A new gravitational wave generation algorithm for particle perturbations of the Kerr spacetime

Enno Harms[1], Sebastiano Bernuzzi[1], Alessandro Nagar[2], and Anıl Zenginoğlu[3]

[1]Theoretical Physics Institute, University of Jena, 07743 Jena, Germany
[2]Institut des Hautes Etudes Scientifiques, 91440 Bures-sur-Yvette, France
[3]Theoretical Astrophysics, California Institute of Technology, Pasadena, CA 91125, USA

Abstract. We present a new approach to solve the 2+1 Teukolsky equation for gravitational perturbations of a Kerr black hole. Our approach relies on a new horizon penetrating, hyperboloidal foliation of Kerr spacetime and spatial compactification. In particular, we present a framework for waveform generation from point-particle perturbations. Extensive tests of a time domain implementation in the code *Teukode* are presented. The code can efficiently deliver waveforms at future null infinity. The accuracy and convergence of the waveforms' phase and amplitude is demonstrated.

As a first application of the method, we compute the gravitational waveforms from inspiraling and coalescing black-hole binaries in the large-mass-ratio limit. The smaller mass black hole is modeled as a point particle whose dynamics is driven by an effective-one-body-resummed analytical radiation reaction force. We compare the analytical, mechanical angular momentum loss (computed using two different prescriptions) to the gravitational wave angular momentum flux. We find that higher-order post-Newtonian corrections are needed to improve the consistency for rapidly spinning binaries. We characterize the multipolar waveform as a function of the black-hole spin. Close to merger, the subdominant multipolar amplitudes (notably the $m = 0$ ones) are enhanced for retrograde orbits with respect to prograde ones. We argue that this effect mirrors nonnegligible deviations from circularity of the dynamics during the late-plunge and merger phase. For the first time, we compute the gravitational wave energy flux flowing into the black hole during the inspiral using a time-domain formalism proposed by Poisson.

Finally, a self-consistent, iterative method to compute the gravitational wave fluxes at leading-order in the mass of the particle is developed. The method can be used alternatively to the analytical radiation reaction in cases the analytical information is poor or not sufficient. Specifically, we apply it to compute dynamics and waveforms for a rapidly rotating black hole with dimensionless spin parameter $\hat{a} = +0.9$. For this case, the simulation with the consistent flux differs from the one with the analytical flux by ~ 35 gravitational wave cycles over a total of about 250 cycles. In this simulation the horizon absorption accounts for about +5 gravitational wave cycles.

1. Introduction

In this work we develop a new, accurate and efficient time domain wave generation algorithm for perturbations of the gravitational field around the Kerr spacetime. We focus on point-particle perturbations, and implement the algorithm to calculate gravitational waveforms generated by a particle in equatorial motion. The method is then applied to modeling the late-inspiral–merger waveforms from large-mass-ratio black hole binaries, using an effective-one-body (EOB) approach for the particle trajectory. We report new results on the multipolar structure of the waveforms, EOB radiation reaction, and horizon-absorbed fluxes.

Perturbations of a field with spin s (integer or semi-integer) on a Kerr background are typically described by a master equation derived by Teukolsky [1, 2]. The Teukolsky equation (TE) is separable in the frequency domain, therefore solutions to this equation have been historically obtained first in the frequency domain (see e.g. [3, 4, 5]). Time domain methods are appealing, however, for the ease of treating non-circular orbits. The first numerical computation of gravitational perturbations of Kerr spacetime in the time domain is by Krivan et al. in 1997 [6]. They use Boyer–Lindquist coordinates and solve the TE as a fully first-order system in 2+1 dimensions after a decomposition into azimuthal angular modes. The system is discretized using a second-order convergent Lax–Wendroff scheme which has favorable dissipative properties for numerical stability. The computational boundaries are placed at finite radii and close to the horizon using a tortoise transformation of the radial coordinate. The errors from the boundaries are mitigated by using a large computational domain. Applications and improvements of this scheme can be found, for example, in [7, 8, 9, 10, 11, 12, 13, 14].

Even though the 2+1 TE is a linear equation, its time-domain integration with $s \neq 0$ is challenging and nontrivial due to stiff terms, exponentially growing continuum modes, and boundary treatment [6, 15]. Until recently, the method of [6] was the only successful approach to gravitational ($|s| = 2$) perturbations. However, developments in computational and geometric methods suggest that this problem should be revisited. On the geometric side, the accuracy ‡ of the numerical approach can be significantly improved by the use of horizon-penetrating coordinates and hyperboloidal compactification [16]. The idea is to use a spacelike foliation that penetrates the horizon in a regular way and asymptotically approaches null infinity so that one can compute both the ingoing and the outgoing radiation. This technique removes the two largest systematic uncertainties of Krivan's method (and of any other based only on Boyer–Lindquist coordinates), namely (i) the inner and outer numerical boundary errors, and (ii) finite-radius-extraction (and/or extrapolation) errors. On the computational side, our approach is efficient because the computational domain is much smaller than the one needed in standard Boyer–Lindquist tortoise coordinates. The efficiency of the method can be further enhanced by employing high order numerical discretization techniques (spectral methods or finite differencing). The accuracy and efficiency resulting from the application of the above techniques have been demonstrated in a calculation of gravitational tail decay rates in Schwarzschild

‡ We distinguish between accuracy and efficiency, referring, respectively, to systematic and truncation errors. When the only source of error is discretization, an efficient code is also accurate because it allows the use of more computational resources thereby reducing the truncation error. In general, however, numerical calculations of gravitational waveforms include *systematic* errors due to inaccurate numerical boundary treatment or extrapolation from finite radii. Our claim of accuracy refers to the removal of these latter sources of error.

spacetime [17], and for generic spin fields in Kerr spacetime, including the extreme case [15].

In this paper we develop the work presented in [15] in two directions. First, we introduce a new foliation of Kerr spacetime which leads to more efficient numerical computations than in [15]. Second, we extend the approach to point particle perturbations and test applications relevant for the binary black-hole problem.

Historically, particle perturbations of black hole spacetimes played a crucial role in understanding the general relativistic two-body problem and the related problem of modeling the emission of gravitational waves (GWs). Perturbation theory gave access to the first waveforms from the strong-field/fast-motion regime (e.g. [18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28]), and has been used to study the radiation reaction problem including black hole absorption (e.g. [29, 30, 31, 32, 33]). Perturbative methods interface with post-Newtonian (PN) calculations in the test-mass limit [34], and have led to a very high-order PN expansion of the circular GW flux [35, 36]. Also, wave generation algorithms and the numerical evolution of linear perturbations sourced by particles are key steps for the self-force problem [37, 38, 39, 40].

Most important for this work, linear point-particle perturbations provide a natural tool to model black-hole binaries in the large-mass-ratio limit. Various strategies have been proposed to model *adiabatic* inspirals of large-mass-ratio binaries [41, 42, 43, 44, 45, 10, 11, 46]. More recently, time domain solutions of the Regge-Wheeler-Zerilli equations (RWZE) [47, 48] with a particle source term have been used as a building block of a hybrid method that models *non-adiabatic* inspirals and the transition inspiral-plunge-ringdown, shortly *insplunge* [49, 50]. The method combines the wave generation algorithm with an analytical, effective-one-body prescription for the particle dynamics. The conservative Hamiltonian equations of motion for the particle are augmented by a radiation-reaction term, $\hat{\mathcal{F}}$ linear in the symmetric mass-ratio, i.e. of order $\mathcal{O}(\nu)$. In this paper, following standard practice in Numerical Relativity, we define the mass ratio as a quantity larger than 1, that is $q = M/\mu$, where μ is the mass of the smaller black hole and M the one of the larger black hole. In our discussion, we will also use the quantity $\nu \equiv 1/q = \mu/M$, that we address, with a slight abuse of language, "symmetric" mass ratio §. In the following, we will often set $M = 1$, which allows to identify ν with μ. The radiation reaction is built from the factorized and resummed PN waveform for circular orbits of [51]. The hybrid RWZE/EOB method proved to be a valuable "binary black-hole laboratory" and led to (i) a detailed analysis of multipolar merger/ringdown waveforms [52]; (ii) an improved analytical description of horizon-absorbed fluxes [53]; (iii) an improved EOB waveform [54, 55]; and (iv) an accurate calculation of the gravitational recoil (kick) [52, 56]. Remarkably, the perturbative kick calculation can be analytically extrapolated to finite mass ratios and provides quantitative answers also for that case [57]. Most of the results mentioned above rely on an accurate time-domain RWZE solver that employs the hyperboloidal layer method to include null infinity in the computational domain [56, 58, 59]. The hybrid, perturbative method has been applied to the Kerr case in [60, 14, 61], where the TE time domain solver of [6, 10] is employed.

In the remainder of this paper, we extend the work in [52, 54, 56] to include the spin of the massive black hole. We characterize the multipolar waveform at

§ The actual symmetric mass ratio of a system of two masses m_1 and m_2 is defined as $\nu = m_1 m_2/(m_1+m_2)^2$. In the large-mass-ratio limit $q = m_1/m_2 \gg 1$ this corresponds to $m_1 m_2/(m_1 + m_2) \approx m_2 = \mu$ and $(m_1 + m_2) \approx m_1 = M$ so that $\nu \to \mu/M$.

null infinity, quantify finite radius extraction and extrapolation uncertainties, and investigate the performance of analytical EOB prescriptions for the radiation reaction during the insplunge as a function of the black hole spin. For the first time, we compute the gravitational wave energy flux into the black hole during the insplunge using a time-domain formalism, proposed by Poisson [62]. Finally, we present a numerical iterative method to compute gravitational wave fluxes at scri and at the horizon that are consistent to first order in the symmetric mass ratio. The method is applied to a rapidly spinning case for which the analytical EOB prescription is inaccurate. We discuss significant differences between the use of the consistent radiation reaction and the EOB one as well as the effect of including horizon-absorbed fluxes.

This paper is organized as follows. In Sec. 2 we introduce a new, horizon-penetrating, hyperboloidal foliation of Kerr building on [56, 58, 59], and compare the new coordinates to those in [63, 15]. In Sec. 3 the inhomogeneous 2+1 TE is rewritten using these new coordinates. The derivation of the point particle source term is presented in detail. In Sec. 4 we summarize the numerical method employed in the *Teukode*. In particular, we discuss the advantages of the new coordinates for the numerics and the strategies to represent the particle as a Dirac-distribution. In Sec. 5 we describe the EOB dynamics used for the nonadiabatic insplunge experiments and compare it with the one of [60, 64, 61]. In Sec. 6 we assess the implementation by considering simple geodetic motion (circular orbits and radial plunge), insplunge waveforms for the nonrotating background, and self-convergence tests for waveforms. In Sec. 7 we present the main results on large-mass-ratio insplunge waveforms. First, we study the consistency between two different analytical prescriptions for the EOB fluxes and the one numerically computed by solving the 2+1 TE for the given dynamics. Second, we present new multipolar merger waveforms at scri for various values of the background spin up to $|\hat{a}| = 0.9999$, and discuss the structure of amplitudes in relation to the particle dynamics. Third, we quantify finite radius extraction errors and discuss the performance of the extrapolation procedure commonly employed in numerical relativity when applied to our setup, e.g. [65]. In Sec. 8 we present the calculation of the energy absorbed by the black hole during the insplunge. In Sec. 9 we describe the self-consistent numerical method for the GW fluxes. A complete calculation for a case study of a rapidly rotating hole is presented. Conclusions are given in Sec. 10. In Appendix A we collect in tables quantitative information on the used dynamics and the produced waveforms. Appendix B reviews the Hamiltonian formulation of the particle dynamics. Appendix C presents a leading-order analysis of next-to-quasi-circular effects in the quadrupolar waveform.

Geometric units ($c = G = 1$) are employed throughout the paper.

2. Horizon-penetrating, hyperboloidal foliations of Kerr spacetime

It has long been argued that hyperboloidal foliations should have favorable properties for numerical calculations of outgoing gravitational radiation [66]. Until recently, however, it was not clear how to use hyperboloidal surfaces in black hole perturbation theory. An important step was presented by Moncrief in a talk [67], wherein he showed how to fix the coordinate location of null infinity (scri-fixing) and gave an example in Minkowski spacetime (see also [68]). This construction was later used in numerical studies in Minkowski spacetime [69]. For black hole spacetimes there were no similarly convenient coordinates available.

A general framework for the construction of explicit, hyperboloidal coordinates

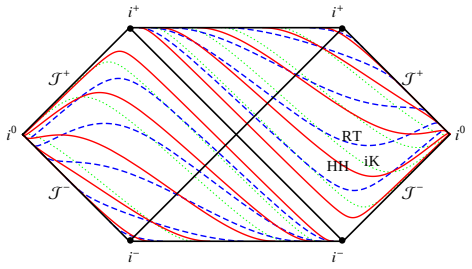

Figure 1. Foliations of Schwarzschild spacetime ($a = 0$): ingoing Kerr (iK; green dotted), Racz–Tóth (RT; blue dashed), and horizon-penetrating-hyperboloidal (HH; red solid) coordinates with $S = 10$.

with scri-fixing for stationary, weakly asymptotically flat spacetimes (including black hole spacetimes) was presented in [16]. The idea is to introduce a new time coordinate with a height function that satisfies certain asymptotic properties, and to introduce a compactifying coordinate in the radial direction along with an *explicit* conformal factor. The transformation from standard coordinates $\{t, r\}$ to hyperboloidal coordinates $\{\bar{t}, \bar{r}\}$ can be written as

$$t = \bar{t} + h(\bar{r}), \qquad r = \frac{\bar{r}}{\Omega(\bar{r})}\,, \tag{1}$$

where $h(\bar{r})$ is the height function and the *explicitly prescribed* $\Omega(\bar{r})$ acts both as a radial compactification and a conformal factor. Asymptotic conditions derived in [16] make sure that the resulting metric is regular. This method has been successfully demonstrated in various examples and today there are many choices available for $h(\bar{r})$ and $\Omega(\bar{r})$ in Minkowski, Schwarzschild, Reissner–Nordström, and Kerr spacetimes. How these functions are chosen beyond their asymptotic behavior is important for the efficiency of the related numerical computation (see Sec. 4.2).

In the following, we discuss two such choices for the Kerr family. The first one has been constructed by Racz and Tóth in [63] ("RT" coordinates hereafter) and has been used in a numerical study of tail decay rates in Kerr spacetime. The second one follows the general ideas of [56] ("HH" for horizon-penetrating, hyperboloidal) and has been used in numerical studies of quasinormal mode behavior in nearly extremal Kerr spacetimes in [70]. Both coordinate systems start from the ingoing Kerr metric, which we review now.

Consider the Kerr metric in Boyer–Lindquist (BL) coordinates $\{t, r, \theta, \phi\}$

$$\begin{aligned} g_{\mathrm{BL}} = &-\left(1 - \frac{2Mr}{\Sigma}\right) dt^2 - \frac{4aMr}{\Sigma}\sin^2\theta\, dt\, d\phi + \frac{\Sigma}{\triangle}\, dr^2 \\ &+ \Sigma\, d\theta^2 + \left(r^2 + a^2 + \frac{2Ma^2 r \sin^2\theta}{\Sigma}\right) \sin^2\theta\, d\phi^2\ , \end{aligned} \tag{2}$$

where the two parameters are the mass of the Kerr spacetime M and its angular momentum parameter a, so that its angular momentum is given as $j = aM$. We further have $\Sigma \equiv r^2 + a^2\cos^2\theta$ and $\Delta \equiv r^2 + a^2 - 2Mr = (r - r_+)(r - r_-)$. The ingoing Kerr (iK) coordinates $\{\tilde{t}, r, \theta, \varphi\}$ are obtained through the following transformation

$$d\tilde{t} = dt + \frac{2Mr}{\Delta} dr, \qquad d\varphi = d\phi + \frac{a}{\Delta} dr\ , \tag{3}$$

i.e.

$$\tilde{t} = t - r + \int \frac{a^2 + r^2}{\Delta}\, dr, \qquad \varphi = \phi + a \int \frac{dr}{\Delta}\ . \tag{4}$$

The resulting Kerr metric reads

$$\begin{aligned} g_{\mathrm{iK}} = &- \left(1 - \frac{2Mr}{\Sigma}\right) d\tilde{t}^2 - \frac{4aMr}{\Sigma} \sin^2\theta\, d\tilde{t}\, d\varphi - 2a \sin^2\theta \left(1 + \frac{2Mr}{\Sigma}\right) dr d\varphi + \\ &+ \frac{4Mr}{\Sigma}\, d\tilde{t} dr + \left(1 + \frac{2Mr}{\Sigma}\right) dr^2 + \Sigma d\theta^2 + \left(r^2 + a^2 + \frac{2Ma^2 r \sin^2\theta}{\Sigma}\right) \sin^2\theta d\varphi^2. \end{aligned} \tag{5}$$

This representation of the Kerr metric is the starting point for both the RT and HH coordinates, which we discuss next.

2.1. RT coordinates

The RT coordinates presented in [63] are very similar to Moncrief's scri-fixing coordinates in Minkowski spacetime [67, 68, 69] but with an additional, necessary term to satisfy the asymptotic properties presented in [16]. Denoting the time and space RT coordinates by $\{T, R\}$, the transformation from the ingoing Kerr coordinates reads

$$\tilde{t} = T - 4M \ln\left(1 - R^2\right) + \frac{1 + R^2}{1 - R^2}\ , \quad r = \frac{2R}{1 - R^2}\ . \tag{6}$$

So the height function and the conformal factor are given by

$$h(R) = \frac{1 + R^2}{2\Omega} - 4M \ln 2\Omega\ , \qquad \Omega(R) = \frac{1 - R^2}{2}\ . \tag{7}$$

Note that the height function blows up at infinity, where the conformal factor vanishes, in a suitable way. The first term in the height function is the same term as in Minkowski spacetime; the second term is needed due to the presence of the black hole. The resulting hyperboloidal foliation of Kerr spacetime is horizon penetrating and smoothly reaches scri at $R = 1$. The event horizon R_+ in the new radial coordinate R is located at

$$R_+ = \frac{2\sqrt{2M\sqrt{M^2 - a^2} - a^2 + 2M^2 + 1} - 2}{2\left(\sqrt{M^2 - a^2} + M\right)}\ . \tag{8}$$

Figure 1 shows the time surfaces of the RT coordinates in a conformal diagram for the non-rotating $a = 0$ case (blue lines). We mention that RT coordinates can be modified to allow a prescribed scri position [15].

2.2. HH coordinates

We present a coordinate system which includes a free parameter and is more efficient for numerical calculations as we argue in Sec. 4.2.

An intuitive way to construct hyperboloidal coordinates is to demand invariance of the coordinate expression for outgoing characteristics in spatially compactifying coordinates [56]. This requirement is beneficial for numerical purposes because, for a prescribed choice of spatial compactification, the outgoing characteristic speeds do not impose strong restrictions on the allowed time step due to the Courant-Friedrich-Lewy

(CFL) limit. We are mainly interested in the asymptotic behavior of outgoing null surfaces and therefore ignore the angular dependence by setting $a = 0$.

The outgoing radial characteristics for vanishing specific angular momentum in the ingoing Kerr coordinates are given by $u = \tilde{t} - (r + 4M \ln(r - 2M))$. For a further simplification, we drop the subtraction term in the logarithm that is only relevant near the horizon. Denoting the time and space HH coordinates by $\{\tau, \rho\}$, our requirement becomes

$$\tilde{t} - (r + 4M \ln r) \stackrel{!}{=} \tau - (\rho + 4M \ln \rho) \ . \tag{9}$$

For any choice of spatial compactification through ρ, the above requirement determines the foliation. We choose the simplest conformal factor with a variable scri location S and set

$$h(\rho) = \frac{\rho}{\Omega} - \rho - 4M \ln \Omega, \qquad \Omega(\rho) = 1 - \frac{\rho}{S} \, . \tag{10}$$

Note that the asymptotic behavior of these functions near scri is similar to the one given for the RT coordinates in (7). The event horizon ρ_+ in the new radial coordinate ρ is located at

$$\rho_+ = \frac{a^2 S + M S^2 + \sqrt{M^2 S^4 - a^2 S^4}}{a^2 + 2MS + S^2} \, . \tag{11}$$

We will indicate with HH_S these coordinates with a specific choice of S, e.g. HH_{10} refers to $S = 10$. The foliation HH_{10} is shown in a conformal diagram in Fig. 1 (red lines) for $a = 0$. It is qualitatively similar to the RT coordinates [63] with the main differences that the location of null infinity can be chosen freely and the outgoing radial characteristic speeds are similar to those of the ingoing Kerr coordinates.

2.3. Advanced and retarded time coordinates

Gravitational waves propagate along null geodesics. The structure of null geodesics in Kerr spacetime is rather complicated, but their approximations by Schwarzschild null geodesics is sufficient for our purposes. The retarded and advanced time coordinates in Schwarzschild spacetime, u and v, are defined by $u = t - r_*$ and $v = t + r_*$, where $r_* = r + 2M \ln(r/2M - 1)$ is the Schwarzschild tortoise coordinate. We use these coordinates in Kerr spacetime to connect the particle's dynamics with the measured gravitational radiation. This approximation agrees with the general practice in numerical relativity, where null geodesics in a binary black-hole spacetime are approximated by their Schwarzschild counterparts for extrapolating gravitational waveforms [65].

Here, we give the retarded and advanced time coordinates in the horizon-penetrating, hyperboloidal coordinates used in the simulations. We get for the retarded coordinate

$$u(\tau, \rho) = \tau - \rho - 4M \ln \left(\frac{S\rho + 2M\rho - 2MS}{S} \right) + 2M \ln 2M \ , \tag{12}$$

and for the advanced coordinate

$$v(\tau, \rho) = \tau + \rho \, \frac{S + \rho}{S - \rho} - 4M \ln \left(\frac{S - \rho}{S} \right) - 2M \ln(2M) \ . \tag{13}$$

The constant term $2M \ln 2M$ comes from different conventions in the tortoise coordinate. All plots in this paper employ the above coordinates for visualizing

waves and fluxes at scri or the horizon. This allows a direct comparison of dynamical quantities (e.g. the particle's orbital frequency) with the measured wave signal because the BL-t used for the particle's dynamics can be identified with the retarded time at scri and the advanced time at the horizon.

3. Inhomogeneous 2+1 Teukolsky equation with a particle source

In this section we briefly outline the main steps in the derivation of the 2+1 Teukolsky equation in horizon-penetrating, hyperboloidal coordinates. The calculation of the stress-energy tensor with a particle source term is presented in detail. Formulas for gravitational wave fluxes at scri and the horizon in terms of 2+1 fields are collected.

3.1. The Teukolsky equation

The Teukolsky equation describes the evolution of perturbations of certain Weyl tensor components on a Kerr background [2]. The equation is derived using the Newman–Penrose formalism [71] which relies on the choice of a tetrad and a coordinate system. In the original calculation Teukolsky used the Kinnersley tetrad in BL coordinates which reads

$$\ell^\mu = \frac{(r^2+a^2, \Delta, 0, a)}{\Delta}\,, \quad n^\mu = \frac{(r^2+a^2, -\Delta, 0, a)}{2\Sigma}\,, \quad m^\mu = \frac{(\mathrm{i}a\sin\theta, 0, 1, \mathrm{i}\csc\theta)}{\sqrt{2}(r+\mathrm{i}a\cos\theta)}\,. \tag{14}$$

In practice the equivalent, rationalized versions of m^μ and its complex conjugate $m^{\mu *}$ are preferable, e.g.

$$m^\mu = \frac{(r-\mathrm{i}a\cos\theta)}{\sqrt{2}\Sigma}\,(\mathrm{i}a\sin\theta, 0, 1, \mathrm{i}\csc\theta)\,. \tag{15}$$

One straightforward method to obtain the Teukolsky equation in horizon-penetrating, hyperboloidal coordinates is to substitute the transformation formulas (1) directly into the equation as given by Teukolsky [2]. Subsequently, the unknown variable is rescaled for regularity of the coefficients at the horizon and at null infinity [72, 15]. The field ψ with spin weight s behaves as Δ^s towards the horizon and as r^{2s+1} towards null infinity. Therefore, the rescaling goes as

$$\psi \mapsto \Delta^{-s} r^{-1} \psi\,. \tag{16}$$

This is the approach followed in [15] for the transformation of the homogeneous Teukolsky equation into RT coordinates.

