$  was independently derived. When expressed in our notation with polar motion included, this relationship is given by:

$$\begin{aligned}\Phi_{mnk} &= m\tilde{\varphi} + n\tilde{q}_r + k\tilde{q}_z \\ &\quad + (m\Omega_\phi^{(0)} + n\Omega_r^{(0)} + k\Omega_z^{(0)})(t - \tilde{t}) + \mathcal{O}(\epsilon)\end{aligned}\quad (\text{B7})$$

Using the relationship between  $t$  and  $\tilde{t}$  given by Eq. (42c), the result for  $\Delta t$  given by Eq. (49a), and the relationship between the NIT action angles given by Eq. (46b), one can obtain to leading order in  $\epsilon$ :



$$\begin{aligned}
\Phi_{mnk} &= m(\tilde{\phi} - \Omega_\phi^{(0)}(t - \tilde{t})) + n(\tilde{q}_r - \Omega_r^{(0)}(t - \tilde{t})) + k(\tilde{q}_z - \Omega_z(t - \tilde{t})) + \mathcal{O}(\epsilon) \\
&= m(\tilde{\phi} - \Omega_\phi^{(0)} Z_t^{(0)}) + n(\tilde{q}_r - \Omega_r^{(0)} Z_t^{(0)}) + k(\tilde{q}_z - \Omega_z^{(0)} Z_t^{(0)}) + \mathcal{O}(\epsilon) \\
&= m(\tilde{\phi} + \Omega_\phi^{(0)} \Delta t) + n(\tilde{q}_r + \Omega_r^{(0)} \Delta t) + k(\tilde{q}_z + \Omega_z^{(0)} \Delta t) + \mathcal{O}(\epsilon) \\
&= m(\tilde{\phi} + \Delta\varphi_\phi) + n(\tilde{q}_r + \Delta\varphi_r) + k(\tilde{q}_z + \Delta\varphi_z) + \mathcal{O}(\epsilon) \\
&= m\varphi_\phi + n\varphi_r + k\varphi_z + \mathcal{O}(\epsilon)
\end{aligned} \tag{B8}$$

One can also freely replace the dependence of the evolving orbital elements  $\vec{P}(t)$  with either  $\vec{P}(t)$  or  $\vec{\mathcal{P}}(t)$ , as at leading order in  $\epsilon$  they are identical. As a result, we can now relate a solution to the EMRI's inspiral trajectory to its associated Teukolsky based waveform.

### Appendix C: Gauge completion integrands

Here we give the explicit integrands for the gauge completion integrals (18) and (19),

$$\begin{aligned}
\mathcal{I}_\alpha &= \mathcal{E}^2 r_p (1 - z_p^2) \left[ 3r_p^{10} + 3r_p^9 (1 - z_p^2) - 6r_p^8 a^2 (1 + z_p^2 - 3z_p^4) - 2r_p^7 a^2 (5 + 4z_p^2 + 3z_p^4) \right. \\
&\quad - r_p^6 a^2 (3a^2 - (48 - 31a^2)z_p^2 + (24 - 47a^2)z_p^4 - 3a^2 z_p^6) - r_p^5 a^4 (5 - 9z_p^2 + 41z_p^4 + 3z_p^6) \\
&\quad - 2r_p^4 a^4 z_p^2 (2(1 + 2a^2) - (4 + 21a^2)z_p^2 - (a^2 + 12)z_p^4) + 2r_p^3 a^6 z_p^2 (11 - 16z_p^2 + z_p^4) \\
&\quad \left. - r_p^2 a^6 z_p^2 (10a^2 - 15a^2 z_p^2 + 5a^2 z_p^4 + 8z_p^4) - r_p a^8 z_p^4 (5 - 13z_p^2) - 2a^{10} z_p^4 (1 - 2z_p^2) (1 - z_p^2) \right] \\
&\quad + (\mathcal{L}_z - a\mathcal{E})^2 (1 - z_p^2)^2 \left[ 3r_p^8 - 3r_p^7 (4 - a^2 + 5a^2 z_p^2) - 6r_p^6 a^2 (3 + z_p^2) \right. \\
&\quad + r_p^5 a^2 (4 + 6a^2 + 3(12 - 13a^2)z_p^2 - 3a^2 z_p^4) - r_p^4 a^4 (13 - 12z_p^2 + 9z_p^4) \\
&\quad \left. + 3r_p^3 a^4 (a^2 - (4 + 11a^2)z_p^2 - 2a^2 z_p^4) + 2r_p^2 a^6 z_p^2 (9 - z_p^2) - 3r_p a^8 z_p^2 (3 + z_p^2) - a^8 z_p^4 \right] \\
&\quad - \mathcal{L}_z^2 r_p \left[ 3r_p^8 (1 + z_p^2)^2 - 3r_p^7 (1 + 3z_p^2) (3 + z_p^2) - r_p^6 (a^2 - (48 - 2a^2)z_p^2 - 47a^2 z_p^4 + 12a^2 z_p^6) \right. \\
&\quad - r_p^5 a^2 (5 - 8z_p^2 + 65z_p^4 + 18z_p^6) - r_p^4 a^2 z_p^2 (16 + 26a^2 - 57a^2 z_p^2 - (48 - 4a^2)z_p^4 + 3a^2 z_p^6) \\
&\quad + r_p^3 a^4 z_p^2 (22 - 21z_p^2 - 8z_p^4 - 9z_p^6) - r_p^2 a^4 z_p^2 (10a^2 - 15a^2 z_p^2 + 8(2 + a^2)z_p^4 - 3a^2 z_p^6) \\
&\quad \left. - r_p a^6 z_p^4 (1 - 5z_p^2) (5 - z_p^2) + 2a^8 z_p^4 (1 - 4z_p^2 + z_p^4) \right] \\
&\quad - r_p (3r_p^2 - a^2) (1 - z_p^4) (a^2 z_p^2 - \mathcal{Q}) \Delta_p^2 \Sigma_p,
\end{aligned} \tag{C1}$$