For this paper we found that deriving the Teukolsky equation from scratch using the new coordinates, while equivalent to the above transformation, yields better results in the source term because the cancellations for regularity at the horizon are implied automatically. To this end, we first perform a null rotation that corresponds to the rescaling given above. For example, for regularity at the horizon a null rotation with $\Lambda = \Delta$ can be performed as in [73]. The rotation gives the tetrad fields

$$\ell^\mu = (\Delta + 4Mr, \Delta, 0, 2a)\,, \quad n^\mu = \frac{(1, -1, 0, 0)}{2\Sigma}\,. \tag{17}$$

Note that, under a Λ rotation, $\ell^\mu \to \Lambda \ell^\mu$, $n^\mu \to \Lambda^{-1} n^\mu$ and m^μ stays unchanged. With $\Lambda = \Delta$ and $s = -2$, the Weyl scalar $\psi_4 = -n^\alpha m^{*\beta} n^\gamma m^{*\delta} C_{\alpha\beta\gamma\delta}$ transforms as

$\psi_4 \to \Delta^{-2}\psi_4$. For general s, the null rotation with $\Lambda = \Delta$ corresponds to a rescaling with Δ^{-s}. To obtain regularity at scri (in the homogeneous part of the equation) a subsequent rescaling by r^{-1} or an additional null rotation must be performed. Then the tetrad is transformed to horizon-penetrating, hyperboloidal coordinates and used to calculate the coefficients and the source term for the Teukolsky equation. A more general approach can also be taken by incorporating free coordinate functions in the tetrad before the derivation of the TE [17].

Finally the Teukolsky equation is transformed in 2+1 form separating each Fourier m-mode in the azimuthal direction. The resulting equation has the general form

$$C_{\tau\tau}\partial_{\tau\tau}\psi + C_{\tau\rho}\partial_{\tau\rho}\psi + C_{\rho\rho}\partial_{\rho\rho}\psi + C_{\theta\theta}\partial_{\theta\theta}\psi + C_{\tau}\partial_{\tau}\psi + C_{\theta}\partial_{\theta}\psi + C_{\rho}\partial_{\rho}\psi + C_0\psi = S_s \ , \tag{18}$$

with coefficients $C(\rho, \theta; m, s)$ depending on the background coordinates, the spin weight s, and the azimuthal mode-index m. The index m in the variable ψ_m has been suppressed for brevity. The general formulas for the coefficients under time transformation, spatial compactification, and rescaling can be found in [74].

3.2. Particle source term

The calculation of the stress-energy tensor for a particle source is well documented for the frequency-domain Teukolsky equation (see [34]). For time domain applications, however, we could not find a complete description. Therefore we present the calculation in detail for a Hamiltonian formulation of the particle dynamics (see Appendix B).

The source term S_s depends on the spin weight s, the background metric, and the stress-energy tensor, $T_{\mu\nu}$, of the matter perturbation. The general form of S_s is given in [2]. Here we discuss gravitational perturbations ($s = \pm 2$). In BL coordinates,

$$S_{-2} = 8\pi\Sigma(r - \mathrm{i}a\cos\theta)^4 T_4 \ , \tag{19}$$

$$S_{+2} = 8\pi\Sigma T_0 \ , \tag{20}$$

where T_4 and T_0 are expressions involving contractions of the stress-energy tensor with tetrad vector fields and their first and second partial derivatives with respect to the background coordinates. Specifically, the contractions involved are

$$T_{mm*} \equiv T_{\mu\nu}m^\mu m^{*\nu} \ , \quad T_{nm*} \equiv T_{\mu\nu}n^\mu m^{*\nu} \ , \quad T_{nn} \equiv T_{\mu\nu}n^\mu n^\nu \ . \tag{21}$$

For example, the term T_4 is given by

$$\begin{aligned} T_4 = {} & (\underline{\Delta} + 3\gamma - \gamma^* + 4\mu + \mu^*)(\delta^* - 2\tau^* + 2\alpha)T_{nm^*} \\ & - (\underline{\Delta} + 3\gamma - \gamma^* + 4\mu + \mu^*)(\underline{\Delta} + 2\gamma - 2\gamma^* + \mu^*)T_{m^*m^*} \\ & + (\delta^* - \tau^* + \beta^* + 3\alpha + 4\pi)(\underline{\Delta} + 2\gamma + 2\mu^*)T_{nm^*} \\ & - (\delta^* - \tau^* + \beta^* + 3\alpha + 4\pi)(\delta^* - \tau^* + 2\beta^* + 2\alpha)T_{nn} \ , \end{aligned} \tag{22}$$

where $\underline{\Delta}, \delta, \gamma, \mu, \tau, \alpha, \beta$ are the complex Newmann-Penrose operators, e.g. [2]. In Eq. (22) $\underline{\Delta} = n^\alpha\partial_\alpha$ (the underbar is introduced to distinguish it from our previous definition of Δ) and $\delta^* = m^{*\alpha}\partial_\alpha$ are differential operators that depend on the coordinates. They do not commute, that is, $[\underline{\Delta}, \delta^*] \neq 0$. For example, using the

Kinnersley tetrad in BL coordinates they read

$$\underline{\Delta} = \frac{1}{2\Sigma}\left(\left(r^2+a^2\right)\partial_t - \Delta\partial_r + a\ \partial_\phi\right) \tag{23}$$

$$\delta^* = \frac{1}{\sqrt{2}\Sigma}\left((a^2\cos\theta\sin\theta - \mathrm{i}\ r\ a\sin\theta)\partial_t + (r+\mathrm{i}\ a\cos\theta)\partial_\theta + \left(a\frac{\cos\theta}{\sin\theta} - \mathrm{i}\frac{r}{\sin\theta}\right)\partial_\phi\right) . \tag{24}$$

The related 2+1 decomposed operators are obtained by the substitution $\partial_\phi \to \mathrm{i}\ m$.

The stress-energy tensor for a point particle with mass μ can be written as

$$T_{\mu\nu} = \mu\int \frac{d\lambda}{\sqrt{-g}}\ u_\mu\, u_\nu\ \delta^4(x^\alpha - X^\alpha(\lambda)) \tag{25}$$

with λ the proper time, $X^\alpha(\lambda)$ the particle's worldline and $u^\alpha(\lambda) = dX^\alpha/d\lambda$ the 4-velocity. Assuming the particle's motion is described by Hamiltonian dynamics (Appendix B), we write the 4-velocity in terms of the reduced momenta, $u_\alpha = (-\hat{H}, p_i)$, and the coordinates as $x^\alpha = (t, q^i)$. Here, $t \equiv x^0$ is a generic time coordinate, not necessarily the BL time. Replacing the affine parameter $\lambda(t)$ with t and integrating Eq. (25) one gets

$$T_{\mu\nu} = \frac{\mu}{\sqrt{-g}}\frac{d\lambda}{dt}p_\alpha p_\beta \delta^3(x^i - q^i(t)) . \tag{26}$$

For the calculation of the 2+1 source term the equation above is mode decomposed. In BL coordinates, we write $T_{\mu\nu} = \sum_m T^m_{\mu\nu}e^{\mathrm{i}m\phi}$, and using

$$\delta(\phi - \phi(t)) = (2\pi)^{-1}\sum_m e^{\mathrm{i}m(\phi-\phi(t))} , \tag{27}$$

the $T^m_{\mu\nu}$ components read (dropping the m-index for brevity)

$$T_{00} = \frac{\mu A\ \hat{H}^2}{2\pi\Sigma\sin\theta\,(\hat{H} - \omega p_\phi)}\ \delta(r-r(t))\delta(\theta-\theta(t))e^{-\mathrm{i}m\phi(t)} , \tag{28a}$$

$$T_{0i} = \frac{\mu A\ (-\hat{H})p_i}{2\pi\Sigma\sin\theta\,(\hat{H} - \omega p_\phi)}\ \delta(r-r(t))\delta(\theta-\theta(t))e^{-\mathrm{i}m\phi(t)} , \tag{28b}$$

$$T_{ij} = \frac{\mu A\ p_i p_j}{2\pi\Sigma\sin\theta\,(\hat{H} - \omega p_\phi)}\ \delta(r-r(t))\delta(\theta-\theta(t))e^{-\mathrm{i}m\phi(t)} , \tag{28c}$$

where we have specified to the Kerr metric, $\sqrt{-g} = \Sigma\sin\theta$, and used $d\lambda/dt = A/(\hat{H} - \omega p_\phi)$ (see Eq. (B.9)).

Inspecting Eq. (22) and Eq.'s (28), we define the following functions

$$L(t,r,\theta) = \frac{\mu}{2\pi\Sigma\sin\theta}\frac{d\lambda}{dt} = \frac{\mu}{2\pi\Sigma\sin\theta}\frac{A}{\hat{H} - \omega p_\phi} \tag{29a}$$

$$M(t,r,\theta) = m^{\mu *}p_\mu = \frac{(r+\mathrm{i}a\cos\theta)}{\sqrt{2}\Sigma}\left(\mathrm{i}a\sin\theta\ \hat{H} + p_\theta - \mathrm{i}\frac{p_\phi}{\sin\theta}\right) \tag{29b}$$

$$N(t,r,\theta) = n^\mu p_\mu = \frac{1}{2\Sigma}\left(-\hat{H}(r^2+a^2) - \Delta p_r + a p_\phi\right) , \tag{29c}$$

and use them to express the contractions of the stress-energy tensor with the tetrad, e.g. $T_{nm*} = L\ M\ N\ \delta^2(...)\ e^{(\cdots)}$. The quantity T_4 in Eq. (22) is then written in

terms of L, M, N and their derivatives using the chain rule. In practice, in the code, the algebraic complexity of the source is greatly reduced by computing and storing L, M, N and derivatives, and combining them.

In order to compute the derivatives in Eq. (22) one needs to make a choice because the delta functions formally identify the background coordinates with the particle's position coordinates. The result is theoretically independent of the particular choice, but the explicit expressions can differ and, numerically, there may be differences due to truncation errors (see discussion in [49]). Also, one has to perform the calculation consistently. We leave all the background coordinates in the L, M, N functions unchanged and consider functions of time only the reduced momenta $p_i(t)$, the Hamiltonian $\hat{H}(t)$ and the trajectory $q^i(t)$. The time dependence in the momenta and the Hamiltonian is assumed to account for dissipative forces (radiation reaction). The time derivatives of the $q^i(t)$ are systematically substituted with the right-hand-side (r.h.s.) of the Hamiltonian equation of motion, so they do not appear explicitly but only $\dot{p}_i(t), \dot{\hat{H}}(t), \ddot{p}_i(t), \ddot{\hat{H}}(t)$ remain.

We want to mention here a difference between the calculation of the TE source and the RWZE source. Differently from the TE, in the RWZE case it is possible to remove explicit time derivatives from the source by using the (linearized) Bianchi identities in the metric perturbation framework. Specifically and referring to equations in Ref. [75], this can be accomplished by substituting the Eq. (4.21) into the even parity source in Eq. (4.27) and the Eq. (5.12) into the odd parity source in Eq. (5.16). The substitution is used in the calculation of [76] although not explicitly stated.

Let us comment on the use of HH coordinates. First the NP operators in Eq. (22) and the tetrad in Eq. (17) have to be rewritten in the HH-coordinate system. For iK coordinates explicit expressions from [73] can be used as a starting point to apply the transformation of Eq. (10). Because the particle's trajectory is usually computed in BL coordinates we also have to apply the coordinate transformations to the trajectory and to the momenta (i.e. to $T^{\mu\nu}$). We emphasize that the rewriting of the source term in HH-coordinates as sketched above is important in our approach, even though the terms $T_{4,0}$ are invariant under coordinate transformations (being tetrad scalars). The reason is that in the code the particle event at each time step must be located in hyperboloidal coordinates, say $\rho(\tau)$, and any discrete representation of the delta function (see Sec. 4.3) involves a few grid points around $\rho(\tau)$. The transformation $t = \tau + h(\rho)$ introduces a grid-point dependent BL-time t at a given evolution slice τ. Using a trajectory in BL coordinates would mean to have a non-unique particle's position at the evolution slices τ of the code. Re-writing the source term in the evolution coordinates removes the ambiguities and thus greatly simplifies the source treatment.

3.3. Gravitational strain

We describe the relation between the master variables of the Teukolsky equation and the gravitational strain. At scri, the $s = -2$ master variable ψ_m of the 2+1 TE written in HH (or RT) coordinates is $r\psi_{4\ m}$, i.e. the m-mode of the Weyl scalar (in the Kinnersley tetrad) describing asymptotically outgoing radiation multiplied by r. The $s = +2$ master variable ψ_m corresponds instead to the m-mode of the Weyl scalar $r\psi_{0\ m}$ also (referring to the tetrad of eq. (17)).

The GW strain $h = h_+ - \mathrm{i}\, h_\times$ is found by integrating the asymptotic relation

$$\ddot{h} = 2\psi_4 \tag{30}$$

for each m-mode. The integration gives $r\, h_m(u,\theta)$ along scri. We also compute the (spin weighted with $s=-2$) multipoles $r\, h_{\ell m}(u)$, defined through

$$r\, h = \sum_{\ell=2,m} r\, h_{\ell m}\, {}_{-2}Y_{\ell m}(\theta,\phi) = \sum_{\ell=2,m} \sqrt{\frac{(\ell+2)!}{(\ell-2)!}}\; \Psi_{\ell m}\; {}_{-2}Y_{\ell m}(\theta,\phi)\ , \tag{31}$$

by mode-projecting $r\, h_m(u,\theta)$ in the θ-direction. The complex quantities $\Psi_{\ell m} = \Psi^{(\mathrm{e})}_{\ell m} + \mathrm{i}\Psi^{(\mathrm{o})}_{\ell m}$ in Eq. (31) are the RWZE variables [76]. The energy flux at scri is given by

$$\dot{E} = \frac{1}{16\pi}\int_{S_2} d\Omega\ |r\,\dot{h}|^2 = \frac{1}{16\pi}\sum_m \int_{-1}^{1} d\xi\ |r\,\dot{h}_m|^2\ , \tag{32}$$

where in the last expression we have used $\xi = \cos\theta$ and introduced the mode-decomposition of h to express the flux in terms of the 2+1 fields. The angular momentum $\vec{J} = (J_x, J_y, J_z)$ flux is given by

$$\dot{J}_i = -\frac{1}{16\pi}\Re\left\{\int_{S_2} d\Omega\ (r\,\dot{h})^*\mathcal{J}_i(r\,h)\right\}\ , \tag{33}$$

where $\mathcal{J}_i$ are the spin 2 quantum mechanical operators, in particular $\mathcal{J}_z = \partial_\phi$. For equatorial orbits $J_x = J_y = 0$, so the relevant quantity is

$$\dot{J}_z = \frac{1}{16\pi}\Im\left\{\sum_m m\int_{-1}^{1} d\xi\ (r\,\dot{h}_m)^*(r\,h_m)\right\}\ . \tag{34}$$

Similarly, the linear momentum $\vec{P} = (P_x, P_y, P_z)$ flux can be computed from

$$\dot{P}_i = \frac{1}{16\pi}\int_{S_2} d\Omega\ n_i |r\,\dot{h}|^2\ , \tag{35}$$

where $n_i = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$. For equatorial orbits $P_z = 0$.

The horizon-absorbed GW energy and angular momentum are defined using the first law of black-hole mechanics [77, 78] $\frac{\kappa}{8\pi}\dot{A}_{\rm H} = \dot{M}_{\rm H} - \Omega_{\rm H}\dot{J}_{\rm H}$, where $\kappa = (r_+ - M)/(r_+^2 - a^2)$ is the surface gravity and $\Omega_{\rm H} = a/(2Mr_+)$ is the angular velocity of the horizon. Considering the equations for the horizon generator dynamics, the variation of the horizon mass and angular momentum can be expressed as [79, 80, 62],

$$\dot{M}_{\rm H} = \frac{1}{16\pi}\int dS\ \sigma^{AB}\mathcal{L}_t\gamma_{AB}\quad ,\quad \dot{J}_{\rm H} = \frac{1}{16\pi}\int dS\ \sigma^{AB}\mathcal{L}_\phi\gamma_{AB}\ , \tag{36}$$

where dS is the horizon area element of the induced 2-metric, σ^{AB} is the 'shear tensor' (see Eq. (3.4) in [62]), and $\mathcal{L}_{t,\phi}$ are Lie derivatives with respect to the Killing vectors of the background. A similar equation holds for the area variation $\dot{A}_{\rm H}$, see Eq. (4.24) in [62]. The equations above are derived considering a particular coordinate system on the horizon (v, X^B) $(B = 2, 3)$, where $v = t + \int dr\Delta^{-1}(r^2 + a^2)$ is the advanced time as in (13) (connected to the ingoing Kerr time coordinate from (4) via $v = \tilde{t} + r$),

and $X^B = (\theta, \phi - \Omega_{\rm H} v)$ in terms of BL coordinates. Reference [62] specifies the final flux equations for the 2+1 fields, which read

$$\dot{M}_{\rm H} = \frac{r_+^2 + a^2}{4\kappa} \sum_m \left[2\kappa \int_{-1}^{1} d\xi \ |f^+_{{\rm H}\,m}|^2 - {\rm i}m\Omega_{\rm H} \int_{-1}^{1} d\xi \ \left(f^{+\,*}_{{\rm H}\,m} f^-_{{\rm H}\,m} - f^+_{{\rm H}\,m} f^{-\,*}_{{\rm H}\,m} \right) \right] , \tag{37}$$

$$\dot{J}_{\rm H} = -\frac{r_+^2 + a^2}{4\kappa} \sum_m {\rm i}m \left[\int_{-1}^{1} d\xi \ \left(f^{+\,*}_{{\rm H}\,m} f^-_{{\rm H}\,m} - f^+_{{\rm H}\,m} f^{-\,*}_{{\rm H}\,m} \right) \right] . \tag{38}$$

The complex quantities $f^{\pm}_{{\rm H}\,m}$ are defined as integrals of the m-mode components of the Weyl scalar ψ_0 at the horizon

$$f^+_{{\rm H}\,m}(v,\theta) = -e^{\kappa v} \int_v^\infty dv' e^{-(\kappa - {\rm i}m\Omega_{\rm H})v'} \psi_{0\ m}(v', r_+, \theta) \tag{39}$$

$$f^-_{{\rm H}\,m}(v,\theta) = -\int_{-\infty}^{v} dv' e^{{\rm i}m\Omega_{\rm H} v'} \psi_{0\ m}(v', r_+, \theta) \ . \tag{40}$$

The Weyl scalar $\psi_{0\ m}$ is understood as the one defined by the Hawking-Hartle tetrad. Our $s = +2$ TE master variable must be divided by $(4r_+(r_+^2 + a^2)^2)$ because we use the tetrad in Eq. (17) and rescale by r. Note also that $f^+_{{\rm H}\,m}(v)$ depends on the future behavior of the field. As mentioned in [62], the formalism is not yet optimal for 2+1 simulations; a certain drawback for its use in our setup will be pointed out in Sec. 8. Nonetheless, we employ it for the first time (to our knowledge) in this work and leave the development of a more practical method to the future.

4. Numerical method

In Sec. 4.1, we describe the numerical discretization of the TE (see also [15]). In Sec. 4.2 RT and HH coordinates are compared in their efficiency. Strategies for implementing the Dirac δ-functions are discussed in Sec. 4.3.

4.1. Discretization

For numerical integration the 2+1 TE is written as a first-order in time, and second-order in space system with reduction variables $u = \{\psi, \ \partial_\tau \psi\}$. The domain $(x, \theta) \in [x_+, x_S] \times (0, \pi)$, where x is the radial coordinate and x_+ (x_S) is the location of the horizon (null infinity), is uniformly discretized with $N_x \times N_\theta$ points. The spatial derivatives are represented by finite differences up to eighth order of accuracy. The stencils in the radial direction are centered in the bulk of the domain and lop-sided/sided at the boundaries (we also tested ghost points at radial boundaries filled by extrapolation and found no advantage). The angular grid is staggered and ghost points are employed to implement the boundary conditions on the axis. The ghost points are filled according to the parity condition $\pi = (-1)^{m+s}$. Artificial dissipation operators are implemented but not used for our results unless mentioned explicitly.

A standard fourth order Runge-Kutta integrator is employed for time advancing the solution. The time step is chosen according to a CFL condition of type $\triangle t = C_{\rm CFL} \min(h_x, h_\theta)$, where h_x is the grid spacing in direction x and the factor $C_{\rm CFL}$ accounts for the maximum coordinate speed of the PDE system. Even the most expensive simulations of this work (see Sec. 7) were performed on a standard

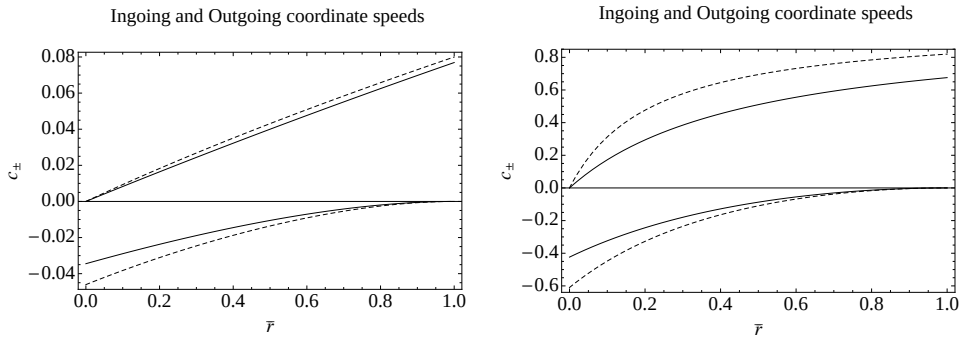

Figure 2. Visualization of ingoing (c_-, negative) and outgoing (c_+, positive) radial coordinate speeds. Setting $a = 0$, on the left panel we use HH_1 (solid) and RT (dashed) coordinates and on the right panel we use HH_{10} (solid) and HH_{20} (dashed) coordinates. The incoming (with respect to the domain) characteristic speeds vanish at both boundaries (horizon and scri) so no boundary conditions are needed. Note the radial domains are rescaled to $[0, 1]$ for the comparison.

Linux desktop computer using a serial implementation and the GNU C compiler. For example, EOB insplunge simulations at the typical resolution $N_x \times N_\theta = 3600 \times 160$ took about 2 weeks to reach the final time $5000\,M$.

The angular integrals for ℓ-mode projections are computed with the Simpson rule, while integrals in time with the trapezoidal rule. The integrals for the absorbed fluxes are calculated in post-processing; accuracy is thus dependent on the sampling of the output in time and angular direction. Note that the $f^+_{\mathrm{H}\,m}$ integrals at a given time v_* are mostly determined by the master function at times $(v - v_*) \sim \kappa^{-1}$, due to the exponential function. Therefore we restrict the integrations to times v such that $e^{-\kappa(v-v_*)} > 10^{-6}$. Since $\kappa \in [0.25, 0.35]$ (for $|\hat{a}| \in [0, 0.9]$) the interval is typically $v \in [v_* - 50, v_* + 50]$. We use Gaussian quadratures in the angular direction interpolating to ~ 50 Gauss-Lobatto output points, and manually vary the time output sampling until results are satisfactory (see Sec. 6.1).

4.2. Numerical efficiency of background coordinates

Both RT and HH coordinates have desirable properties for numerical treatment: (i) the horizon and scri are included in the computational domain, (ii) outgoing radial coordinate speeds c_+ vanish at the horizon and ingoing radial coordinate speeds c_- vanish at scri, so no particular treatment at the boundaries is needed (see Fig. 2).

The HH_S coordinates, however, are more flexible due to the presence of the free parameter S, and can therefore be tuned to be more efficient than the RT coordinates. To understand the effect of this parameter on the simulations, it is useful to view hyperboloidal surfaces as mediating between characteristic and Cauchy surfaces [16].

Geometrically, S is inversely correlated with the mean curvature of the foliation. Remember that the asymptotic mean curvature vanishes for Cauchy surfaces and is unbounded for characteristic ones. Therefore, a large S gives a more Cauchy-like behavior, whereas a small S gives a more characteristic-like behavior.

Numerically, the mean extrinsic curvature is related to the characteristic coordinate speeds and the spatial wavelength of waves propagating across the grid [16]. A large S implies a low characteristic speed and a small spatial wavelength,

whereas a small S implies a high characteristic speed and a large spatial wavelength. The trade-off is between the time step size restricted by the CFL condition due to the characteristic coordinate speeds and spatial resolution restricted by the wavelengths to be represented. The free parameter allows us to find a good balance between the time stepping and the spatial resolution requirements.

To find this balance, we compared the performances of RT and HH_S coordinates with $S = 1, 10, 20$ for $a = 0$ in wave scattering numerical experiments (no particle, as in [15]), geodesic dynamics (Sec. 6), and EOB insplunge simulations (Sec. 7) using the finite difference algorithm described in Sec. 4.1. Note that the RT slices have high curvature, which implies that there are hardly any waves to be resolved on the grid but the time stepping requirement becomes very restrictive. The HH_1 and the RT coordinates behave similarly with respect to numerical efficiency. Increasing S allows us to use larger time steps, but the number of waves on the spatial grid also increases. When S is too large (above $S = 20$), waves get blue shifted near the compactification boundary and artificial dissipation becomes necessary for stability. Artificial dissipation not only reduces numerical accuracy, but more importantly, it adds a significant computational cost. Therefore it is preferable to choose a value for S that does not require dissipation.

We found that $S = 10$ provides a good balance between time stepping and spatial resolution requirements without artificial dissipation. The HH_{10} coordinates give a speed up of ~ 2 with respect to HH_1 and RT coordinates. We adopted HH_{10} in all the simulations presented in the following.

4.3. Representation of δ functions

A key point in the numerical algorithm is the discrete representation of the distributional δ-functions appearing in the source term. The main requirements are numerical stability and accuracy. We implemented and tested two different methods.

The first method is the narrow Gaussian representation,

$$\delta(x - q(t)) \to \delta_\sigma(x - q(t)) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - q(t))^2}{2\sigma^2}\right] , \tag{41}$$

where $\sigma \sim n_\sigma h \ll M$, $n_\sigma \in \mathbb{N}$ as in e.g. [49, 52]. This method is very simple, smooth, and analytical but, in principle, computationally expensive since (i) the Gaussian must be well resolved on the grid ($n_\sigma \gtrsim 4$); (ii) exponential functions must be often evaluated during the time evolution.

The second method is a $2n$-points discrete δ as described in [81, 82, 10]. The prescriptions discussed therein comprise (a) an order $\mathcal{O}(h^2)$ with $n = 1$, (b) an order $\mathcal{O}(h^4)$with $n = 2$ and (c) an order $\mathcal{O}(h^2)$ with variable $2n$-points representation for the δ-function. Let us sketch the main ideas, for more details we refer to Sec. III.A of [10]. Assume that the position of the particle α lies between two grid points, $\alpha \in [x_k, x_{k+1}]$. Then, the discrete δ has support only for $2n$-points δ_i around α. These values are given requiring that the integral properties of the δ-function, e.g.

$$f(\alpha) = \int dx f(x)\delta(x - \alpha) \approx \sum h f_i \delta_i , \tag{42}$$

are preserved also on the discrete level. If by chance $\alpha = x_k$, setting $\delta_k = 1/h$ and $\delta_i = 0$ elsewhere solves the problem. In general, α does not lie on a grid point so

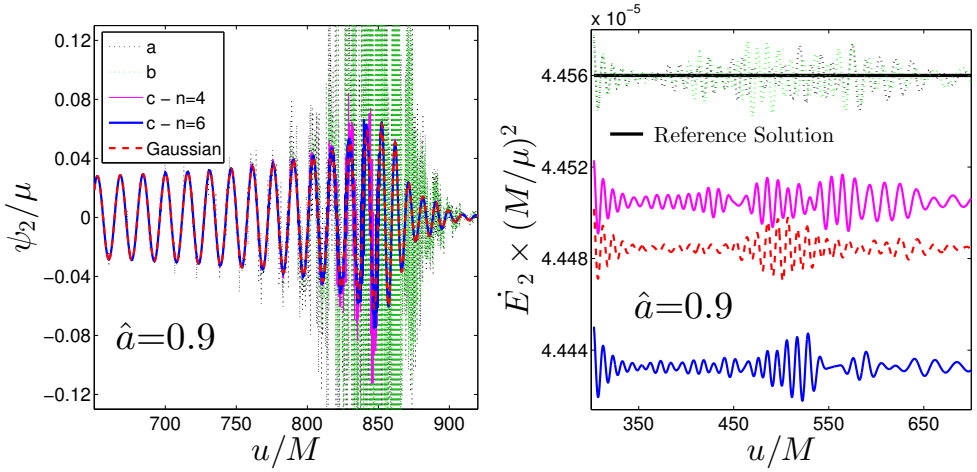

Figure 3. Comparison of different δ-function representations in the two different cases of an EOB insplunge (left panel) and a circular orbit (right panel). Note the two panels shows different quantities. The left panel shows the evolution of the $m = 2$ master variable $\Re(\psi_2)$. The right panel shows the evolution of the GW fluxes. The left panel illustrates how the discrete representations lead to instabilities in the simulation if few points are employed. Options (a), (b) and (c) with $n = 4$ described in Sec. 4.3 are very noisy, while option (c) with $n \gtrsim 6$ gives results comparable to the smooth, analytical Gaussian representation. The right panel illustrates the accuracy of the discrete representations. The reference solution is taken from the frequency domain results of Ref. [10] and shown as a black solid line. Option (c) with $n \gtrsim 6$ is less accurate than the Gaussian representation. Options (a) and (b) are instead very accurate in this case. Thus we use a Gaussian for any trajectory moving in the (r, θ) plane but the discrete representation of option (a) for circular orbit simulations.

that interpolation has to be used. Linear interpolation leads to option (a) and (c), the more accurate cubic interpolation yields option (b). The discrete representation uses (a) a 2-points support, (b) a 4-points support, and (c) a variable $2n$-points support. For example, considering a 2-points support, linearly interpolating at α, and enforcing Eq. (42) leads to option (a)

$$\delta(x - q(t)) \to \delta_{(a)\ j} = \begin{cases} \gamma\, h^{-1} & , j = k \\ (1-\gamma)\, h^{-1} & , j = k+1 \\ 0 & , \text{otherwise} \ , \end{cases} \tag{43}$$

where $\gamma = (x_{k+1} - \alpha)/h$. Similar formulas can be derived for the first two derivatives [10]. Overall, this method is expected to be computationally more efficient than the Gaussian. However (i) high-order accuracy requires larger stencils, increasing the cost; (ii) too narrow/lower order representation may lead to instabilities, as we shall discuss next (see also [10, 11]).

We have tested the different δ-representations for various numerical setups in the cases of circular orbits and EOB insplunge simulations. Summarizing our findings, the tests indicate that the discrete δ method in its simplest representation, option (a), is efficient and accurate for simulating a source that is effectively *not* moving on the computational domain, e.g. circular, geodesic orbits. In cases like inspiral-plunge,

instead, for the same computational cost the most accurate, stable simulations are achieved with the Gaussian method. An example is given in Fig. 3 where we plot for the different options on the left panel the $m = 2$ field $\Re(\psi_2)$ for an EOB insplunge configuration and on the right the GW energy fluxes from circular orbit experiments. In case the source is effectively moving on the computational domain (left panel) the discrete representations (a), (b) and (c) with $n = 4$ lead to instabilities towards merger. A larger support makes the discrete representation smoother and only option (c) with $n \gtrsim 6$ gives results comparable to the smooth, analytical Gaussian representation. On the contrary, in case of circular orbits (right panel) options (a) and (b) are both more accurate and efficient. The accuracy of the circular fluxes is here evaluated against a reference solution taken from the frequency domain results of Ref. [10] and shown as a black solid line. Note that here the Gaussian is more accurate than option (c) with $n = 6$. Finally, because the insplunge motion considered later on in this work is fixed on the $\theta = \pi/2$ plane, the best option is to use a combination of the two methods: a Gaussian representation in radial direction and a discrete representation in θ direction (we adopted option (b)).

5. Dynamics of an inspiraling and plunging, nonspinning point-particle

Let us now discuss our prescription to model the dynamics of an inspiraling and plunging particle. The dynamics of the particle is described using an Hamiltonian approach for the conservative part and an analytical radiation reaction force that accounts for the losses of angular momentum through GWs for the dissipative part. This analytic, EOB-resummed, but approximate, radiation reaction force will be discussed in detail and we will review two different choices that were proposed in the literature to define it. We will use the TE data to argue that one of the two may be preferable for counterrotating orbits.

It is convenient to describe the Hamiltonian dynamics of a particle (a short review is attached in Appendix B) using dimensionless quantities like $\hat{H} \equiv H/\mu$. We also denote "reduced" quantities with respect to the background with a hat, e.g. $\hat{t} \equiv t/M$, although they coincide for $M = 1$ as used in the simulations. We focus on equatorial motion only, so that the spin of the black hole is either aligned (corotating case) or antialigned (counterrotating case) with the orbital angular momentum. The particle Hamiltonian, Eq. (B.15), specified to the equatorial motion reads,

$$\hat{H} \equiv \hat{H}_{\rm SO} + \hat{H}_{\rm orb} = \omega p_\phi + \sqrt{A\left(1 + \frac{p_\phi^2}{\varpi^2}\right) + A\frac{\Delta}{\Sigma}p_{\hat{r}}^2}\,, \tag{44}$$

where $p_\phi \equiv P_\phi/(\mu M)$, $\hat{r} = r/M$, $p_{\hat{r}} \equiv P_{\hat{r}}/\mu$ and the functions $(\omega, A, \Delta, \Sigma, \varpi)$ are given by Eqs. (B.10)-(B.14) specified to the equatorial plane ($\theta = \pi/2$). In Eq. (44) we separate the Hamiltonian in a formally "pure orbital" part $\hat{H}_{\rm orb}$ and a "pure spin-orbit" part $\hat{H}_{\rm SO}$. This formal separation will be used in Sec. 7.2 below. In the general case of nonconservative dynamics the radiation reaction force enters in both the Hamiltonian equations for the momenta, yielding both energy and angular

momentum losses. The Hamilton's equations read

$$\dot{\hat{r}} = \partial_{p_{\hat{r}}} \hat{H}, \tag{45}$$

$$\dot{\phi} = \partial_{p_\phi} \hat{H} = \hat{\Omega}, \tag{46}$$

$$\dot{p}_{\hat{r}} = -\partial_{\hat{r}} \hat{H} + \hat{\mathcal{F}}_{\hat{r}}, \tag{47}$$

$$\dot{p}_\phi = \hat{\mathcal{F}}_\phi \ , \tag{48}$$

where the overdot stands for $d/d\hat{t}$. In the following, we neglect the radial flux contribution and set $\hat{\mathcal{F}}_{\hat{r}} = 0$. The reason is that a robust strategy to resum the $\hat{\mathcal{F}}_{\hat{r}}$ in the strong-field regime is, at present, not available, even though the PN expansion of $\hat{\mathcal{F}}_{\hat{r}}$ is known up to 2PN order [83]. As a consequence, $\hat{\mathcal{F}}_{\hat{r}}$ may be ill-behaved even for non-extremal values of $\hat{a}$. Although $\hat{\mathcal{F}}_{\hat{r}}$ has been considered at leading, Newtonian order in [60], we believe it is still premature to include this term in the analytical model of the dynamics. We postpone to future work a detailed analysis of its properties needed to devise a robust resummation procedure.

By contrast, the analytic expression of $\hat{\mathcal{F}}_\phi$ has been thoroughly used in recent years. Nonetheless, a careful inspection of the literature indicates that there are two ways of writing the flux during the plunge. During the latter, the Kepler's constraint, $1 = \hat{\Omega}^2 \hat{r}^3$, is not satisfied [84, 85] (only valid for the quasi-adiabatic, circular inspiral) and the two descriptions differ in their way of relaxing the constraint. We will contrast the original implementation of $\hat{\mathcal{F}}_\phi$ (see Ref. [86] and references therein), here specified to the test-mass case in Kerr spacetime, to the different one proposed in [60, 64, 61]. The differences are essentially in the choices of the arguments of certain functions so to incorporate the non-Keplerian behavior during the plunge.

Let us discuss these differences in detail. In the circular approximation the mechanical angular momentum loss is given by

$$\hat{\mathcal{F}}_\phi = -\frac{1}{\nu \hat{\Omega}} \dot{E} \ , \tag{49}$$

where the energy flux $\dot{E}$ is resummed according to the multipolar waveform resummation introduced in [51] for nonspinning binaries and extended in [87] to the spinning case. We consider multipoles up to $\ell = 8$ and the energy flux is given by

$$\dot{E} = \sum_{\ell=2}^{8} \sum_{m=1}^{\ell} \dot{E}_{\ell m} = \frac{1}{8\pi} \sum_{\ell=2}^{8} \sum_{m=1}^{\ell} (m\hat{\Omega})^2 |\mathcal{R} h_{\ell m}(x)|^2 \ , \tag{50}$$

where we used the usual PN-expansion variable $x = \hat{\Omega}^{2/3}$ and $\mathcal{R}$ is the distance ‖ from the source. The multipoles $h_{\ell m}(x)$ are written in factorized form as

$$h_{\ell m}(x) = h^{(N,\epsilon)}_{\ell m}(x) \hat{S}^{(\epsilon)} T_{\ell m}(\hat{\Omega}) \left[\rho_{\ell m}(x)\right]^\ell e^{\mathrm{i}\delta_{\ell m}} \ , \tag{51}$$

where $\epsilon = (0, 1)$ is the parity of $\ell + m$,

$$h^{(N,\epsilon)}_{\ell m}(x) = \frac{\nu}{\mathcal{R}} n^{(\epsilon)}_{\ell m} c_{\ell+\epsilon} x^{\frac{\ell+\epsilon}{2}} Y^{\ell-\epsilon,-m}\left(\frac{\pi}{2}, \phi\right) \tag{52}$$

is the Newtonian contribution, $Y^{\ell m}(\theta, \phi)$ are the scalar spherical harmonics, and the functions $n^{(\epsilon)}_{\ell m}$ and $c_{\ell+\epsilon}$ are given in Eqs. (4a), (4b) and (5) of Ref. [51]. $T_{\ell m}(\hat{\Omega})$ is the

‖ Here, we use $\mathcal{R}$ instead of r, as in Eq. (31), to avoid confusion with the relative separation r.

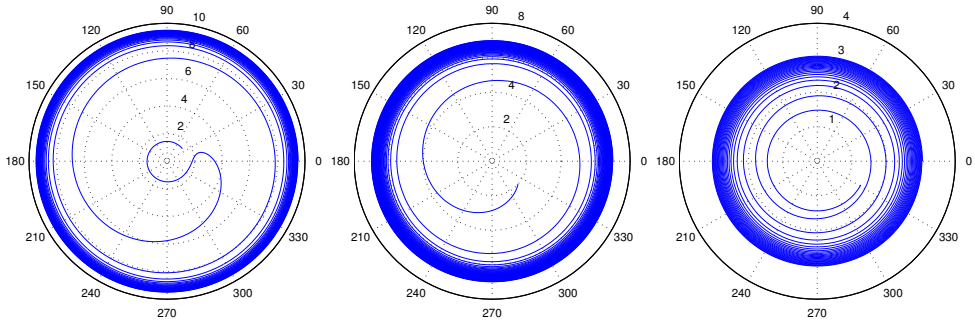

Figure 4. Examples of insplunge trajectories for $\hat{a} = -0.9$ (left), $a = 0$ (center), and $\hat{a} = +0.9$ (right). The last stable circular orbits (LSO) are located, respectively, at $\hat{r}_{\rm LSO} \simeq 8.697, 6.0, 2.424$. The light rings (LR) are located at $\hat{r}_{\rm LR} \simeq 3.910, 3.0, 1.558$.

tail factor, the source term $\hat{S}^{(\epsilon)}$ is specified later, and $\rho_{\ell m}$ and $\delta_{\ell m}$ are the residual amplitude and phase corrections. We will specify their structure below. Note here that the tail factor $T_{\ell m}(\hat{\Omega})$ explicitly depends on the orbital frequency $\hat{\Omega}$, while all other factors depend on $x = \hat{\Omega}^{2/3}$ because of Kepler's constraint for circular orbits. On a Kerr background, the Kepler constraint may be written as

$$\hat{r}^3 \left(1 + \hat{a}\hat{r}^{-3/2}\right)^2 \hat{\Omega}^2 = \hat{r}_\Omega^3 \hat{\Omega}^2 = 1 \ , \tag{53}$$

where we defined

$$\hat{r}_\Omega = \hat{r} \left(1 + \hat{a}\hat{r}^{-3/2}\right)^{2/3} \ . \tag{54}$$

Since the circular Kepler constraint is not satisfied during the plunge, the usual practice is to modify the standard $\hat{\mathcal{F}}_\phi$ derived along circular orbits so that it is not imposed explicitly in the analytical expressions during the plunge (though it is automatically recovered during the quasi-circular inspiral). In the literature, one finds two prescriptions to construct $\hat{\mathcal{F}}_\phi$'s such that the circular Kepler's constraint is relaxed during the plunge.

(A) – *the v_ϕ-prescription:* The prescription introduced in [50, 86] is to impose that the argument x entering in the functions $h_{\ell m}^{(N,\epsilon)}(x)$ and $\rho_{\ell m}(x)$ is the non-Keplerian (squared) tangential velocity, $x = v_\phi^2$, where

$$v_\phi = r_\Omega \hat{\Omega} \ . \tag{55}$$

When the black hole is spinning, semi-integer powers of x appear in the $\rho_{\ell m}$'s. This introduces a subtlety when the particle is counterrotating with respect to the black hole, so that, to keep the correct sign, one should see the $\rho_{\ell m}$'s as functions of $\sqrt{x}$ (with the correct sign) and not of x. The sources $\hat{S}^{(\epsilon)}$ are computed along the dynamics and we use $\hat{S}^{(0)} = \hat{H}$ and $\hat{S}^{(1)} = p_\phi v_\phi$ (the Newton-normalized orbital angular momentum). The $\rho_{\ell m}$-functions are given by Eqs. (29a)-(29i) of Ref. [88], augmented by the full 5PN-accurate terms computed in Ref. [89]. Note that the PN-based calculations have been pushed to 22 PN for the nonspinning case [35] and to 20 PN for the spinning case [36], so that more analytical information is in principle available and will be incorporated in the model in the future.

(B) –*the v_Ω-prescription:* Following Ref. [60], the first difference with (A) is that the argument of the $h^{(N,\epsilon)}_{\ell m}(x)$ prefactors for the $(2,1)$ and $(4,4)$ multipoles differs from the general prescription introduced above. The straightforward rewriting of Eqs. (13a) and (13b) of Ref. [60] in our notation ¶ gives

$$x = v_\phi^2 \qquad (\ell, m) \neq (2,1),\ (4,4) \tag{56}$$

$$x = r_\Omega^{-\frac{2}{\ell+\epsilon}} v_\phi^{\frac{2(\ell+\epsilon-2)}{\ell+\epsilon}} \qquad (\ell, m) = (2,1),\ (4,4)\ , \tag{57}$$

with v_ϕ defined by Eq. (55). Explicitly for $(\ell, m) = (2,1)$ $(\epsilon = 1)$ and $(\ell, m) = (4,4)$ $(\epsilon = 0)$ the expression Eq. (57) yields

$$x = \hat{\Omega}^{2/3} = v_\Omega^2 \qquad (\ell, m) = (2,1), \tag{58}$$

$$x = r_\Omega^{1/2}\hat{\Omega} = r_\Omega^{1/2} v_\Omega^3 \qquad (\ell, m) = (4,4), \tag{59}$$

where we have introduced the *circular* velocity $v_\Omega = \hat{\Omega}^{1/3}$. A second difference to (A) is that the odd-parity source is normalized to the inverse of v_Ω instead of v_ϕ; namely one uses $\hat{S}^{(1)} \equiv p_\phi v_\Omega$. In other words, the "Newtonian" angular momentum used to normalize the odd-parity source is computed imposing the Kepler's constraint. Finally, the argument of the $\rho_{\ell m}$ functions given by Eqs. (29a)-(29i) of Ref. [88] is taken to be v_Ω and not v_ϕ. Note that here the $\rho_{\ell m}$'s contain all the 5PN-accurate nonspinning terms. This is done for consistency with case (A) and differs from Ref. [60] where only part of the 5PN-accurate information was retained.

In summary, prescription (A) is "less Keplerian" than prescription (B) because of the explicit relaxation of the Kepler's constraint when choosing the argument of the various functions. A priori, one expects (A) to give a more accurate representation of the "actual" flux when the deviations from circularity are larger; notably when the black hole spin is high and antialigned with p_ϕ. However, (A) and (B) should be essentially equivalent when the plunge is more circularized, that is, when the black hole spin is high and aligned with p_ϕ. In this work, we consider (A) as our *standard* way of implementing the non-Keplerian behavior in an analytical expression of the radiation reaction. On top of its simplicity and uniformity of implementation (the circular velocity $v_\Omega = \hat{\Omega}^{1/3}$ is ubiquitously replaced by v_ϕ and Kepler's constraint is always relaxed), it offers a more consistent analytical representation of the loss of mechanical angular momentum during the plunge.

Before contrasting the two flux prescriptions with TE fluxes, we show in Fig. 4 examples of particle trajectories, as obtained with (A), for $\hat{a} = -0.9$ (left) $\hat{a} = 0$ (center) and $\hat{a} = +0.9$ (right). The qualitative differences of the dynamics as a function of the rotation parameter $\hat{a}$ are mainly due to the spin-orbit coupling in the Kerr potential for geodesics. The latter determines the positions of the last stable orbit (LSO), $\hat{r}_{\rm LSO}(\hat{a})$, and the light ring (LR), $\hat{r}_{\rm LR}(\hat{a})$. Compared to the nonrotating case ($\hat{a} = 0$), positive spins (aligned to the orbital angular momentum) move the LSO and LR closer to each other and closer to the horizon, e.g. $\hat{r}_{\rm LSO}(\hat{a} > 0) < \hat{r}_{\rm LSO}(0)$. As a consequence, the plunge becomes progressively "more circular" as $\hat{a} \to 1$, and the radial momentum of the particle attains a minimum for $\hat{a} = 1$. Negative spins (antialigned with p_ϕ) move the LSO and the LR farther from each other, and farther from the horizon, i.e. $\hat{r}_{\rm LSO}(\hat{a} < 0) > \hat{r}_{\rm LSO}(0)$. Retrograde plunges are characterized

¶ Reference [60] uses the notation V_ϕ^ℓ, where the ℓ is a *label*, and *not* an exponent, to denote $x^{(\ell+\epsilon)/2}$.