and

$$\begin{aligned}
\mathcal{I}_\beta &= \mathcal{E}^2 a^2 r_p (1 - z_p^2) \left[ r_p^9 (1 - z_p^2) (1 - 9z_p^2) - r_p^8 (6 - (18 + a^2)z_p^2 + (12 - a^2)z_p^4) \right. \\
&\quad - r_p^7 a^2 (2 + 35z_p^2 - 32z_p^4 + 3z_p^6) - r_p^6 a^2 (6 - (46 + a^2)z_p^2 + 2(22 - a^2)z_p^4 - (12 + a^2)z_p^6) \\
&\quad + r_p^5 a^4 (1 - 28z_p^2 + 33z_p^4 - 14z_p^6) - r_p^4 a^4 z_p^2 (1 - z_p^2) (4 + a^2 + (26 + a^2)z_p^2) \\
&\quad \left. + r_p^3 a^6 z_p^2 (1 + 18z_p^2 - 11z_p^4) - r_p^2 a^6 z_p^2 (a^2 + 2(3 + a^2)z_p^2 + (2 + a^2)z_p^4) + (4r_p - a^2)a^8 z_p^4 (1 + z_p^2) \right] \\
&\quad + (\mathcal{L}_z - a\mathcal{E})^2 (1 - z_p^2)^2 \left[ 3(r_p - 2)r_p^9 + 3r_p^8 a^2 (4 - 3z_p^2) - 2r_p^7 a^2 (11 - 6z_p^2) + r_p^6 a^4 (7 - 33z_p^2) \right. \\
&\quad \left. - 2r_p^5 a^4 (4 - 18z_p^2 + 3z_p^4) + r_p^4 a^6 (2 - 27z_p^2 - z_p^4) - 6r_p^3 a^6 z_p^4 - r_p^2 a^8 z_p^2 (3 - 2z_p^2) - z_p^4 a^{10} \right] \\
&\quad - \mathcal{L}_z^2 r_p \left[ 3r_p^9 (1 - z_p^2)^2 - r_p^8 (6 - a^2 - 2(6 + a^2)z_p^2 + (6 - a^2)z_p^4) + r_p^7 a^2 (2 - 29z_p^2 + 20z_p^4 - 9z_p^6) \right. \\
&\quad - r_p^6 a^2 (6 + a^2 - (40 + 3a^2)z_p^2 + 5(6 - a^2)z_p^4 - (12 + a^2)z_p^6) + r_p^5 a^4 (1 - 14z_p^2 + 25z_p^4 - 28z_p^6) \\
&\quad - 2r_p^4 a^4 z_p^2 (1 - z_p^2) (2 + 2a^2 + (17 + 2a^2)z_p^2 - 3z_p^4) + r_p^3 a^6 z_p^2 (2z_p^6 - 19z_p^4 + 32z_p^2 + 1) \\
&\quad \left. - r_p^2 a^6 z_p^2 (a^2 + (6 + 5a^2)z_p^2 + (4 + 3a^2)z_p^4 + (6 - a^2)z_p^6) + (4r_p - a^2)a^8 z_p^4 (1 + z_p^2)^2 \right] \\
&\quad - a^2 r_p (r_p^2 - a^2) (1 - z_p^4) (a^2 z_p^2 - \mathcal{Q}) \Delta_p^2 \Sigma_p.
\end{aligned} \tag{C2}$$