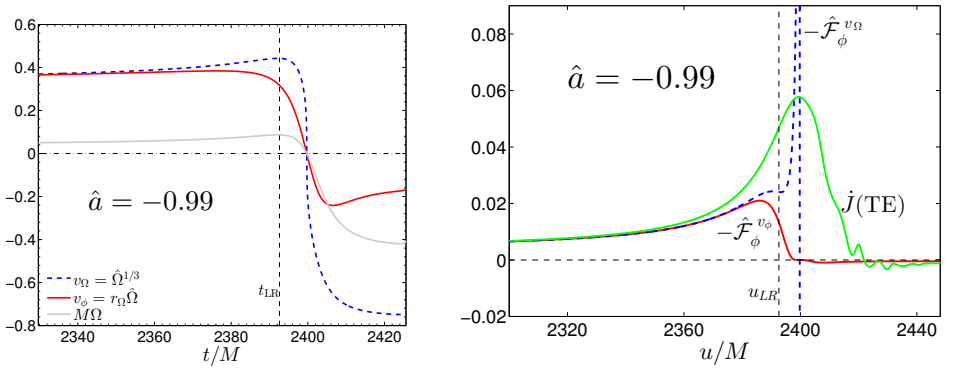

Figure 5. Comparing two different prescriptions for the mechanical angular momentum loss, $-\hat{\mathcal{F}}_\phi^{v_\phi}$ and $-\hat{\mathcal{F}}_\phi^{v_\Omega}$, with the TE angular momentum flux $\dot{J}^{\rm TE}$ (all values normalized with (M/μ^2)). We set $\hat{a} = -0.99$ and focus on late-plunge and merger phase. The left panel compares v_Ω and v_ϕ; the right panel compares the fluxes. The vertical dashed line marks the light-ring crossing. The prescription $\hat{\mathcal{F}}_\phi^{v_\phi}$ captures the correct qualitative behavior of the TE flux. Remaining differences are due to next-to-quasi-circular effects not included in the modeling of $\hat{\mathcal{F}}_\phi$ and to QNMs ringdown oscillations.

by a turning point, $\hat{\Omega} = 0$, after which $\hat{\Omega}$ changes sign until the particle locks to the horizon. This effect is due to frame dragging and happens at $\hat{r} < \hat{r}_{\rm LR}$. The radial momentum of the particle during plunge, as a function of the rotation parameter, attains a maximum for $\hat{a} = -1$.

Now that we have discussed qualitatively the structure of insplunge trajectories, let us motivate why we use prescription (A) for the radiation reaction force. We focus on the case $\hat{a} = -0.99$ and perform a consistency comparison between EOB and TE fluxes. The particle inspirals from $\hat{r}_0 = 9.5$. The LSO is at $\hat{r}_{\rm LSO} = 8.97$, the light-ring at $\hat{r}_{\rm LR} = 3.99$ and the outer horizon at $\hat{r}_+ = 1.14$. The left panel of Fig. 5 compares the time evolution of $\hat{\Omega}$, v_Ω, and v_ϕ close to merger; the light-ring crossing time is identified by the dashed vertical line. The right panel of the figure presents, versus time, the triple comparison between: (i) the mechanical angular momentum loss $-\hat{\mathcal{F}}_\phi^{v_\phi}$ computed via prescription (A); (ii) $-\hat{\mathcal{F}}_\phi^{v_\Omega}$ obtained from prescription (B); (iii) the TE angular momentum flux computed from the waves extracted at future null infinity along dynamics (A). Figure 5 highlights that the differences in the analytical fluxes only occur in the very late plunge phase, very close to the LR crossing (when $\hat{r} = \hat{r}_{\rm LR}$) at $t_{\rm LR}/M = 2392.6$. As a consequence, the plunge dynamics is unaffected by the differences between prescriptions (A) and (B). This is why in Fig. 5 we report only one TE curve: the dynamics is practically the same and so are the TE fluxes. However, the comparison of Fig. 5 indicates that $-\hat{\mathcal{F}}_\phi^{v_\phi}$ is qualitatively closer to the numerical flux than $-\hat{\mathcal{F}}_\phi^{v_\Omega}$ around and after the light-ring crossing. The main reason for this difference resides in the behavior of $\hat{\Omega}$ and its effect on the $(2,1)$ and $(4,4)$ Newtonian contributions to the analytical flux, that are proportional to certain powers of Eqs. (58) and (59). The difference between $-\hat{\mathcal{F}}_\phi^{v_\phi}$ and $-\hat{\mathcal{F}}_\phi^{v_\Omega}$ is then easily explained from the different behavior of v_Ω and v_ϕ. Referring to the left panel of Fig. 5, one sees that, after the $\hat{\Omega} = 0$ point, v_Ω keeps decreasing monotonically, while v_ϕ is limited and eventually its derivative changes sign. Since $\dot{E}_{21} \propto x^6$ and $\dot{E}_{44} \propto x^7$ one understands

the origin of the unphysical growth of the $-\hat{\mathcal{F}}_\phi^{v_\Omega}$ when $x = v_\Omega^2$ for $(2,1)$ and by $x = r_\Omega^{1/2} v_\Omega^3$ for $(4,4)$, since one gets $\dot{E}_{21} \propto \hat{\Omega}^4$ and $\dot{E}_{21} \propto r_\Omega^{7/2}\hat{\Omega}^7$. By contrast, the milder behavior of v_ϕ when $\hat{\Omega} \to 0$ produces limited fluxes $(\dot{E}_{21}, \dot{E}_{44})$ and a globally more consistent behavior of $-\hat{\mathcal{F}}_\phi^{v_\phi}$. It is also worth noting the consistency between $-\hat{\mathcal{F}}_\phi^{v_\phi}$ and $\dot{J}$ after the turning point $\hat{\Omega} = 0$, with both quantities having approximately the same magnitude. On the contrary, $-\hat{\mathcal{F}}_\phi^{v_\Omega}$ is very different there because of the vertical tangent of v_Ω at $\hat{\Omega} = 0$. The residual differences between $-\hat{\mathcal{F}}_\phi^{v_\phi}$ and $\dot{J}$ seen in the right panel of Fig. (5) are due to: (i) next-to-quasi-circular terms depending explicitly on p_{r_*} that are not included in the analytical modeling and (ii) effects of the ringdown. Though we will quantify the amount of noncircularity of the dynamics versus $\hat{a}$ in Sec. 7 and connect it with the multipolar structure of the waveform around the light-ring crossing, the proper modeling of next-to-quasi-circular effects in the EOB waveform and flux and ringdown is out of the scope of this paper.

When the black hole spin is aligned with the orbital angular momentum, we shall show below, in Sec. 7.1, that the differences between $-\hat{\mathcal{F}}_\phi^{v_\Omega}$ and $-\hat{\mathcal{F}}_\phi^{v_\phi}$ become much smaller and practically negligible as $\hat{a} \to 1$. Actually, when $\hat{a} \to 1$ one finds that they are equally inaccurate with respect to TE fluxes, because of missing higher PN (spin-dependent) terms in the expansion of the $\rho_{\ell m}$'s.

6. Code tests: accuracy of gravitational waveforms and fluxes

Before discussing in detail the structure of the multipolar waveforms for $\hat{a} \neq 0$ and their properties when $|\hat{a}| \to 1$, we present several tests of our new computational infrastructure. In Sec. 6.1 we first consider circular, equatorial orbits for different values of $\hat{a}$, calculate the GW energy fluxes emitted to future null infinity and to the horizon, and compare with the results of Ref. [10, 14]. In Sec. 6.2 we discuss the radial geodetic plunge (no radiation reaction) for $\hat{a} = 0$ and compare with the RWZE results of [52]. This test also gives us the opportunity to discuss the numerical treatment of the source during the plunge. Self-convergence is studied in Sec. 6.3 for insplunge waveforms. We show that the expected convergence rate is attained already at low resolutions and the absolute phase errors are small. Note that this is a challenging test of the main physical application of the method. In Sec. 6.4 we compare the multipolar insplunge waveform for a nonrotating background computed with the 2+1 *Teukode* and with the 1+1 RWZE code from [54, 56, 53]. All the fluxes shown are normalized by the appropriate powers of M, μ.

6.1. GW fluxes from circular orbits

This test considers circular equatorial trajectories at various radii $\hat{r}_0 = 4, 6, 8, 10$, for $\hat{a} = 0, 0.9$ and $m = 2, 3$. GW energy fluxes at scri are compared with the frequency domain results kindly provided to us by Scott Hughes using an improved version of his frequency domain code from [10, 14].

GW fluxes at the horizon are calculated in two different ways: (i) using the formulas presented in Sec. 3.3 and performing in postprocessing the integral of Eq. (37), and (ii) using the frequency domain formulas of [79]. The latter calculation is possible because we are considering circular orbits. It is performed as a check of method (i) to verify its accuracy and robustness. The data of the horizon fluxes are compared with the corresponding ones also provided to us by Scott Hughes [14]. The

Table 1. GW energy fluxes at scri, $\dot{E}^{\infty}_m$, and at the horizon, $\dot{E}^{H}_m$, for circular, equatorial orbits at various $\hat{r}_0$ for $m = 2, 3$ and background rotations $\hat{a} = 0.0, 0.9$. The values are normalized by $(M/\mu)^2$. Radii below the LSO are marked with $*$. The resolution used for the shown results is $N_x \times N_\theta = 2400 \times 200$. The horizon fluxes are computed with two different methods: the usual frequency domain formula [79] applicable in our time domain setup because of circular orbits, and the time domain formula in Eq. (37) (in brackets). $\Delta\dot{E}^{\infty,H}_m/\dot{E}^{\infty,H}_m$ are the percentual relative differences to the frequency domain values of [10, 14]. Note that our results include *all* the ℓ-mode contributions, while the reference solution truncates the sums at $\ell = 8$.

$\hat{a}$	m	$\hat{r}_0$	$\dot{E}^{\infty}_m$	$\Delta\dot{E}^{\infty}_m/\dot{E}^{\infty}_m[\%]$	$\dot{E}^{H}_m$		$\Delta\dot{E}^{H}_m/\dot{E}^{H}_m[\%]$
0	2	4*	8.580479 e-03	8.33e-03	5.64953e-04	(5.64849e-04)	2.03e-02
0	2	6	7.368338 e-04	3.58e-04	2.62484e-06	(2.62443e-06)	3.91e-03
0	2	8	1.650495 e-04	1.52e-03	1.09970e-07	(1.09953e-07)	4.00e-05
0	2	10	5.373492 e-05	2.75e-03	1.13139e-08	(1.13122e-08)	2.14e-03
0.9	2	4	2.661563 e-03	2.57e-03	-5.28423e-05	(-5.28346e-05)	4.72e-03
0.9	2	6	4.621241 e-04	2.81e-03	-3.98467e-06	(-3.98441e-06)	1.30e-03
0.9	2	8	1.254217 e-04	3.44e-03	-5.68006e-07	(-5.67988e-07)	9.13e-04
0.9	2	10	4.455909 e-05	3.49e-03	-1.19689e-07	(-1.19702e-07)	1.36e-03
0	3	4*	2.710318 e-03	7.95e-03	6.92585e-05	(6.92581e-05)	5.34e-03
0	3	6	1.459721 e-04	1.22e-02	5.41814e-08	(5.41814e-08)	8.86e-03
0	3	8	2.449258 e-05	1.31e-02	8.61375e-10	(8.61376e-10)	1.17e-02
0	3	10	6.434177 e-06	1.34e-02	4.69154e-11	(4.69155e-11)	1.28e-02
0.9	3	4	6.466345 e-04	1.37e-02	-3.00663e-06	(-3.00675e-06)	8.98e-03
0.9	3	6	8.042190 e-05	1.34e-02	-1.17094e-07	(-1.17111e-07)	1.11e-02
0.9	3	8	1.717198 e-05	1.35e-02	-1.00392e-08	(-1.00421e-08)	1.20e-02
0.9	3	10	5.043443 e-06	1.34e-02	-1.40038e-09	(-1.40118e-09)	1.21e-02

Table 2. GW energy fluxes for a circular, equatorial orbit at $\hat{r}_0 = 6$ for $\hat{a} = 0.9$ in the $l = m = 2$ mode at different finite extraction radii, for waves extrapolated using Eq. (62) and $K = 2$, and for waves at null infinity. The values are normalized by $(M/\mu)^2$.

$\hat{r}$	100	200	300	500	740	1000	Extrp. $(K = 2)$	$\mathscr{I}$
$\dot{E}_{22} \times 10^4$	4.546	4.595	4.604	4.608	4.610	4.610	4.611	4.611

relevant results are collected in Table 1. Using $N_x \times N_\theta = 2400 \times 200$ we reproduce three digits of the frequency domain data, that is, the agreement is around 0.01%. We mention that already a resolution of 1200×100 (runtime ~ 0.5 hours on a standard desktop machine) suffices to obtain agreement up to two digits. These results rely on the waveform extraction at scri. For completeness, Table 2 lists, for the $\ell = m = 2$ mode, the energy flux computed from waveforms extracted at finite radii as well as the value obtained by extrapolation (see Eq. (62) below), that coincides with the one computed from scri waveforms.

Similar results are obtained for the horizon fluxes. Table 1 shows that the time domain calculation in post-processing is indeed accurate. Note that the agreement with the frequency domain data is obtained also below the last stable orbit, e.g. $\hat{r}_0 = 4$ for $\hat{a} = 0$.

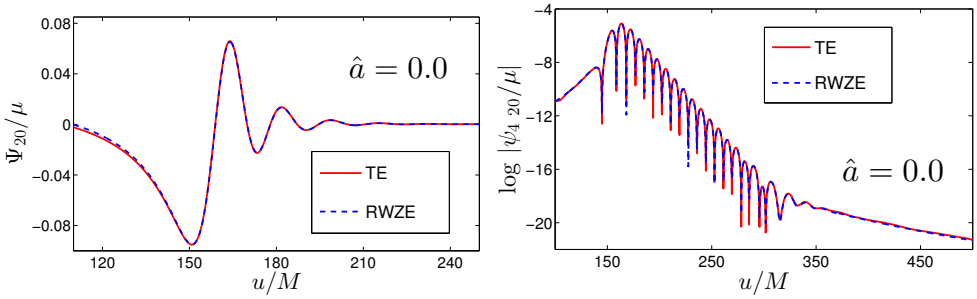

Figure 6. Comparison of 2+1 TE/1+1 RWZE code waveforms for geodetic radial infall dynamics. The particle is falling from $\hat{r}_0 = 25$ along the x-axis onto a Schwarzschild black hole. Left: The $\ell = 2$, $m = 0$ waveform $\Psi_{20}(u)$ (RWZE variable) at $\mathscr{I}$ (*Teukode*) and large finite radius (RWZE code). Right: The $\ell = 2$, $m = 0$ component of the Weyl scalar $\psi_{4\ 20}$ at $\mathscr{I}$ (*Teukode*) and large finite radius (RWZE code) in logarithmic scale.

6.2. Geodetic radial infall for $\hat{a} = 0$: TE vs. RWZE waveforms

This test considers a radial trajectory from $\hat{r}_0 = 25$ along the x-axis in Schwarzschild background ($\hat{a} = 0$). For simplicity, we focus only on the $\ell = 2$, $m = 0$ multipole. But contrarily to previous calculations of the waveform from the radial geodetic plunge along the z-axis [25], also the multipoles with $m \neq 0$ are switched on since the particle is moving in the equatorial plane.

The TE waveforms are compared with the RWZE ones computed as described in [52]. Waveforms from the RWZE code are extracted at large finite radius $\hat{r} \sim 2200$, TE waveforms are extracted at $\mathscr{I}$. The initial data for the RWZE code solve the linearized Hamiltonian constraint [25]. On the contrary, initial data for the TE are not solving the constraints and we trivially set $\psi = \dot{\psi} = 0$. This produces an initial burst of junk radiation which is radiated away after $\sim 200M$. In Fig. 6 we compare the outcome of the two codes in the $\ell = 2$, $m = 0$ RWZE variable $\Psi_{20}(u)$ (left) and the Weyl mode $\psi_{4\ 20}(u)$ (right). Despite the differences in the setup, we find visual agreement; quantitative differences are below a few percent. In particular the $\psi_{4\ 20}(u)$ variables (right panel) agree also during the tail phase and both codes capture the correct tail decay. The simulation of the correct tail phase requires artificial dissipation. The tail cannot be captured in the 2+1 RWZE variable because the integration to Ψ_{20} of the TE data produces inaccuracies.

Let us discuss the behavior of the particle at the horizon. In BL coordinates the source smoothly "switches-off" when approaching the horizon due to a red-shift effect driven by the term $d\lambda/dt \to 0$ in Eq. (26). If, in addition, the computational domain does not include the horizon (e.g. the tortoise coordinate is employed) no particular treatment is needed for the source approaching the horizon, see e.g. [12]. The situation is different in horizon penetrating coordinates. We observe a red-shift "shrinking" effect but the source does not approach zero towards the horizon, instead it reaches a finite limit. When the particle has reached the last point before the horizon, we remove the remaining half of the Gaussian by simply advecting it out of the computational domain. This procedure is somehow unphysical but does not affect the waveform since it involves very few points close to the horizon. Furthermore (i) the inspiral-plunge waveform is mainly determined by the particle crossing the light-ring

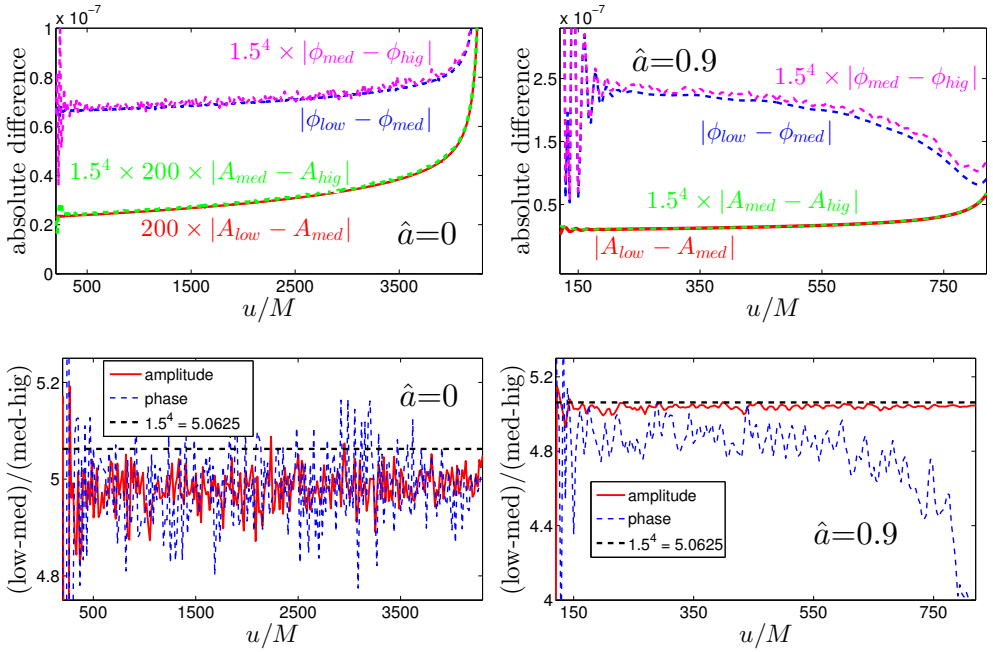

Figure 7. Self-convergence for the amplitude and phase at the extraction point $(\rho, \theta) = (10, \pi/2)$ for insplunge waveforms. Left: $\hat{a} = 0$, right: $\hat{a} = 0.9$. The triplet uses radial resolutions $N_x = (400, 600, 900)$, with $n_\sigma = (4, 6, 9)$ and $N_\theta = 30$. The expected scaling for the error is $1.5^4 \approx 5.0625$ for 4th-order finite differences. Top panels: absolute differences in phase $\Delta\phi$ and amplitude ΔA between various resolutions. The differences between medium and high resolution are rescaled by the expected factor assuming convergence and lay on top of the differences between low and medium resolution. For visualization the differences in amplitude are rescaled by an arbitrary factor 200 in the left panel. Bottom panels: ratios of absolute differences. Convergence is thus observed already at very low resolutions.

and entering the potential well, (ii) the ringdown part of the waveforms is essentially particle-independent.

6.3. Self-convergence of insplunge waveforms

To test the accuracy of the code, Fig. 7 shows self-convergence tests for $\hat{a} = 0$ (left) and $\hat{a} = 0.9$ (right) obtained at resolutions $N_x = (400, 600, 900)$ (all with $N_\theta = 30$), using 4th order finite differencing and a minimal number of points for the Gaussian, i.e. $\sigma = n_\sigma h_x$ with $n_\sigma = (4, 6, 9)$. The absolute differences in phase and amplitude between the low and medium resolutions are at the level of $\Delta\phi \sim 10^{-7}$ and $\Delta A \sim 10^{-8}$ (top panel). In both cases we obtain the expected 4th order convergence up to merger (bottom panels). Convergence is slightly worse in the $\hat{a} = 0.9$ case. The plots are noisy in the ringdown phase (not shown), where the fields exponentially decrease by several orders of magnitude. Hence, we expect larger relative errors during the ringdown. For the science runs of the next section a much higher resolution is employed.

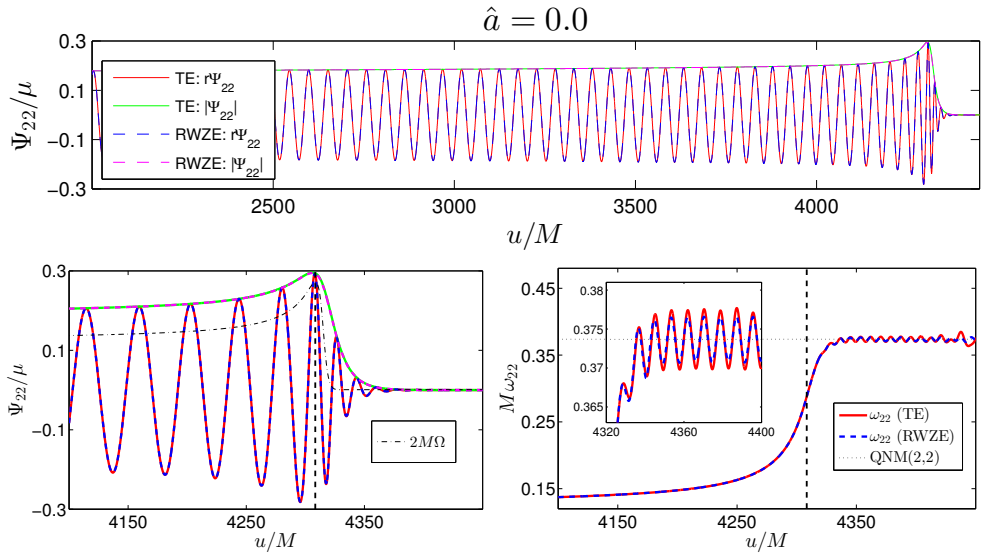

Figure 8. Case $\hat{a} = 0$: comparison between the 2+1 TE and the 1+1 RWZE code for insplunge-ringdown waveforms with $\mu/M = 10^{-3}$. Shown is the real part of the Ψ_{22} metric waveform extracted at $\mathscr{I}$ together with its amplitude and frequency. The vertical lines mark the time of the light-ring crossing ($t_{\rm LR} = u_{\rm LR} = u_{\Omega^{\rm max}} = 4308.39M$). Twice the orbital frequency $2\hat{\Omega}$ is represented with a dash-dotted black line (bottom left panel). The dotted horizontal line in the right panel marks the fundamental QNM frequency [90].

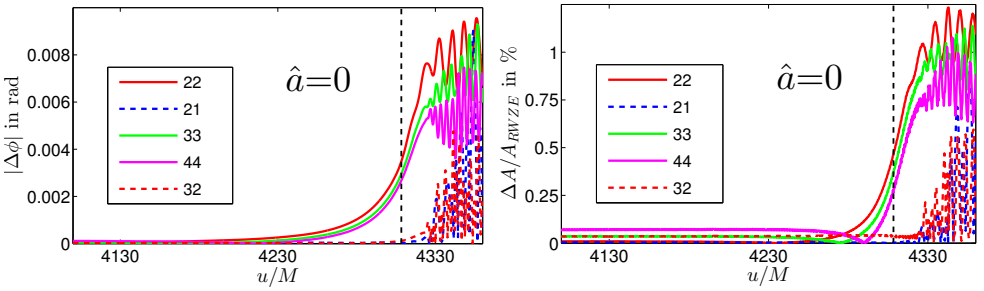

Figure 9. Case $\hat{a} = 0$: comparison between the 2+1 TE and the 1+1 RWZE code for insplunge-ringdown waveforms with $\mu/M = 10^{-3}$. Absolute phase (left) and relative amplitude (right) differences for the dominant multipoles at $\mathscr{I}$. The absolute phase differences are $\lesssim 2 \times 10^{-3}$ rad until the time of the light-ring crossing ($u_{\rm LR} = u_{\Omega^{\rm max}} = 4308.39M$) and remain $\lesssim 0.01$ during the ringdown ($u > 4308.39M$). The relative amplitude differences are at the order of $\sim 0.25\%$ until $u_{\rm LR}$ and remain $\lesssim 1.25\%$ afterwards.

6.4. TE vs. RWZE insplunge waveforms for $\hat{a} = 0$

We compare the 2+1 TE data for $\hat{a} = 0$ with the 1+1 RWZE data of [54, 56]. We use exactly the same dynamics for both sets of simulations, with $\hat{r}_0 = 7$. Figure 8 shows the RWZE Ψ_{22} complete (inspiral-plunge-ringdown) waveform extracted at $\mathscr{I}$ as computed from two simulations with $\mu/M = 10^{-3}$. The data are in excellent visual agreement also during the ringdown. A quantitative comparison for various

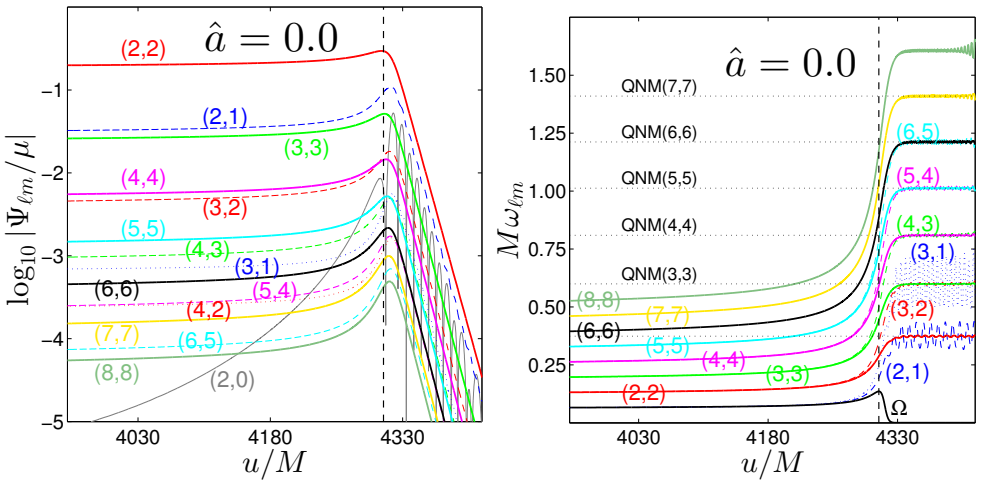

Figure 10. Multipolar structure around merger of the $\Psi_{\ell m}$ insplunge waveforms with $\mu/M = 10^{-3}$ and $\hat{a} = 0$ obtained from the 2+1 TE. Amplitudes (left) and GW frequencies (right). The vertical lines correspond to the maximum of the orbital frequency $u_{\Omega^{\rm max}} = u_{\rm LR} = 4308.39M$ (the orbital frequency $\hat{\Omega}$ is also shown as a black line in the right panel). The horizontal lines in the right panel mark the QNM frequencies of the black hole [90].

Table 3. Properties of multipolar waveforms at merger for $\hat{a} = 0$. The retarded time at the crossing of the light ring $u_{\rm LR}$, coincides with the time of the maximum of the orbital frequency, $u_{\rm LR} = u_{\Omega^{\rm max}} = 4308.39M$ (for $\hat{a} \neq 0$ they can differ, see Table A1 below) , with $M\Omega^{\rm max} = 0.136$. The peak of each multipolar amplitude divided by μ, $\hat{A}^{\rm max}_{\ell m} \equiv A^{\rm max}_{\ell m}/\mu$ occurs at another time $u_{A^{\rm max}_{\ell m}} \neq u_{\Omega^{\rm max}}$. The table lists the differences $\Delta t_{\ell m} = u_{A^{\rm max}_{\ell m}} - u_{\Omega^{\rm max}}$. For completeness we also state the peak values $\hat{A}^{\rm max}_{\ell m}$ and the frequencies at that time $M\omega_{\ell m}^{A^{\rm max}_{\ell m}}$. Values in brackets refer to 1+1 RWZ simulations of [54]. The table confirms that the peak of the $\ell = m = 2$ multipole occurs before the light ring crossing as pointed out in [54].

ℓ	m	$\Delta t_{\ell m}$	$\hat{A}^{\rm max}_{\ell m}$	$M\omega_{\ell m}^{A^{\rm max}_{\ell m}}$
2	2	-2.38 (-2.56)	0.29589 (0.29472)	0.27335 (0.27213)
2	1	9.41 (9.37)	0.10694 (0.10692)	0.29067 (0.29064)
3	3	1.11 (1.00)	0.051673 (0.051456)	0.45462 (0.45321)
3	2	6.85 (6.84)	0.018170 (0.018174)	0.45181 (0.45174)
3	1	10.55 (10.54)	0.0056954 (0.0056872)	0.41176 (0.41129)
4	4	2.90 (2.82)	0.014581 (0.014523)	0.63541 (0.63400)
4	3	7.22 (7.21)	0.0049634 (0.0049653)	0.63686 (0.63668)
4	2	9.54 (9.51)	0.0016570 (0.0016543)	0.62603 (0.62533)
5	5	4.18 (4.12)	0.0052278 (0.0052093)	0.81811 (0.81672)
5	4	7.63 (7.63)	0.0017267 (0.0017277)	0.82170 (0.82148)
6	6	5.20 (5.14)	0.0021703 (0.0021636)	1.00027 (1.00013)
6	5	8.09 (8.09)	0.00069673 (0.00069726)	1.00079 (1.00077)

multipoles is shown in Fig. 9, that reports phase and amplitude differences. Note

that no time/phase alignment is required to perform such a comparison since both sets of waveforms are extracted at scri and generated from the same dynamics. Phase differences are $\Delta\phi_{\ell m} = |\phi^{\rm TE}_{\ell m} - \phi^{\rm RWZ}_{\ell m}| \lesssim 10^{-3}$ rad until the time of the light ring crossing ($u_{\rm LR} = u_{\Omega^{\rm max}} = 4308.39M$) and remain below 0.01 during the ringdown ($u > u_{\rm LR}$). The relative amplitude differences $\Delta A_{\ell m}/A_{\ell m}$ are at the order of 0.25% until $u_{\rm LR}$ and remain $\lesssim 1.25\%$ during the ringdown.

These differences are larger than those estimated from self-convergence tests, so they have systematic origin. The two codes are independent, use different coordinate systems, solve different equations, and differ in many implementation details. Furthermore, the *Teukode* is a 2+1 code and the 2+1 RWZ waveforms are reconstructed from ψ_4 by integration. Remarkably, these systematic differences in phase and amplitude are small enough to be negligible for many practical purposes. We note in particular that the $\Delta\phi_{\ell m}$ are significantly smaller than the differences between RWZE and EOB waveforms found in [54, 55]. Thus, the result of that analysis is robust and confirmed here with an independent waveform data set. In particular, in [54] it was reported for the first time that the peak of the $\ell = m = 2$ multipole is located earlier in time than the peak of the orbital frequency (for $\hat{a} = 0$ the peak coincides with $u_{\rm LR}$). This observation is confirmed also by the 2+1 TE data, see Table 3.

The multipolar amplitudes and frequencies near merger are shown in Fig. 10. The data agree with the analysis of the multipolar structure performed in [52] within the 1+1 RWZE approach. Note in particular the oscillations in the quasi normal mode (QNM) frequencies, e.g. in modes $(2,1)$,$(2,2)$ and $(3,1)$. As explained in [50, 52], these oscillations arise from the interference between positive ($m > 0$) and negative ($m < 0$) QNM frequencies.

We analyzed also the effect of finite-radius extraction on the waveforms; these results are reported in Sec. 7.4 together with $\hat{a} \neq 0$ data.

7. From quasi-circular inspiral to merger and ringdown: dynamics and waveforms for $\hat{a} \neq 0$

In this Section we discuss the structure of the waveforms emitted from inspiraling and coalescing configurations with spin $\hat{a} \neq 0$. The underlying dynamics of the particle is computed according to Sec. 5. The purpose of this Section is threefold: (i) check the consistency of the (two) analytical expressions of the mechanical angular momentum loss with the angular momentum flux computed from the waves, Sec. 7.1; (ii) characterize quantitatively the multipolar waveform around merger, as obtained with $\hat{\mathcal{F}}^{v_\phi}_{\phi}$, for $|\hat{a}| \leq 0.9$ in Sec. 7.2 and for nearly-extremal configurations $0.9 < |\hat{a}| \leq 0.9999$ in Sec. 7.3, (iii) quantify waveform extrapolation errors in Sec. 7.4.

All simulations in this work refer to $\mu/M = 10^{-3}$ and post-circular initial data. Simulations are done for different values of black hole spins $\hat{a} \in [-0.9999, +0.9999]$ and $m = 0, 1, ..., 8$. For $|\hat{a}| < 0.99$ the initial separations $\hat{r}_0$ were chosen such that the systems perform about ~ 25 orbits before merger. This implies that the inspiral is "very-strong-field" when the spins are highly positive (with $\hat{r}_0 \sim 3$), and relatively "less-strong-field" when spins are highly negative (with $\hat{r}_0 \sim 10$). For nearly-extremal simulations the pronounced potential-well close to the black hole "traps" the junk radiation for long times (e.g. $1000M$ for $\hat{a} = 0.9999$), so that we needed bigger separations/longer inspirals for $|\hat{a}| \geq 0.99$. We collect in Table A1 in Appendix A detailed information about the configurations we consider.

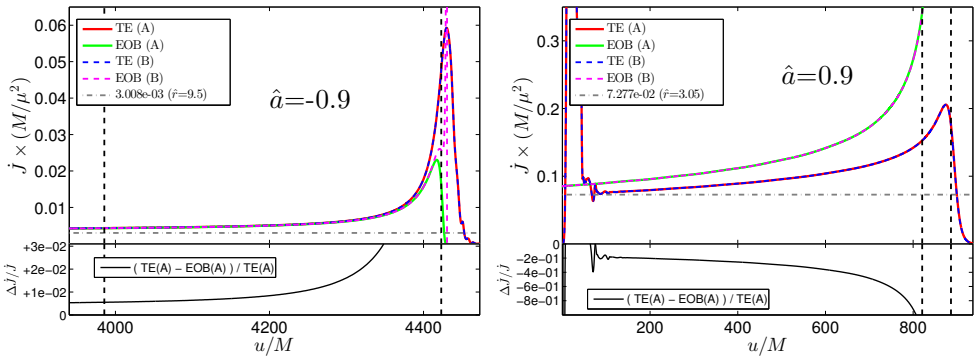

Figure 11. Consistency of numerical and EOB fluxes for $\hat{a} = -0.9$ (left) and $\hat{a} = +0.9$ (right). The top panels show the two analytical EOB-flux-prescriptions (A) and (B) (discussed in Sec. 5) and the respective numerical fluxes produced using those prescriptions. The vertical lines correspond to the LSO and the LR crossing. The horizontal line is the circular flux corresponding to the initial separation. The bottom panels show the difference between the numerical and the analytical flux when using prescription (A) (the respective line for (B) would lie on top in the plotted sector). Looking at the bottom left panel the analytical prescriptions for $\hat{a} = -0.9$ match the numerical fluxes within $\lesssim 1\%$ until well beyond the LSO. Instead, for $\hat{a} = 0.9$ the analytical information is less accurate ($\sim 100\%$ off at LSO). The numerical fluxes (A) and (B) are visually the same in both plots though the flux prescriptions differ significantly at LR.

The simulations discussed below use 6th-order finite differences. The resolution employed is $N_x \times N_\theta = (3600 \times 160)$. This setup is chosen to guarantee that the truncation errors, as estimated from self-convergence tests (Sec. 6.3), are around 10^{-10}, at least for the dominant multipoles and for most of the configurations.

7.1. Checking the consistency of the analytical radiation reaction

We compare the two prescriptions for the radiation reaction $\hat{\mathcal{F}}_\phi$ described in Sec. 5 with the numerical fluxes of the TE, to complement the analysis in Sec. 5. For both $\hat{\mathcal{F}}_\phi^{v_\phi}$ and $\hat{\mathcal{F}}_\phi^{v_\Omega}$ we compute an insplunge trajectory and perform a TE simulation. The consistency between the analytical flux and the numerical one (or "exact", for the given dynamics) is a crucial test of the consistency of the analytical model. Note that the comparison presented here is meaningful because in our setup the TE waveforms are extracted at scri. In the nonspinning case the accuracy of the 5PN-accurate, resummed, analytical radiation reaction has been tested in [50, 52, 56]. In these references it was shown that (for $\mu/M \leq 10^{-3}$) the 5PN-accurate information yields a radiation reaction consistent up to a few percent even below the LSO crossing.

When $\hat{a} \neq 0$ things are more complicated since the accuracy of the analytical flux, that is based on a limited PN knowledge of the $\rho_{\ell m}$'s, actually depends on the value of $\hat{a}$. As already pointed out in previous works, the analytical information we are employing here is not sufficient to guarantee agreement between analytical and numerical fluxes for $\hat{a} > 0.7$. Practically speaking, the analytical radiation reaction is too large and yields a smaller number of GW cycles up to merger than what it should be. This discrepancy was pointed out in Ref. [88], and thoroughly analyzed in follow-up works [91, 60, 14, 61]. In particular, Ref. [14] proposed to fit several high-order coefficients entering the $\rho_{\ell m}$'s to the circular flux in order to improve the behavior

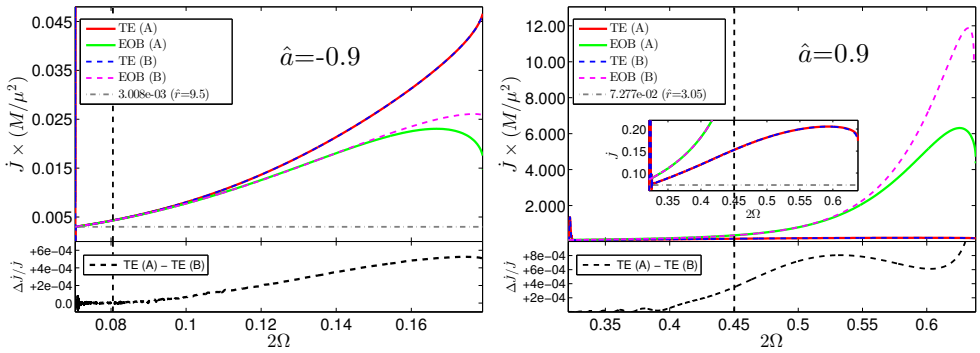

Figure 12. Consistency of numerical and EOB analytical fluxes for $\hat{a} = -0.9$ (left) and $\hat{a} = +0.9$ (right). The two major panels show the EOB-fluxes as obtained with prescriptions (A), i.e. $-\hat{\mathcal{F}}_\phi^{v_\phi}$, and (B), i.e. $-\hat{\mathcal{F}}_\phi^{v_\Omega}$, (discussed in Sec. 5) and the two respective numerical fluxes as functions of twice the orbital frequency $\hat{\Omega} = M\Omega$. The bottom panels show the difference between the numerical flux as produced with prescription (A) and (B). The vertical lines correspond to the LSO, the plots terminate at the LR. The horizontal line is the circular flux corresponding to the initial separation, which was chosen to provide ~ 25 orbits before merger. Note that all orbits are spent before the LSO, $\hat{\Omega} < \hat{\Omega}_{\rm LSO}$.

of the purely analytical $\rho_{\ell m}$'s as $\hat{a} \to 1$ (, see their Fig. 13). Here, for simplicity, we rely exclusively on analytical information, though we do not expect good consistency between the analytical and numerical fluxes for large, positive values of the black hole spin. In Sec. 9 we present a method to calculate the *consistent* GW flux at linear order in ν.

Let us compare the GW fluxes for spins $\hat{a} = \pm 0.9$ (with $\hat{r}_0$ as in Table A1). Figure 11 displays the comparison in the time domain. In the top panels we contrast $\dot{J}$ with $-\hat{\mathcal{F}}_\phi^{v_\phi}$ and $-\hat{\mathcal{F}}_\phi^{v_\Omega}$ while the bottom panels show the relative difference between $\dot{J}$ and $-\hat{\mathcal{F}}_\phi^{v_\phi}$. The two vertical lines on the plot indicate, from left to right, the LSO and the LR crossing. As expected, we find excellent agreement between the analytical and numerical flux when $\hat{a} = -0.9$ (also beyond the LSO and almost up to merger), whereas the agreement is poor for $\hat{a} = +0.9$ (the analytical flux is off by a factor 2 at LSO crossing) due to the lack of higher PN information in the $\rho_{\ell m}$'s. This discrepancy holds for both implementations of the non-Keplerian behavior we discussed in Sec. 5. Instead, the differences in the prescriptions manifest in the behavior of the fluxes for $\hat{a} = -0.9$ around the light ring crossing time. $-\hat{\mathcal{F}}_\phi^{v_\Omega}$ (dashed magenta line in Fig. 11) deviates from $-\hat{\mathcal{F}}_\phi^{v_\phi}$, consistently with the $\hat{a} = -0.99$ case analyzed above, with $-\hat{\mathcal{F}}_\phi^{v_\phi}$ remaining close, in shape, to $\dot{J}$.

Figure 12 illustrates the same comparison under a different and complementary point of view: the fluxes are plotted versus the orbital frequency $\hat{\Omega} = M\Omega$. Note that the ~ 25 orbits (most of the simulated time) have frequencies $\hat{\Omega} < \hat{\Omega}_{\rm LSO}$ (the LSO is marked by vertical lines). For $\hat{a} < 0$ (left panel), the analytical prescription (A) ((B)) deviates from the numerical flux by about 1% (1%) at the LSO and 60% (30%) at the LR. This is roughly comparable to the nonspinning case [56]. The unphysical feature of prescription (B) for $\hat{a} < 0$ as discussed in Sec. 5 happens at $r < r_{\rm LR}$ and does therefore not affect the waveforms: both analytical fluxes generate almost identical numerical fluxes. Differences between the two TE fluxes are below 0.2% during the

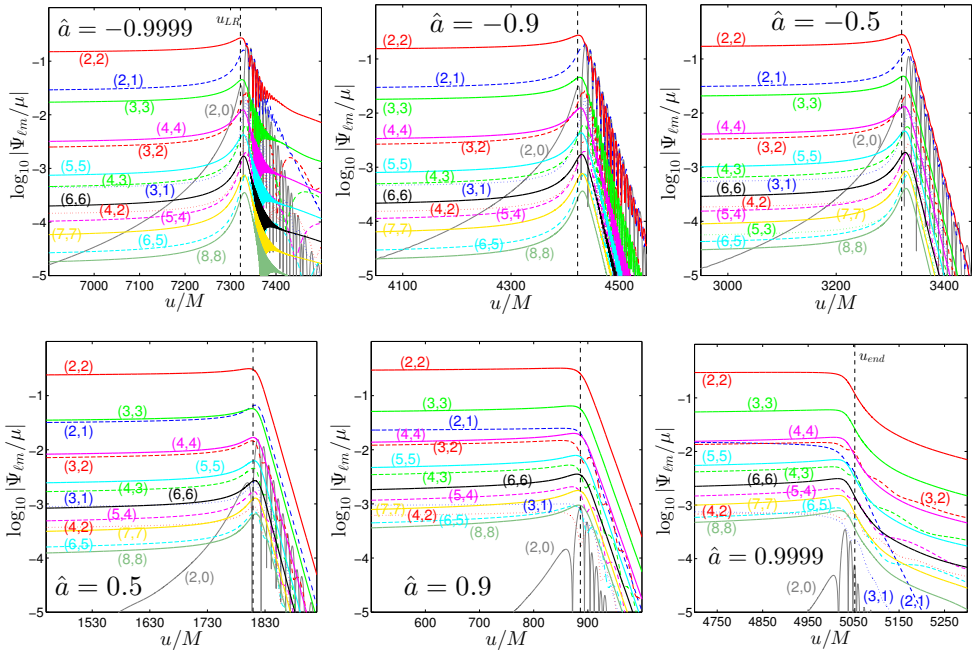

Figure 13. Multipolar amplitudes of the $\mu/M = 10^{-3}$ insplunge waveforms for certain values of $\hat{a}$. The vertical line on each plot marks the crossing of the LR, except for $+0.9999$, where it marks the end of the trajectory (in this case our trajectory stops slightly before the horizon, see discussion in text).

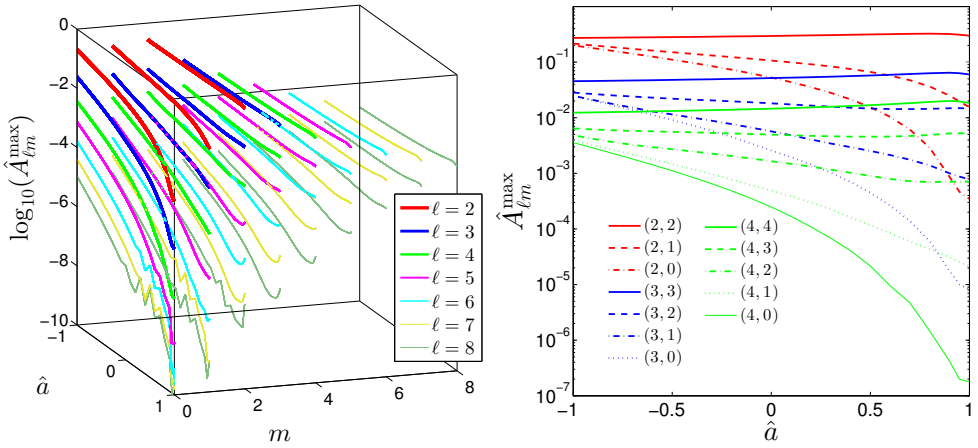

Figure 14. Peaks of multipolar amplitudes as functions of the black hole spin. Left: 3D visualization of all the data up to $\ell = m = 8$. ℓ-modes are identified with colors, the spin dependence is marked by solid lines. Right: peaks of the dominant multipoles as functions of the spin. This plot is a 2D cut of the left panel (restricting to $\ell \leq 4$). The $m = 0$ and $m = 1$ modes have the largest amplitudes for $\hat{a} \to -1$. The diagonal modes have weak dependence on the spin and attain a maximum for spin values $0.8 < \hat{a} < 1$.

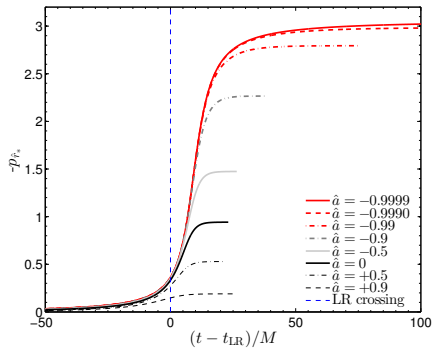

Figure 15. Time evolution of the radial momentum p_{r_*} around the light-ring crossing time as a function of $\hat{a}$. The magnification of p_{r_*} around merger when $\hat{a} \to -1$ is responsible for the sharpness of the peak of the multipolar waveform and the increased importance of subdominant multipoles.

whole insplunge. The reason for this is simply that during the plunge (i.e. after the LSO crossing) the radiation reaction does not play a significant role (quasi-geodetic plunge.) Although the standard flux can still be improved, we consider the results for $\hat{a} \leq 0$ quite satisfactory for studying large-mass-ratio mergers. For $\hat{a} = 0.9$ the figure illustrates clearly how both analytical prescriptions systematically deviate from the numerical fluxes already at the beginning of our (rather strong-field) inspiral. Comparing the deviations for different $\hat{a} > 0$ runs we find that higher spins suffer from larger deviations. For instance, for $\hat{a} = 0.5$ both flux prescriptions deviate from the numerical outcomes by about 2% at the LSO and more than 60% at the LR; for $\hat{a} = 0.9$ the deviation is about a factor two already at the LSO. The reason is that, differently from $\hat{a} < 0$, the plunge dynamics for $\hat{a} \to 1$ is very circular and the radiation reaction plays a significant role also in the strong-field regime. We conclude that an urgent step for future analytical developments (beyond using the effective fits of [14]) is the inclusion of higher PN corrections to the $\rho_{\ell m}$'s residual amplitude corrections [36].

7.2. Waveforms for $\hat{a} \neq 0$: multipolar hierarchy at merger

Gravitational waveforms for a particle inspiraling and plunging into a Kerr black hole have been computed for the first time by Sundararajan et al. [12] and then used and updated in [60, 61]. In particular, the recent Taracchini et al. [61] paper (that appeared while the current study was being finalized) presented a detailed analysis of the ringdown waveforms for modes $(\ell, m) = (2,2), (2,1), (3,3), (3,2), (4,4), (5,5)$. Our study confirms previous findings but also extends/complements them because it (i) computes TE data for all multipoles up to $\ell = m = 8$; (ii) considers higher values of the black hole spin; (iii) presents a detailed analysis of $m = 0$ modes; and (iv) explicitly connects the structure of the multipolar waveform around merger with the noncircularity of the plunge.

Let us focus first on the dependence of the waveform amplitude on $\hat{a}$. Figure 13 gathers the time evolutions of the (μ-normalized) amplitudes $\hat{A}_{\ell m} \equiv A_{\ell m}/\mu \equiv |\Psi_{\ell m}|/\mu$ of several multipoles for representative values of $\hat{a}$. This figure is complemented by

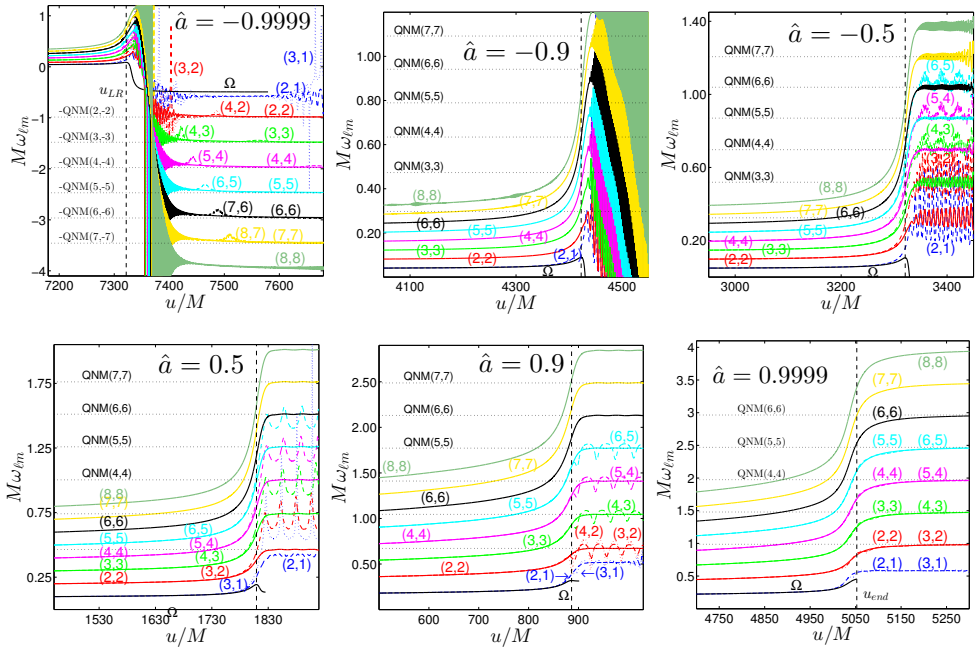

Figure 16. Multipolar GW frequencies of the $\mu/M = 10^{-3}$ insplunge waveforms for certain values of $\hat{a}$. The vertical line on each plot marks the crossing of the LR, except for $+0.9999$, where it marks the end of the trajectory (in this case our trajectory stops slightly before the horizon, see discussion in text). Horizontal lines are the QNM frequencies of [90]. For nearly-extremal negative $\hat{a}$ we find negative frequencies due to the inversion of the trajectory after the LSO-crossing.

Fig. 14 that highlights certain properties of the maxima of the $\hat{A}_{\ell m}$'s. The left panel of Fig. 14 shows the global variation of $\hat{A}^{\rm max}_{\ell m} = \max(\hat{A}_{\ell m})$ versus $(\ell, m, \hat{a})$. The right panel of the figure is the vertical plane projection of the left panel restricted to modes up to $\ell = m = 4$ for clarity. We make the following observations:

(i) In the quasi-circular inspiral, the hierarchy between the $(2,1)$ and $(3,3)$ modes is inverted as the spin goes from negative to positive values. The $(2,1)$ mode becomes progressively less important for $\hat{a} \to 1$. More precisely, looking at the maxima (right panel of Fig. 14), $\hat{A}^{\rm max}_{21} \geq \hat{A}^{\rm max}_{33}$ for $\hat{a} \lesssim 0.6$ and $\hat{A}^{\rm max}_{21} < \hat{A}^{\rm max}_{33}$ for $\hat{a} \gtrsim 0.6$. Similarly, $\hat{A}_{31}$ is close to $\hat{A}_{43}$ during the inspiral for $\hat{a} = -0.9$, but at least an order of magnitude smaller for $\hat{a} = +0.9$. This feature is even more prominent for higher spins with $\hat{A}_{43}$ and $\hat{A}_{31}$ practically coinciding for $\hat{a} = -0.9999$ (see top left plot of Fig. 13) . This behavior is due to the analytical structure of the leading-order spin-dependent term entering the $(2,1)$ and $(3,1)$ waveform (see for example Eqs. (38b) and (38e) of Ref. [88]). When taking the limit of a nonspinning point particle, the leading, spin-dependent PN correction to the Newtonian prefactor is of the form $(1 - \frac{3}{2}\hat{a}v)$ for $(2,1)$ and $(1 - 2\hat{a}v^3)$ for $(3,1)$, which explains why their amplitudes are larger for $\hat{a} < 0$ and smaller for $\hat{a} > 0$.

(ii) Concerning multipoles with $m \neq 0$, the peak of each waveform flattens for $\hat{a} \to 1$, and becomes sharper and more pronounced for $\hat{a} \to -1$. This effect was pointed out in Ref. [61] for the therein computed modes, though explicitly

shown only for $\ell = m = 2$. Here we complete the results of Ref. [61], showing how this flattening is a general property of all multipoles, as it is related to the (absence of) next-to-quasi-circular (NQC) corrections in the waveform. Defining the function $B = \Sigma/\Delta$, it is meaningful to use for this analysis the radial momentum $p_{r_*} = \sqrt{A/B}p_r$, i.e., the momentum conjugate to the definition of a tortoise-like coordinate as $dr_*/dr \equiv \sqrt{B/A}$ (note that this r_* is different from the usual tortoise coordinate introduced above). The important point is that using p_{r*} the p_r-dependent part of the Hamiltonian just reads $p^2_{r_*}$, analogously to the Schwarzschild case [92]. Consistently with the discussion in [61], the magnitude of p_{r_*} becomes very small when $\hat{a} \to 1$ and so does the corresponding effect on the waveform amplitude. On the contrary, due to the larger values of p_{r_*} around light-ring crossing when $\hat{a} \to -1$, noncircular effects are nonnegligible and yield the sharpness of the peak of each multipolar amplitude. The dependence of p_{r_*} on $\hat{a}$ is illustrated in Fig. 15. In Appendix C we present an heuristic discussion based on the leading-order Newtonian waveform to practically illustrate how NQC corrections can shape the multipolar waveform peaks close to merger.

(iii) The effect of p_{r_*} is particularly evident when inspecting the $m = 0$ modes, that only depend on the radial part of the motion. For example, for $\hat{a} = 0.9$, $\hat{A}^{\rm max}_{20} = 0.0009$, which is negligible compared to the dominant mode's peak $\hat{A}^{\rm max}_{22} = 0.3212$. By contrast, for $\hat{a} = -0.9$ we have $\hat{A}^{\rm max}_{20} = 0.1788$, which is of the same order as $A^{\rm max}_{22} = 0.2738$ and $\hat{A}^{\rm max}_{33} = 0.1996$. This kind of information is gathered in Fig. 14, where the right panel of the figure clearly shows the "growth" of the subdominant modes over $1 > \hat{a} > -1$. This is particularly striking for the $m = 0$ modes, that for $\hat{a} \to -1$ attain values that are comparable to the $\ell = m$ ones. The consequence of this growth is that the well-known hierarchy of the modes at merger in the $\hat{a} = 0$ case (discussed in Appendix A of [54]) does no longer hold when $\hat{a} \neq 0$.

Let us now turn to discuss the dependence of the multipolar frequencies $M\omega_{\ell m}$ on $\hat{a}$, which is illustrated in Fig. 16 for the same $\hat{a}$'s as above. The plots show that the QNM interference phenomenon mentioned in Sec. 6.4 is greatly enhanced for $\hat{a} < 0$, i.e. retrograde plunges. This is due to the progressively larger excitation of $m < 0$ QNMs as $\hat{a} \to -1$. These results are consistent with the findings of Refs. [60, 61].

Finally, though we do not discuss EOB waveform calibration in this work, we list in Table A2 useful information extracted from the TE waveforms for a few values of $\hat{a}$. The numerical relativity completion of the EOB waveform requires the calibration of next-to-quasicircular (NQC) corrections to data and the attachment of the QNM waveform to the merger one. Both NQC corrections and QNM attachment are usually performed after a careful analysis of the properties of the merger waveform, i.e. around the peak of $\hat{A}_{22}$. The table lists the values of $\hat{A}^{\rm max}_{\ell m}$ and the corresponding GW frequencies together with the time lag between the peak of the orbital frequency, $t_{\Omega^{\rm max}}$, and the peak of each multipole $u_{A^{\rm max}_{\ell m}}$,

$$\Delta t_{\ell m} = u_{A^{\rm max}_{\ell m}} - t_{\Omega^{\rm max}}. \tag{60}$$

We confirm the finding of [60] that Δt_{22} (as well as all $\Delta t_{\ell m}$'s) strongly depends on the spin for $\hat{a} \to 1$. The comparison of our $\Delta t_{\ell m}$ with the values stated in Tab.III of [60] shows excellent agreement between the two codes, considered that they are completely independent and use different prescriptions for the radiation reaction. Besides $\Delta t_{\ell m}$ also the peak of $\hat{\Omega}$ is spin-dependent, because the peak of $\hat{\Omega}$ becomes progressively

less visible for $\hat{a} \to 1$, so that $\hat{\Omega}$ becomes effectively monotonic in time for large spins. As advocated already in [60], $\hat{\Omega}$ is not a good quantity to identify an "anchor" point (like its maximum, that is a robust choice in the nonspinning case) for EOB modeling purposes, namely for determining effective next-to-quasi-circular corrections and for matching the insplunge EOB waveform to the ringdown part. Reference [60] suggested to use the time when the EOB insplunge waveform peaks. An alternative approach currently under investigation [92] is to use the peak of the (formally) "pure orbital" frequency. This frequency is defined as

$$\hat{\Omega}_{\rm orb} = \partial_\phi \hat{H}_{\rm orb}, \tag{61}$$

that, differently from $\hat{\Omega}$, always has a neat isolated peak (before the particle gets to the horizon) due to the absence of spin-orbit effects. Table A3 in Appendix A displays the time lag $u_{A_{22}^{\rm max}} - \hat{t}_{\Omega_{\rm orb}^{\rm max}}$ between the peak of $\hat{A}_{22}$ and the peak of $\hat{\Omega}_{\rm orb}$. Interestingly, this time difference is small (order unity) and has a very mild dependence on spin up to $\hat{a} = 0.8$.

7.3. Waveforms for nearly-extremal configurations

The analytical and numerical setup of this work allows us to explore nearly extremal configurations. In this section we discuss the multipolar hierarchies at merger for the cases $|\hat{a}| = 0.99, 0.999, 0.9999$. To our knowledge these are the first results available of this kind (compare with [61].)

Since these simulations are technically more challenging than those for $|\hat{a}| \leq 0.99$, we needed artificial dissipation operators for stability (using the same resolution as before). Also, for such high positive spins, the analytical radiation reaction (in both prescriptions) is not only inaccurate but grows very rapidly around the LR and corrupts the numerical calculation of the trajectory. In order to perform the simulations one can either (i) stop the particle dynamics before it reaches the horizon or the LR and advect the source off the domain (for example, for $\hat{a} = 0.999$ (0.9999) the LR is at $\hat{r}_{\rm LR} \sim 1.05$ (1.02) while our trajectory stops at $\hat{r}_{end} = 1.09$); or (ii) smoothly "switch off" the fluxes after the LSO (similarly to the procedure of Sec. 9). Both procedures lead to qualitatively the same results, and, although they might introduce a systematic effect in the QNM waveforms, we do not observe any obvious unphysical features in using (i).

Figures 13 and 16 include the amplitudes $\hat{A}_{\ell m}$ and frequencies $M\omega_{\ell m}$ for $|\hat{a}| = 0.9999$ for various dominant multipoles. (The analogous plots for $|\hat{a}| = 0.99, 0.999$ resemble that one and do not convey more information).

The features discussed for lower spins remain valid, i.e. flattening of the waveform amplitude around merger for $\hat{a} \to 1$ and sharpening as $\hat{a} \to -1$. The most interesting features are observed during the QNM ringing. First, the top left plot of Fig. 16 shows that for nearly-extremal negative $\hat{a}$ the QNM waveforms are characterized by negative frequencies. The frequencies correspond to the QNM ringing modes with $m < 0$ [90]. As noted in [61], they are excited by the change of sense of rotation of the particle ($\hat{\Omega}$ has a zero) during the plunge. Our analysis extends the previous one by considering higher spins and including all the multipoles up to $\ell = m = 8$. Second, for nearly-extremal cases the trapping of modes leads to very weakly damped QNM's in the case of free QNM ringing (no particle), see e.g. [93, 94, 90, 15, 70]. We qualitatively confirm this behavior also for the QNM's excited by a particle. Damping times can be measured by fitting the exponential decays of Fig. 13. Note that in

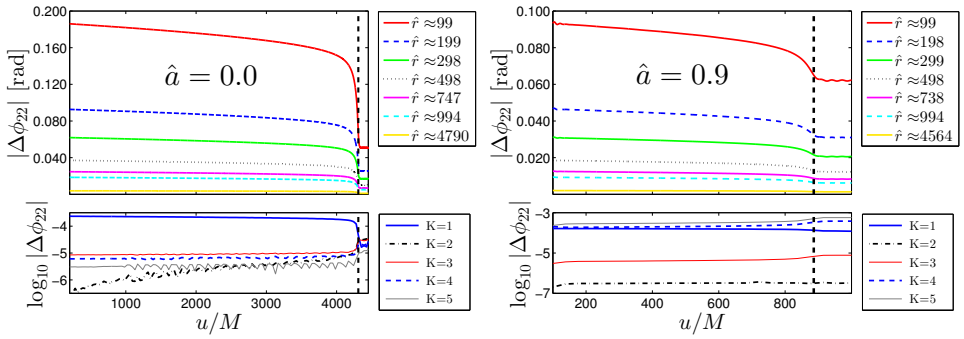

Figure 17. Finite-extraction-radius effects on the phase of the $\psi_{4\,22}$ waveforms from insplunges with $\hat{a} = 0$ (left) and $\hat{a} = 0.9$ (right). Top panels refer to absolute differences between scri and finite radii $\hat{r} \sim (100, 200, 500, 750, 1000, 4800)$. Bottom panels refer to absolute differences between the scri values and waveforms extrapolated according to Eq. (62) with various polynomial orders $K = 1, 2, 3, 4, 5$. Note the $\log_{10}$ in the y-axis. The vertical line represents the crossing of the LR.

the nearly-extremal regime the exponential decay holds for much longer times than shown in the plots, e.g. for $\hat{a} = 0.9999$ we fit over $[5400M, 6800M]$ for the $\ell = m = 2$ mode. The damping exponents of the dominant overtones computed for $\hat{a} = 0.9999$ are $\{-3.52, -39.11, -3.61\} \times 10^{-3}$ for $(\ell, m) = \{(2, 2), (2, 1), (3, 3)\}$ respectively. These numbers match at $\lesssim 4\%$ with the corresponding values for free QNM ringing [90].

7.4. Finite extraction errors in waveforms

To estimate finite extraction effects on the waveforms, we have investigated systematic phase and amplitude differences between scri and finite radius waveforms, as well as the performance of an extrapolation procedure commonly used in numerical relativity.

Figure 17 (top panels) shows phasing results for two case studies: the $\hat{a} = 0$ insplunge $\ell = m = 2$ waveform of Sec. 6.4 (left) and the $\hat{a} = 0.9$ insplunge $\ell = m = 2$ waveform (right). In the $\hat{a} = 0$ plot finite extraction errors at $\hat{r} \sim 100$ amount to $\Delta\phi_{22} \sim 0.16$ *rad*. This value is not only significantly larger than truncation errors, but has a similar size as the dephasing due to horizon absorption fluxes [53]. Hence, extraction at these radii can affect relevant physics. Similarly, phase errors in the $\hat{a} = 0.9$ plot are at the order of $\Delta\phi_{22} \sim 0.08$ *rad* (~ 0.03 *rad*) for $\hat{r} \sim 100$ (~ 300) accumulated to merger over 25 orbits. In all the cases analyzed, finite extraction errors are typically larger at early times. The quantitative difference between finite radius and scri waveforms is relevant when comparing and calibrating the EOB waveform. For example, Ref. [54] has found that no time-phase alignment is needed if waveforms are extracted at scri and that the errors at early times are those expected by the order of the PN approximation of the EOB waveform.

Many numerical relativity (either nonlinear or perturbative) simulations calculate approximate GWs at finite radii and use extrapolation to estimate those at null infinity. Our method allows us to test such a procedure. In order to extrapolate waveforms extracted at finite radius, we assume that the phase (and the amplitude) of the finite-

radii waveform behaves as a Kth order polynomial in $1/r$,

$$f(u,r) = f^{(0)}(u) + \sum_{k=1}^{K} r^{-k} f^{(k)}(u)\,. \tag{62}$$

We fit this model for every radius and some choices of the polynomial order K. The first term approximates the scri waveform. The extrapolation procedure is applied here on the ψ_4 multipolar waveform considering radii $\hat{r} \sim (100, 200, 500, 750, 1000, 4800)$. We find that the leading finite radius effect on $\phi(u,r)$ is the $1/r$ behavior, i.e. the term $K = 1$. Using larger K can reduce the phase errors, but we do not find a single prescription for $K > 1$ that robustly improves the results among different datasets. Figure 17 (bottom panels) shows these findings for the case study. Using linear extrapolation we obtain typical phase differences of $\Delta\phi \lesssim 10^{-3}$ with respect to scri. Different results are obtained using extrapolations with $K > 1$. For example, for $\hat{a} = 0$ the choice $K = 5$ gives the smallest errors, but for $\hat{a} = 0.9$ that choice leads to a large error and $K = 2$ is the optimal one. As suggested by the figure, in many cases $K = 2$ improves the $K = 1$ result. Similar results are found for the amplitude (not shown in the figure). The only important difference in that case is that $A(u,r)$ clearly shows a parabolic behavior, thus $K = 1$ extrapolation cannot be used. We suggest $K = 2$ as a "safe option" also in nonlinear numerical relativity simulations [+].

8. Horizon absorption during insplunge

Here, we discuss the fluxes absorbed by the black hole. The theoretical tools to compute time-domain horizon fluxes are provided by [62] and described in Sec. 3.3. The accuracy of our implementation was tested against frequency domain circular orbit data as presented in Sec. 6.1. In this Section, we compute, for the first time, horizon fluxes for insplunge trajectories around spinning black holes.

Horizon absorption is nonnegligible during the inspiral and can be important at merger already in the nonspinning case [53]. In the spinning case, it is expected to be more relevant because absorption terms enter at 2.5 PN order (4PN for $\hat{a} = 0$), e.g. [14]. So far, no numerical calculation of horizon-absorbed fluxes during the insplunge into a rotating black hole has been performed. The formalism of Poisson [62] allows us to perform this calculation beyond the LSO, but not up to the LR, as we shall see.

In Fig. 18 we compare the $m = 2$ horizon absorbed energy flux with the infinity flux for $\hat{a} = -0.9$ (left) and $\hat{a} = 0.9$ (right). The horizontal lines indicate the fluxes for a circular orbit at the initial separation of the insplunge, $\hat{r}_0 = 9.5$ (3.05) for $\hat{a} = -0.9$ (0.9). The agreement of our fluxes at initial times with the circular fluxes at $\hat{r}_0$ is about $\sim 0.2\%$. During the 25 orbits up to the LSO the ratio of horizon-absorbed fluxes to null infinity fluxes amounts to $\sim 1\%$ ($\sim 3\%$) for $\hat{a} = -0.9$ (0.9). These values are consistent with circular orbits' fluxes of Sec. 6.1.

Between the LSO and the LR we observe a rapid variation of the horizon fluxes. Unfortunately, our calculation becomes inaccurate and we stop it some time before reaching the LR. The reason of the failure is twofold: (i) the wave-extraction at the horizon is corrupted by the source term when the particle reaches the horizon (cf. also Fig. 3 of [53]); (ii) the time-domain formalism by Poisson relies on the calculation

[+] Note in the nonlinear case there are more complications due to the dynamical gauge (no fixed background), dynamical null structure, nonlinear effects, and outer boundary errors.

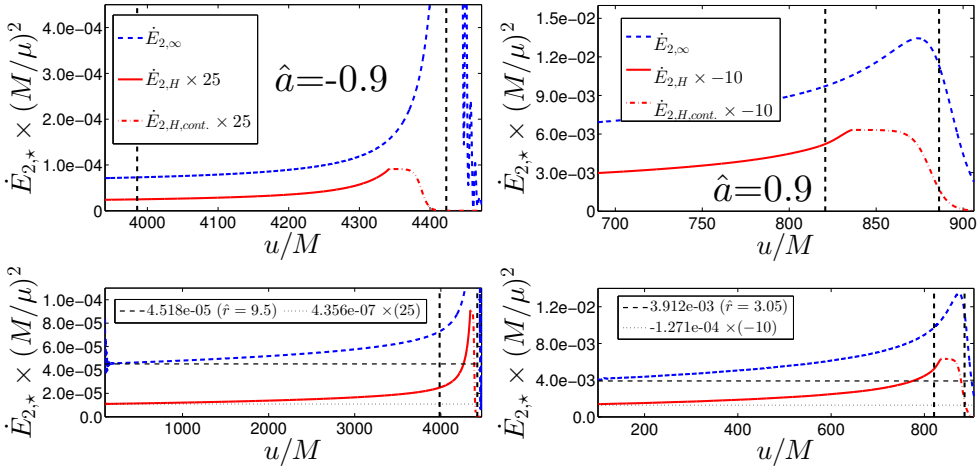

Figure 18. Comparison of $m = 2$ energy fluxes at the horizon (red) and scri (blue dashed) for $\hat{a} = -0.9$ (left panel) and $\hat{a} = 0.9$ (right panel). The time coordinate u refers to the retarded time for scri and the advanced time v for the horizon. The first vertical lines mark the LSO and the second the LR. The horizontal lines show the circular orbits' fluxes at the initial separation. Note the absorbed fluxes are scaled by multiplicative factors, and, in the right panel, the factor is negative, i.e. the two fluxes have different signs. The relative importance of horizon fluxes to infinity fluxes remains around $\sim 1\%$ ($\sim 3\%$) for $\hat{a} = -0.9$ (0.9). The waveforms at the horizon are contaminated by the source at late times. As a consequence, the horizon fluxes are reliable until $v \approx v_{r_+} - 100M$ as marked by the solid line (see discussion in the text). The dashed lines illustrate how we analytically continue the horizon fluxes in order to build a self-consistent radiation reaction for insplunge trajectories (see Sec. 9).

of advanced-time integrals (see Sec. 3.3); the resulting absorbed fluxes are corrupted *earlier* than the waveforms. In order to exclude the waveform corruption, the flux integrals are considered up to $v \sim v_{r_+} - 100M$, where v_{r_+} is the advanced time corresponding to the particle reaching the horizon (see also Sec. 4). For this reason a 2+1 formalism that is "local" in time, i.e. only relies on data on hypersurfaces, seems desirable. We leave the development of such a formalism, that will allow us to compute horizon fluxes up to the LR, to the future.

Finally, we stress that the absorbed fluxes calculated in this Section are inconsistent since they are not taken into account in the trajectory. Analytical results for absorbed fluxes are available, but the development of *resummed* expressions valid in the strong-field–fast-motion regime and for generic spins has just started [95, 14]. In the next Section we develop a numerical method to calculate consistent horizon absorptions and evaluate their impact on the waveform phasing for a test case.

9. A numerical method to compute $\mathcal{O}(\nu)$ self-consistent GW fluxes

In this Section we present a method to compute *exactly* the GW fluxes at linear order in the symmetric mass ratio $\mathcal{O}(\nu)$. The method is a simple and self-consistent procedure that iteratively employs the GW flux extracted from the TE simulations in the particle dynamics. We have tested it for various spin values; here we discuss

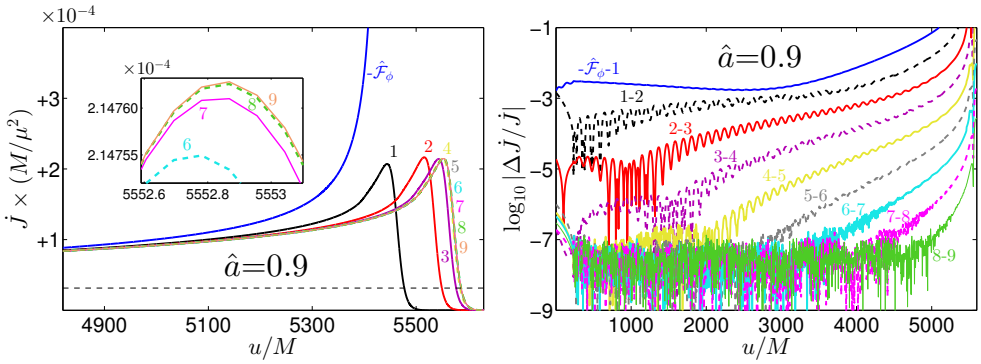

Figure 19. Convergence of the iterative self-consistent procedure for the radiation reaction. Insplunge with $\hat{a} = 0.9$ and $r_0 = 4M$. Left: GW fluxes after each iteration. Note the large differences with respect to the EOB radiation reaction (iteration 0) even though we used the refining fit given in [14]. Right: Convergence of the fluxes after each iteration. After 9 iterations we obtain a self-consistent radiation reaction including both infinity and horizon fluxes.

the case $\hat{a} = 0.9$. Differently from the $\hat{a} = 0.9$ simulation of previous Sections (see also Tab. A1), the simulations of this Section start at $\hat{r}_0 = 4$ corresponding to ~ 120 orbits. We discuss, in particular, the differences with the EOB analytical radiation reaction and the effect of horizon absorbed fluxes.

At the first iteration, the fluxes are calculated by the analytical EOB formula, as described in Sec. 5. Then, we run a series of 16 simulations for $m = 1, ..., 8$ and $s = \pm 2$ and compute the numerical fluxes at scri and at the horizon. These numerical fluxes are then employed for a new calculation of the trajectory. At the next iteration, new fluxes are computed by another series of 16 simulations. If a fixed point exists, and/or the first step is accurate enough, convergence should be observed. This was the case for all the tests we performed, i.e. provided with a "close-enough" guess, the method converges and the differences between the fluxes at each iteration consistently converge to zero. The most difficult cases are those with positive high spins, where EOB fluxes are less accurate. For example, for $\hat{a} = +0.9$ we were able to obtain iterative convergence only when the initial EOB trajectory was computed incorporating the refined $\rho_{\ell m}$'s obtained in Ref. [14] by fitting to numerical data into the radiation reaction.

The result of the iterative procedure is summarized in Fig. 19. The left panel shows the complete (scri and horizon) GW flux at each iteration. The peak position of the flux significantly changes for the first iterations and in total from $u_{peak} \approx 5450M$ for the analytical flux to $u_{peak} \approx 5553M$ for the 9th iteration. On the contrary the peak amplitude remains approximately the same after the first iteration (see inset), consistently with the intuition that it is not a radiation-reaction-driven effect, but rather its structure depends on the plunge phase. The right panel shows that the relative differences in the fluxes between the previous and the next iteration converge very rapidly to zero. At iteration 9 the relative flux differences saturate around 10^{-7} during most of the evolution, and the radiation emitted during the insplunge is consistent with the one used for the particle to this level.

The impact of the consistent flux on the GW modeling is quantified by considering the difference in the number of gravitational wave cycles $\Delta N_{\rm gw}$ between the final

waveform (after 9 iterations) and the starting one (iteration 0): $\Delta N_{\rm gw} \sim 6.7$. Note, however, that the 0th iteration waveform is already very different from the one computed in Sec. 7.2 without the use of the fit of [14]. The latter difference amounts to $\Delta N_{\rm gw} \sim 28.8$; so, overall, the self-consistent simulation differs from the corresponding EOB one by about $\Delta N_{\rm gw} \sim 35.5$.

We tested the importance of horizon absorbed fluxes by performing another self-consistent calculation that neglects these contributions. We found that the final trajectory is *shorter* and the particle reaches the horizon $\sim 117M$ earlier compared to the final trajectory that includes horizon absorption. The fact that in presence of horizon absorption effects the insplunge is longer is explained by superradiance, i.e. the net effect of horizon-absorption for $\hat{a} = 0.9$ is essentially an *emission* of energy/angular momentum. The differences in the dynamics correspond to $\Delta N_{\rm gw} \sim 5.4$, i.e. ~ 127.1 instead of ~ 124.4 orbits before merger. This result highlights the importance of horizon-absorbed fluxes during the insplunge.

We mention two technical details. First, at early times $\lesssim 200M$ our numerical fluxes are corrupted by "junk radiation" due to the imperfect initial data. For their use in the trajectory calculation, we linearly extrapolate the flux from $u \in [250, 500]$ to $u = 0$. This procedure does not introduce a significant error. As a check we compared the extrapolated value at $u = 0$ with the circular flux at the relative radius and find an agreement of $\sim 0.2\%$. Second, as mentioned in the previous Section, our horizon fluxes are only reliable until $v \approx v_{r_+} - 100M$. For the numerical fluxes we therefore switch off the absorbed fluxes at late times by multiplying them with a tanh() function that goes rapidly to zero. The result of this procedure is shown by the dotted red lines of Fig. 18.

10. Summary and Outlook

We summarize the main findings of this paper and give an outlook for future developments.

New approach for time-domain solution of the 2+1 Teukolsky equation. We presented a new approach to compute time domain solutions of the 2+1 Teukolsky equation (TE) based on horizon-penetrating, hyperboloidal foliations of Kerr spacetime and spatial compactification. The coordinates are constructed from the ingoing-Kerr coordinates, introducing a simple compactification and demanding the invariance of the coordinate expression for outgoing radial characteristics (see Sec. 2). The advantages of these coordinates in numerical applications are pointed out in Sec. 4.2.

In Sec. 3 we discussed the derivation of a regular inhomogeneous 2+1 TE as well as the computation of the point-particle source term in the new coordinates. A natural way to derive a regular TE in horizon-penetrating coordinates is to use the Hartle-Hawking tetrad [2] or the one in [73] (we use the latter) with a subsequent rescaling for regularity at scri.

New wave generation algorithm for point-particle perturbations. We implemented and tested a gravitational wave generation algorithm based on the new coordinates. The code *Teukode* [15], employed in this work, uses a standard method-of-lines approach and finite differencing operators of fourth and sixth order. The Dirac δ functions involved in the particle description are handled either using a narrow Gaussian or a

discrete representation [81, 82, 10]. We found that the "best" representation, i.e. the most accurate and computationally most efficient, depends on the trajectory (see Sec. 4.3).

We have tested the code for the simple, but physically relevant, cases of circular orbits (various rotating backgrounds) and a radial geodetic plunge (Schwarzschild). In the computation of circular GW fluxes at scri and at the horizon, our 2+1 simulations obtain a good accuracy as verified in a comparison against frequency domain results.

In Sec. 6.4 the results of the 2+1 *Teukode* for a nonrotating background are directly compared with those of the RWZE 1+1 code [54, 56, 53]. At the resolution employed, small systematic differences in the multipolar phases and amplitudes (i.e. larger then those expected from self-convergence tests) are found between the two codes. These differences are negligible for many practical purposes; notably they do not influence, but confirm, the precise analysis of [54, 55].

Modeling the late-inspiral-merger waveforms from large-mass-ratio spinning black hole binaries. We applied our method to study GWs from large-mass-ratio and spinning black-hole binaries. The plunging particle's trajectory is calculated by an analytic and nonadiabatic effective-one-body approach [49, 50].

Consistency of analytical EOB fluxes. In Sec. 7.1 we have compared and contrasted, for the first time, the numerical GW flux at scri (as computed from the TE) with two analytical prescriptions ("$\hat{\mathcal{F}}_\phi^{v_\phi}$" and "$\hat{\mathcal{F}}_\phi^{v_\Omega}$") for the EOB radiation reaction. For spins anti-aligned with the orbital angular momentum, the current, resummed, PN knowledge suffices to make the analytical fluxes agree well with the numerical data. This is in-line with the nonspinning case [50, 52, 56]. For highly negative spins, the v_ϕ-prescription is qualitatively and quantitatively closer to the numerical data in the late plunge and merger phases than the v_Ω-prescription. The reason is the intrinsic "less Keplerian" character of $\hat{\mathcal{F}}_\phi^{v_\phi}$, represented by the uniform use of $v_\phi = r_\Omega \Omega$ instead of $v_\Omega = \Omega^{1/3}$ as the argument in the Newtonian prefactors of the subdominant $(2,1)$ and $(4,4)$ multipoles. As such, the uniform use of v_ϕ in the EOB resummed (circular) multipolar waveform (and flux) seems the most natural way to incorporate the violation of Kepler's law during the plunge; being qualitatively consistent with the TE data, it offers a more suitable starting point for further improvements of the EOB insplunge waveform that are needed in the late plunge and merger phase (such as the addition of NQC corrections and the matching to the ringdown). In the case of spins aligned with the orbital angular momentum, we found that the dominant source of uncertainty in both EOB fluxes is the lack of high-order PN resummed knowledge. A crucial step for the development of next generation EOB models for spinning binaries will therefore be the update of the EOB resummed multipolar waveforms with higher order PN corrections. Similar conclusions have been drawn recently in [61] in an independent analysis based on the waveform phasing. Note that the necessary PN information to update the $\rho_{\ell m}$'s of the EOB waveform is implicitly available in the work of Shah [36] (a result that complements the 22PN-accurate result for a particle orbiting a Schwarzschild black hole by Fujita [35]). Once the various PN multipoles are available and resummed to obtain the higher order PN corrections to the $\rho_{\ell m}$'s, it will be interesting to revise our study of the span of accuracy for the radiation reaction.

Multipolar insplunge waveforms at scri. We computed multipolar waveforms spanning almost the entire black hole rotation parameter range, i.e. corotating and counterrotating backgrounds up to nearly-extremal spins of $|\hat{a}| = 0.9999$. Results are reported in Sec. 7.2 and Sec. 7.3. The multipolar hierarchy at merger has been investigated in detail as a function of the spin, considering all the multipoles up to $\ell = m = 8$. Close to merger, the subdominant multipolar amplitudes (notably the $m = 0$ ones) are enhanced for retrograde orbits with respect to prograde ones. We argue that this effect mirrors nonnegligible deviations from circularity of the dynamics during the late-plunge and merger phase (see in particular Fig. 14 and Appendix C).

The waveforms presented here are the first test-mass waveforms extracted at scri for inspirals on spinning backgrounds driven by a nonadiabatic EOB radiation reaction *. We believe these results improve quantitatively over previous calculations [60], due to the extraction at scri and the accuracy of the employed high-order finite differencing scheme. In particular, systematic uncertainties due to finite radius extraction and extrapolation are discussed in Sec. 7.4. Future work will be devoted to the analysis and the improvement of the relative EOB analytical multipolar waveform following the lines of [54, 60, 55].

Horizon-absorbed fluxes in the time-domain. For the first time to our knowledge, we have applied the time-domain formalism of Poisson [62] to calculate the GW fluxes absorbed by the horizon during the insplunge. Results are collected in Sec. 8. We calculated the horizon fluxes up to (and slightly beyond) the LSO. The TE solution indicates that during the last 25 orbits and up to the LSO the horizon-absorbed flux contributes to the total GW flux at least 3% (1%) for $\hat{a} = 0.9$ ($\hat{a} = -0.9$). These values are consistent with those from adiabatic circular orbits' fluxes. Notably, it was not possible to obtain reliable results up to the LR due to inaccuracies related to the calculation of the advanced-time integrals and the presence of the particle source term. We argued that, in our setup, a local formalism using data only from the hypersurfaces would be preferable.

In future work we will develop an improved analytical description within the EOB framework extending the work in [95, 14] to evaluate the relevance of absorbed fluxes for GW astronomy with intermediate-mass-ratio astrophysical binaries [53].

A numerical method for a $\mathcal{O}(\nu)$-consistent radiation reaction. We proposed a numerical, iterative method to calculate self-consistently the GW fluxes to first order in the symmetric mass ratio, $\mathcal{O}(\nu)$. As a proof of principle, a calculation of the $\mathcal{O}(\nu)$ consistent radiation reaction for $\hat{a} = 0.9$, including horizon absorption is presented in Sec. 9. The self-consistent simulation produces a waveform that differs by $\Delta N_{\rm gw} \sim 35.5$ gravitational wave cycles from the one using the EOB radiation reaction. Two self-consistent simulations, one with and the other without horizon absorption, differ by $\Delta N_{\rm gw} \sim 5.4$. This result further highlights the importance of horizon-absorbed fluxes during insplunge.

The method can be used alternatively to the EOB analytical radiation reaction when the analytical information is poor or not sufficient. For example, it could be employed for rapidly spinning and/or precessing binaries, for the horizon absorbed fluxes, and for generic orbits, including eccentric/scattering configurations.

* The work of [61] also computed waveforms at scri. Differently from here, however, the Teukolsky waveforms therein are generated by a radiation reaction computed numerically from frequency domain circular orbits' fluxes, and extracted with the algorithm presented in [96]. See also Sec. 7.2.

Acknowledgments

We thank Scott Hughes for making available his data on circular orbits. This work was supported in part by DFG grant SFB/Transregio 7 "Gravitational Wave Astronomy". E.H. thanks IHES for hospitality during the development of part of this work. A.Z. was supported in part by NSF grant PHY-1068881 and by a Sherman Fairchild Foundation grant to Caltech.

Appendix A. Data and Tables

Tab. A1 shows parameters and important times (crossing of LR etc.) and orbital frequencies of the dynamics. Tab. A2 lists important features of the waveforms at merger, which will be useful for waveform modeling and calibration studies. Tab. A3 shows the mild dependence of $u_{A_{22}^{\rm max}} - \hat{t}_{\Omega_{\rm orb}^{\rm max}}$ discussed in Sec. 7.2.

Table A1. Key numbers for the simulations discussed in this work (see Figs. 11, 12, 13, 14, 16). From left to right: $\hat{r}_0$ is the initial separation, $M\Omega_0 = \left(\hat{r}^{3/2} + \hat{a}\right)^{-1}$ the initial (circular) orbital frequency; $\hat{\Omega}_{\rm LSO} \equiv \hat{\Omega}\left(\hat{t}_{\rm LSO}\right)$ and $\hat{\Omega}_{\rm LR} \equiv \hat{\Omega}\left(\hat{t}_{\rm LR}\right)$ refer to the orbital frequency of the particle at the LSO and LR crossing respectively, (at $\hat{t}_{LR}$). The last column of the table lists the time corresponding to $\max(M\Omega)$. For $\hat{a} \to 1$ the trajectories stop slightly outside the LR (see discussion in Sec. 7.3). Note, how (i) for positive spins the inspiral starts already in the strong-field regime and (ii) for $\hat{a} \neq 0$, $\hat{t}_{\Omega^{\rm max}} \neq \hat{t}_{\rm LR}$ with progressively larger differences as $\hat{a} \to 1$.

$\hat{a}$	$\hat{r}_0$	$M\Omega_0$	$\hat{r}_{\rm LSO}$	$M\Omega_{\rm LSO}$	$\hat{t}_{\rm LSO}$	$\hat{r}_{\rm LR}$	$M\Omega_{\rm LR}$	$\hat{t}_{\rm LR}$	$\hat{t}_{\Omega^{\rm max}}$
-0.9999	10.00	0.03266	9.000	0.0385	6858.3	4.000	0.03846	7321.7	7321.3
-0.9995	9.90	0.03317	8.999	0.0385	5541.0	4.000	0.03846	6004.4	6004.0
-0.9990	9.80	0.03369	8.997	0.0385	4382.9	3.999	0.03847	4846.3	4845.8
-0.9950	9.75	0.03396	8.986	0.0385	3963.1	3.996	0.03854	4425.2	4424.8
-0.9900	9.50	0.03535	8.972	0.0386	1931.5	3.991	0.03863	2392.6	2392.2
-0.9700	9.40	0.03591	8.916	0.0390	1629.7	3.973	0.03898	2085.5	2085.1
-0.9500	9.50	0.03530	8.859	0.0393	2747.8	3.955	0.03934	3198.1	3197.7
-0.9000	9.50	0.03523	8.717	0.0403	3985.5	3.910	0.04025	4423.4	4423.0
-0.8000	9.20	0.03689	8.432	0.0422	3668.1	3.819	0.04222	4080.8	4080.4
-0.7000	8.90	0.03868	8.143	0.0444	3397.0	3.725	0.04436	3785.2	3784.8
-0.6000	8.60	0.04062	7.851	0.0467	3168.7	3.630	0.04673	3533.0	3532.7
-0.5000	8.30	0.04271	7.555	0.0493	2980.4	3.532	0.04934	3321.3	3321.0
-0.4000	8.00	0.04499	7.254	0.0522	2829.6	3.432	0.05224	3147.7	3147.5
-0.3000	7.70	0.04747	6.949	0.0555	2714.6	3.329	0.05548	3010.4	3010.2
-0.2000	7.40	0.05018	6.639	0.0591	2634.3	3.223	0.05913	2908.4	2908.3
-0.1000	7.10	0.05314	6.323	0.0633	2588.9	3.113	0.06328	2841.8	2841.8
0.0000	7.00	0.05399	6.000	0.0680	4076.1	3.000	0.06802	4308.4	4308.4
0.1000	6.40	0.06138	5.669	0.0735	2012.0	2.882	0.07352	2224.2	2224.3
0.2000	6.10	0.06551	5.329	0.0800	2088.2	2.759	0.07995	2281.0	2281.1
0.3000	5.80	0.07009	4.979	0.0876	2207.2	2.630	0.08762	2381.0	2381.2
0.4000	5.40	0.07723	4.614	0.0969	1862.8	2.493	0.09694	2018.3	2018.6
0.5000	5.01	0.08537	4.233	0.1085	1671.1	2.347	0.10854	1808.8	1809.2
0.6000	4.70	0.09268	3.829	0.1235	1914.2	2.189	0.12351	2034.5	2035.0
0.7000	4.10	0.11109	3.393	0.1438	1126.9	2.013	0.14379	1230.1	1230.9
0.8000	3.80	0.12184	2.907	0.1736	1571.6	1.811	0.17360	1657.3	1658.5
0.9000	3.05	0.16060	2.321	0.2251	820.7	1.558	0.22514	883.6	886.2
0.9500	3.02	0.16134	1.937	0.2732	1432.9	1.386	0.27316	1472.5	1491.6
0.9700	3.30	0.14358	1.738	0.3037	2813.7	1.296	0.30368	2841.9	2862.4
0.9900	3.01	0.16097	1.454	0.3510	2010.0	1.168	0.35101	2032.6	2058.5
0.9950	3.60	0.12779	1.341	0.3722	4914.9	1.118	0.37215	4941.1	4945.2
0.9990	3.60	0.12772	1.182	0.4137	5018.1	1.052	0.45258	×	5032.5
0.9995	3.60	0.12771	1.140	0.4308	5034.2	1.037	0.45309	×	5043.6
0.9999	3.60	0.12771	1.079	0.4537	×	1.016	0.45368	×	5052.5

Table A2. Properties of multipolar waveforms at merger for representative values of $\hat{a}$. See Table 3 for definitions. The values in brackets are the numbers found in [60].

ℓ	m	$\hat{a}$	$\Delta t_{\ell m}$	$\hat{A}^{\rm max}_{\ell m}$	$M\omega^{A^{\rm max}_{\ell m}}_{\ell m}$	$\hat{a}$	$\Delta t_{\ell m}$	$\hat{A}^{\rm max}_{\ell m}$	$M\omega^{A^{\rm max}_{\ell m}}_{\ell m}$
2	2	0.5	−7.23 (−7.22)	0.3147	0.3396	−0.5	−0.03 (−0.08)	0.2820	0.2378
2	1	0.5	3.83	0.0666	0.2912	−0.5	12.79	0.1508	0.2391
2	0	0.5	8.45	0.0155	×	−0.5	12.81	0.1108	×
3	3	0.5	−1.99	0.0576	0.5678	−0.5	2.76	0.0480	0.3916
3	2	0.5	−1.61	0.0146	0.4262	−0.5	12.15	0.0220	0.4422
3	1	0.5	3.12	0.0025	0.3514	−0.5	12.10	0.0115	0.3162
3	0	0.5	8.94	0.0004	×	−0.5	17.93	0.0091	×
4	4	0.5	0.32	0.0169	0.7954	−0.5	4.40	0.0132	0.5458
4	3	0.5	0.54	0.0046	0.6535	−0.5	10.10	0.0056	0.5934
4	2	0.5	1.06	0.0009	0.5306	−0.5	12.44	0.0027	0.5719
4	1	0.5	4.81	0.0001	0.4958	−0.5	16.38	0.0015	0.4168
4	0	0.5	8.69	2.16e-05	×	−0.5	16.22	0.0011	×
2	2	0.7	−12.74 (−12.77)	0.3228	0.3886	−0.7	0.76 (×)	0.2776	0.2279
2	1	0.7	−0.02	0.0472	0.2950	−0.7	13.62	0.1728	0.2095
2	0	0.7	6.50	0.0061	×	−0.7	13.43	0.1418	×
3	3	0.7	−5.10	0.0611	0.6505	−0.7	3.28	0.0468	0.3729
3	2	0.7	−9.57	0.0143	0.4236	−0.7	13.24	0.0241	0.4015
3	1	0.7	−1.12	0.0018	0.3196	−0.7	18.65	0.0151	0.2230
3	0	0.7	8.30	0.0001	×	−0.7	18.38	0.0138	×
4	4	0.7	−2.12	0.0183	0.9117	−0.7	4.90	0.0128	0.5192
4	3	0.7	−4.66	0.0047	0.6832	−0.7	12.92	0.0058	0.5907
4	2	0.7	−9.29	0.0007	0.4515	−0.7	17.17	0.0032	0.5043
4	1	0.7	−0.39	5.97e-05	0.3849	−0.7	16.79	0.0022	0.3207
4	0	0.7	27.91	4.68e-06	×	−0.7	16.61	0.0018	×
2	2	0.9	−39.16 (−39.09)	0.3212	0.4771	−0.9	1.54 (1.60)	0.2738	0.2198
2	1	0.9	−35.01	0.0249	0.2509	−0.9	14.36	0.1996	0.1738
2	0	0.9	0.09	0.0009	×	−0.9	14.03	0.1788	×
3	3	0.9	−18.03	0.0645	0.8013	−0.9	3.75	0.0459	0.3567
3	2	0.9	−35.46	0.0148	0.4860	−0.9	13.94	0.0267	0.3442
3	1	0.9	−26.01	0.0010	0.2711	−0.9	19.13	0.0209	0.1295
3	0	0.9	5.18	1.75e-05	×	−0.9	18.82	0.0203	×
4	4	0.9	−12.11	0.0201	1.1223	−0.9	5.37	0.0125	0.4960
4	3	0.9	−19.96	0.0052	0.7959	−0.9	13.93	0.0062	0.5318
4	2	0.9	−51.22	0.0007	0.4624	−0.9	17.65	0.0042	0.3645
4	1	0.9	−13.43	3.11e-05	0.3218	−0.9	17.18	0.0032	0.2089
4	0	0.9	21.41	4.17e-07	×	−0.9	16.98	0.0029	×

Table A3. Time lag between $\hat{\Omega}^{\rm max}$, $\hat{\Omega}^{\rm max}_{\rm orb}$ and $\hat{A}^{\rm max}_{22}$ varying $\hat{a}$. The time difference $u_{A^{\rm max}_{22}} - \hat{t}_{\Omega^{\rm max}_{\rm orb}}$ exhibits a mild dependence on $\hat{a}$ up to $\hat{a} = 0.8$.

$\hat{a}$	Δt_{22}	$\hat{t}_{\Omega^{\rm max}} - \hat{t}_{\Omega^{\rm max}_{\rm orb}}$	$u_{A^{\rm max}_{22}} - \hat{t}_{\Omega^{\rm max}_{\rm orb}}$
−0.9999	1.94	−4.09	−2.15
−0.9990	1.93	−4.09	−2.16
−0.9900	1.90	−4.07	−2.17
−0.9500	1.74	−3.97	−2.23
−0.9000	1.54	−3.84	−2.30
−0.8000	1.15	−3.56	−2.41
−0.7000	0.76	−3.26	−2.50
−0.6000	0.37	−2.93	−2.56
−0.5000	−0.03	−2.56	−2.59
−0.4000	−0.43	−2.16	−2.59
−0.3000	−0.86	−1.71	−2.57
−0.2000	−1.33	−1.21	−2.54
−0.1000	−1.83	−0.65	−2.48
0.0000	−2.38	0.00	−2.38
0.1000	−3.02	0.75	−2.27
0.2000	−3.76	1.64	−2.12
0.3000	−4.64	2.73	−1.91
0.4000	−5.76	4.12	−1.64
0.5000	−7.24	6.00	−1.24
0.6000	−9.35	8.80	−0.55
0.7000	−12.74	13.63	0.89
0.8000	−19.36	24.86	5.50
0.9000	−39.16	88.26	49.10
0.9500	−85.79	323.13	237.34
0.9900	−156.26	670.37	514.11
0.9990	−154.70	862.53	707.83
0.9999	−154.96	1122.53	967.57

Appendix B. Hamiltonian dynamics

The geodesic motion of a particle on a fixed background metric can be expressed in Hamiltonian form using coordinates q^α (e.g. $q^\alpha = (t, r, \theta, \phi)$) and the conjugate momenta p_α (e.g. $P_\alpha = (P_t, P_r, P_\theta, P_\phi)$). The affine parameter λ disappears. One starts from the four dimensional Hamiltonian $\mathcal{H}$ written as

$$\mathcal{H} = \frac{1}{2} g^{\alpha\beta} P_\alpha P_\beta = -\frac{1}{2}\mu^2 \ , \tag{B.1}$$

where the second equation comes from using the length of the geodesic as the affine parameter itself. The reduced momenta are defined as

$$\frac{dq_\alpha}{d\lambda} = \frac{P_\alpha}{\mu} \equiv p_\alpha \ , \tag{B.2}$$

which leads to the expression

$$g^{\alpha\beta} p_\alpha p_\beta = -1 \ . \tag{B.3}$$

Since the dynamics does not depend explicitly on the affine parameter λ (,i.e. the proper time) one can use the time coordinate as the integration parameter and thus reduce the geodesic equations generated by the superhamiltonian, Eq. (B.1), to just six Hamilton equations. The Hamiltonian is defined from Eq. (B.3) as

$$\hat{H} \equiv \frac{H}{\mu} = -p_0 = \frac{g^{0i}p_i}{g^{00}} + \left(\frac{g^{0i}p_i}{g^{00}} - \frac{g^{ij}p_i p_j + 1}{g^{00}} \right)^{1/2} \quad . \tag{B.4}$$

Note that $\hat{H}$ is the energy that is conserved for geodesic orbits, i.e. one of the first integrals of motion that one gets from the usual Lagrangian description of the relativistic geodesic motion. The equations of motion follow,

$$\frac{dq^i}{dt} = \frac{\partial \hat{H}}{\partial p_i} \; ; \qquad \frac{dp_i}{dt} = -\frac{\partial \hat{H}}{\partial q^i} + \hat{\mathcal{F}}_i \; . \tag{B.5}$$

We introduced explicitly at the r.h.s. of the second set of equations the dissipative terms $\hat{\mathcal{F}}_i$ that represent the radiation reaction. For example $\hat{\mathcal{F}}_\phi$ is prescribed in the EOB model as a resummed analytical expression for the radiation-reaction. The Hamiltonians relevant for this work are given below. From the Schwarzschild metric

$$ds^2 = -A_S(\hat{r})dt^2 + A_S^{-1}(\hat{r})d\hat{r}^2 + \hat{r}^2(d\theta^2 + \sin^2\theta d\phi^2) \; , \tag{B.6}$$

one gets the well-known Schwarzschild Hamiltonian

$$\hat{H}_{\rm Schw} = \sqrt{A_S(\hat{r}) \left(1 + \frac{p_\phi^2}{\hat{r}^2} \right) + A_S(\hat{r})^2 p_{\hat{r}}^2} \tag{B.7}$$

where $A_S(\hat{r}) = (1 - 2/\hat{r})$. Note that we use normalized coordinates, $\hat{r} \equiv r/M$. Moving to Kerr spacetime, the line element can be written as (we adopt a common notation, see e.g. [97])

$$ds^2 = -Adt^2 + \varpi^2(d\phi - \omega dt)^2 + \frac{\Sigma}{\Delta}d\hat{r}^2 + \Sigma d\theta^2 \; . \tag{B.8}$$

The inverse metric that enters the definition of the Hamiltonian reads

$$g^{\mu\nu} = \begin{pmatrix} -1/A & 0 & 0 & -\omega/A \\ 0 & \Delta/\Sigma & 0 & 0 \\ 0 & 0 & 1/\Sigma & 0 \\ -\omega/\alpha^2 & 0 & 0 & 1/\varpi^2 - \omega^2/A \end{pmatrix}, \tag{B.9}$$

where one has defined

$$\Sigma \equiv \hat{r}^2 + \hat{a}^2\cos^2\theta \tag{B.10}$$

$$\Delta \equiv \hat{r}^2 - 2\hat{r} + \hat{a}^2 \tag{B.11}$$

$$A \equiv \frac{\Sigma\Delta}{\Sigma\Delta + 2\hat{r}(\hat{a}^2 + \hat{r}^2)} \tag{B.12}$$

$$\omega \equiv \frac{2\hat{r}\hat{a}}{\Sigma\Delta + 2\hat{r}(\hat{a}^2 + \hat{r}^2)} \tag{B.13}$$

$$\varpi^2 \equiv \left[\frac{\Sigma\Delta + 2\hat{r}(\hat{a}^2 + \hat{r}^2)}{\Sigma} \right] \sin^2\theta \; . \tag{B.14}$$

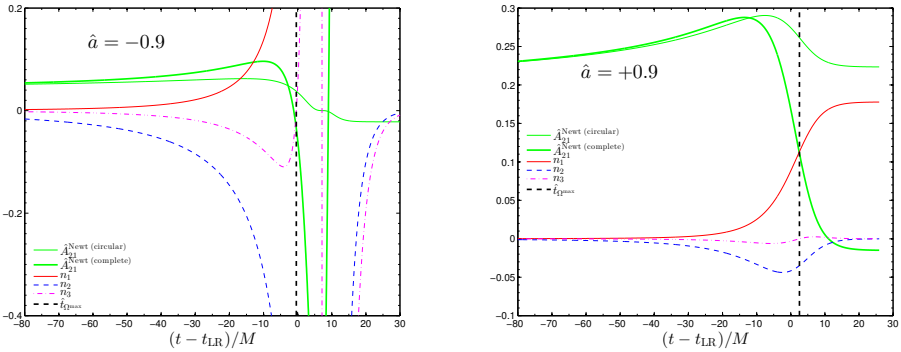

Figure C1. Effect of next-to-quasi-circular correction factors to the Newtonian (circular) waveform for $\hat{a} = \pm 0.9$. NQC terms are responsible for the sharpening of the amplitude's peak when $\hat{a} < 0$.

The Hamiltonian of a nonspinning particle on Kerr spacetime is finally written as

$$\hat{H}_{\rm Kerr} = \omega p_\phi + \sqrt{A\left(1 + \frac{p_\phi^2}{\varpi^2}\right) + \frac{A}{\Sigma} p_\theta^2 + A\frac{\Delta}{\Sigma} p_{\hat{r}}^2}\ . \tag{B.15}$$

Appendix C. Next-to-quasi-circular effects in the multipolar waveform amplitude

In Sec. 7 we pointed out that (consistently with the discussion in Ref. [61]) next-to-quasi-circular (NQC) effects related to the growth of p_{r_*} in the late-plunge and merger phase when $\hat{a} \to -1$ are responsible for the corresponding "sharpening" of the peaks of the multipolar amplitudes $\hat{A}_{\ell m}$. On the contrary, when $\hat{a} \to 1$ and p_{r_*} is almost negligible, the peak is rather flat and barely distinguishable. In other words, when $\hat{a} \to +1$ one has a long, persistent, quasi-adiabatic inspiral until the particle locks to the black hole horizon. In this Appendix we add some details, showing how NQC effects can practically act to shape the waveform amplitude around merger. For pedagogical purposes, we focus on the $\ell = 2$ modes only and use Newtonian-like waveforms to illustrate the effect. The same structural behavior is valid also for the other multipoles.

At the leading-order, Newtonian level the (μ-normalized) quadrupolar mass moments $(\hat{I}_{20}, \hat{I}_{22})$ and the mass current $\hat{S}_{21}$ moment of a particle moving along a trajectory expressed in terms of (r, ϕ) read

$$\hat{I}_{22} = \sqrt{\frac{\pi}{30}}\hat{r}^2 e^{-2{\rm i}\phi}, \tag{C.1}$$

$$\hat{S}_{21} = \sqrt{\frac{16\pi}{45}}\hat{r}^3\hat{\Omega}e^{-{\rm i}\phi}, \tag{C.2}$$

$$\hat{I}_{20} = -\sqrt{\frac{\pi}{15}}\hat{r}^2. \tag{C.3}$$

The leading-order waveform is computed taking two time derivatives of these multipoles. The numerical constants here are chosen consistently with our RWZ

normalizations. Replacing $dr/dt \to p_{r_*}$ (as they coincide in the Newtonian limit) in the two time derivatives of the multipolar moments, one finds

$$\hat{A}^{\rm Newt}_{22} = \hat{A}^0_{22}\hat{A}^{\rm NQC}_{22} = \sqrt{\frac{16\pi}{30}} v_\phi^2 \left(1 - \frac{n_1}{2} - \frac{n_2}{2}\right), \tag{C.4}$$

$$\hat{A}^{\rm Newt}_{21} = \hat{A}^0_{21}\hat{A}^{\rm NQC}_{21} = \sqrt{\frac{16\pi}{45}} v_\phi^3 \left(1 - 6n_1 - 3n_2 - 6n_3\right), \tag{C.5}$$

$$\hat{A}^{\rm Newt}_{20} = \sqrt{\frac{4\pi}{15}} \left(\dot{p}^2_{r_*} + r\dot{p}_{r_*}\right). \tag{C.6}$$

Here, the $m \neq 0$ amplitudes were factorized in a circular prefactor, $\hat{A}^0_{\ell m}$, where we replaced $r\hat{\Omega} \to r_\Omega \hat{\Omega} = v_\phi$, and a NQC correction factor, $\hat{A}^{\rm NQC}_{\ell m}$, that depends on p_{r_*} through the factors

$$n_1 = \frac{p^2_{r_*}}{(\hat{r}\hat{\Omega})^2}, \tag{C.7}$$

$$n_2 = \frac{\dot{p}_{r_*}}{\hat{r}\hat{\Omega}^2}, \tag{C.8}$$

$$n_3 = \frac{p_{r_*}}{\hat{r}\hat{\Omega}^2}\frac{\dot{\hat{\Omega}}}{\hat{\Omega}}. \tag{C.9}$$

Useful conclusions can be driven from this simple Newtonian formulas. Let us first focus on the $(2,1)$ mode end explore its dependence on $\hat{a}$. Figure C1 illustrates the effect of the NQC corrections (n_1, n_2, n_3) on the circular waveform (represented by a thin green line) when $\hat{a} = -0.9$ (left panel) and $\hat{a} = +0.9$ (right panel). Note that the vertical dashed line in the figure marks the location of the peak of $\hat{\Omega}$ and that the plot is done versus $t - t_{\rm LR}$, so to easily identify the merger. When $\hat{a} = -0.9$ one sees that the large effect of the noncircular factors (n_1, n_2, n_3) produces a sharpening of the waveform amplitude, while the time location of the peak moves to the right, closer to the light-ring crossing. On the contrary, for $\hat{a} = +0.9$ the much smaller amplitude of the NQC factor is unable to further sharpen the waveform peaks, but just results in moving it farther to the left of the light-ring crossing. This simple Newtonian-based example is helpful in understanding heuristically the role of the NQC amplitude effects incorporated in the TE waveforms displayed in Fig. 13. A similar heuristic understanding can be driven also for the other multipoles. Evidently, in the actual TE waveform the numerical coefficients in Eq. (C.5) are replaced by functions of $1/r$ so that the actual behavior is more complicated than what is discussed here, though the behavior of the NQC terms is the key element behind the sharpening of the peak of the $\hat{A}_{\ell m}$ as $\hat{a} \to -1$.

The inspection of Figs. 13 also shows that, for a given value of $\hat{a}$ (,i.e. a given behavior of p_{r_*}), and a given value of ℓ the peaks of amplitudes with $\ell = m$ are always less pronounced (with respect to the inspiral part) than the corresponding $0 \leq m < \ell$ subdominant modes. This holds true for any value of $\hat{a}$ and is explained as due to differences between the various $\hat{A}^{\rm NQC}_{\ell m}$'s. Again, the Newtonian analysis helps us understanding the key physical elements. For example, when one looks at the Newtonian NQC corrections, Eqs. (C.4) and (C.5), one sees that, for a given value of p_{r_*}, the absolute values of the coefficients of the n_i factors are *smaller* than 1 in $\hat{A}^{\rm NQC}_{21}$, while they are *larger* than 1 in $\hat{A}^{\rm NQC}_{22}$. This implies that the NQC correction

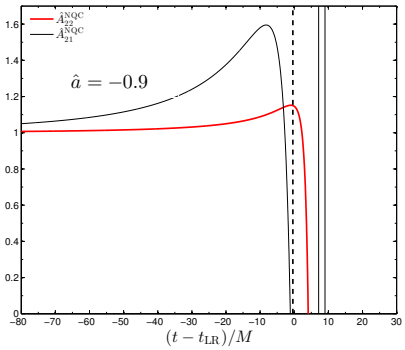

Figure C2. Comparing the NQC correction factors $\hat{A}_{22}^{\rm NQC}$ and $\hat{A}_{21}^{\rm NQC}$ for $\hat{a} = -0.9$.

factor for $(2,1)$ is larger than the one for $(2,2)$, as illustrated in Fig. C2. Note that n_3 is approximately degenerate with n_1 when using the Kepler's constraint for circular orbits: at leading order one has $\dot{\hat{\Omega}}/\hat{\Omega} = -2/3\,(p_{r_*}/r)$ that is $n_3 \approx -2/3\,n_1$, which yields $\hat{A}_{21} \propto (1 - 2n_1 - 3n_2)$. An analogous behavior is found for any other subdominant mode, in the sense that, for any ℓ, the NQC correction for $0 \leq m < \ell$ is always larger than for $\ell = m$, and progressively increasing as $m \to 0$. It is remarkable that the Newtonian analysis suffices in capturing the essential elements of the hierarchical behavior we find in the complete waveform. We postpone to further studies a detailed analysis of the effect on each multipole of the other PN-resummed corrections factors $(\hat{S}^{(\epsilon)}, T_{\ell m}, \rho^{\ell}_{\ell m})$ entering the EOB resummed waveform, Eq. (51), as well as of the corresponding NQC corrections.

References

[1] S.A. Teukolsky. Rotating black holes - separable wave equations for gravitational and electromagnetic perturbations. *Phys.Rev.Lett.*, 29:1114–1118, 1972.

[2] Saul A. Teukolsky. Perturbations of a rotating black hole. 1. Fundamental equations for gravitational electromagnetic and neutrino field perturbations. *Astrophys. J.*, 185:635–647, 1973.

[3] William H. Press and Saul A. Teukolsky. Perturbations of a Rotating Black Hole. II. Dynamical Stability of the Kerr Metric. *Astrophys. J.*, 185:649–674, 1973.

[4] Shuhei Mano, Hisao Suzuki, and Eiichi Takasugi. Analytic solutions of the Teukolsky equation and their low frequency expansions. *Prog.Theor.Phys.*, 95:1079–1096, 1996.

[5] Ryuichi Fujita and Hideyuki Tagoshi. New numerical methods to evaluate homogeneous solutions of the Teukolsky equation. *Prog.Theor.Phys.*, 112:415–450, 2004.

[6] William Krivan, Pablo Laguna, Philippos Papadopoulos, and Nils Andersson. Dynamics of perturbations of rotating black holes. *Phys. Rev.*, D56:3395–3404, 1997.

[7] Ramon Lopez-Aleman, Gaurav Khanna, and Jorge Pullin. Perturbative evolution of particle orbits around Kerr black holes: Time domain calculation. *Class.Quant.Grav.*, 20:3259–3268, 2003.

[8] Gaurav Khanna. Teukolsky evolution of particle orbits around Kerr black holes in the time domain: Elliptic and inclined orbits. *Phys.Rev.*, D69:024016, 2004.

[9] Enrique Pazos-Avalos and Carlos O. Lousto. Numerical integration of the Teukolsky equation in the time domain. *Phys.Rev.*, D72:084022, 2005.

[10] Pranesh A. Sundararajan, Gaurav Khanna, and Scott A. Hughes. Towards adiabatic waveforms for inspiral into Kerr black holes: I. A new model of the source for the time domain perturbation equation. *Phys. Rev.*, D76:104005, 2007.

[11] Pranesh A. Sundararajan, Gaurav Khanna, Scott A. Hughes, and Steve Drasco. Towards adiabatic waveforms for inspiral into Kerr black holes: II. Dynamical sources and generic orbits. *Phys.Rev.*, D78:024022, 2008.

[12] Pranesh A. Sundararajan, Gaurav Khanna, and Scott A. Hughes. Binary black hole merger gravitational waves and recoil in the large mass ratio limit. *Phys.Rev.*, D81:104009, 2010.

[13] Enrico Barausse, Vitor Cardoso, and Gaurav Khanna. Testing the Cosmic Censorship Conjecture with point particles: the effect of radiation reaction and the self-force. *Phys.Rev.*, D84:104006, 2011.

[14] Andrea Taracchini, Alessandra Buonanno, Scott A. Hughes, and Gaurav Khanna. Modeling the horizon-absorbed gravitational flux for equatorial-circular orbits in Kerr spacetime. *Phys.Rev.*, D88:044001, 2013.

[15] Enno Harms, Sebastiano Bernuzzi, and Bernd Brügmann. Numerical solution of the 2+1 Teukolsky equation on a hyperboloidal and horizon penetrating foliation of Kerr and application to late-time decays. *Class.Quant.Grav.*, 30:115013, 2013.

[16] Anil Zenginoğlu. Hyperboloidal foliations and scri-fixing. *Class. Quant. Grav.*, 25:145002, 2008.

[17] Anil Zenginoğlu, Dario Nunez, and Sascha Husa. Gravitational perturbations of Schwarzschild spacetime at null infinity and the hyperboloidal initial value problem. *Class. Quant. Grav.*, 26:035009, 2009.

[18] M. Davis, R. Ruffini, W.H. Press, and R.H. Price. Gravitational radiation from a particle falling radially into a schwarzschild black hole. *Phys.Rev.Lett.*, 27:1466–1469, 1971.

[19] M. Davis, R. Ruffini, and J. Tiomno. Pulses of gravitational radiation of a particle falling radially into a schwarzschild black hole. *Phys. Rev.*, D5:2932–2935, 1972.

[20] Steven L. Detweiler. Black Holes and Gravitational Waves. I. Circular Orbits About a Rotating Hole. *Astrophys.J.*, 225:687–693, 1978.

[21] Steven L. Detweiler and E. Szedenits. BLACK HOLES AND GRAVITATIONAL WAVES. II. TRAJECTORIES PLUNGING INTO A NONROTATING HOLE. *Astrophys.J.*, 231:211–218, 1979.

[22] Misao Sasaki and Takashi Nakamura. Gravitational Radiation From a Kerr Black Hole. 1. Formulation and a Method for Numerical Analysis. *Prog.Theor.Phys.*, 67:1788, 1982.

[23] Eric Poisson. Gravitational radiation from a particle in circular orbit around a black hole. 6. Accuracy of the postNewtonian expansion. *Phys. Rev.*, D52:5719–5723, 1995.

[24] Carlos O. Lousto and Richard H. Price. Headon collisions of black holes: The Particle limit. *Phys.Rev.*, D55:2124–2138, 1997.

[25] Karl Martel and Eric Poisson. A one-parameter family of time-symmetric initial data for the radial infall of a particle into a Schwarzschild black hole. *Phys. Rev.*, D66:084001, 2002.

[26] Karl Martel. Gravitational waveforms from a point particle orbiting a Schwarzschild black hole. *Phys. Rev.*, D69:044025, 2004.

[27] Leor Barack and Carlos O. Lousto. Perturbations of Schwarzschild black holes in the Lorenz gauge: Formulation and numerical implementation. *Phys.Rev.*, D72:104026, 2005.

[28] Carlos F. Sopuerta and Pablo Laguna. A finite element computation of the gravitational radiation emitted by a point-like object orbiting a non- rotating black hole. *Phys. Rev.*, D73:044028, 2006.

[29] Eric Poisson. Gravitational radiation from a particle in circular orbit around a black hole. 1: Analytical results for the nonrotating case. *Phys.Rev.*, D47:1497–1510, 1993.

[30] C. Cutler, Eric Poisson, G. J. Sussman, and L. S. Finn. Gravitational radiation from a particle in circular orbit around a black hole. 2: Numerical results for the nonrotating case. *Phys. Rev.*, D47:1511–1518, 1993.

[31] T. Apostolatos, D. Kennefick, Eric Poisson, and A. Ori. Gravitational radiation from a particle in circular orbit around a black hole. 3: Stability of circular orbits under radiation reaction. *Phys.Rev.*, D47:5376–5388, 1993.

[32] Eric Poisson. Gravitational radiation from a particle in circular orbit around a black hole. 4: Analytical results for the slowly rotating case. *Phys.Rev.*, D48:1860–1863, 1993.

[33] Eric Poisson and Misao Sasaki. Gravitational radiation from a particle in circular orbit around a black hole. 5: Black hole absorption and tail corrections. *Phys.Rev.*, D51:5753–5767, 1995.

[34] Misao Sasaki and Hideyuki Tagoshi. Analytic black hole perturbation approach to gravitational radiation. *Living Rev.Rel.*, 6:6, 2003.

[35] Ryuichi Fujita. Gravitational Waves from a Particle in Circular Orbits around a Schwarzschild Black Hole to the 22nd Post-Newtonian Order. *Prog.Theor.Phys.*, 128:971–992, 2012.

[36] Abhay G Shah. Gravitational-wave flux for a particle orbiting a Kerr black hole to 20th post-Newtonian order: a numerical approach. 2014.

[37] Yasushi Mino, Misao Sasaki, and Takahiro Tanaka. Gravitational radiation reaction to a particle

motion. *Phys.Rev.*, D55:3457–3476, 1997.
[38] Theodore C. Quinn and Robert M. Wald. An Axiomatic approach to electromagnetic and gravitational radiation reaction of particles in curved space-time. *Phys.Rev.*, D56:3381–3394, 1997.
[39] Leor Barack. Gravitational self force in extreme mass-ratio inspirals. *Class.Quant.Grav.*, 26:213001, 2009.
[40] Eric Poisson, Adam Pound, and Ian Vega. The Motion of point particles in curved spacetime. *Living Rev.Rel.*, 14:7, 2011.
[41] Scott A. Hughes. The Evolution of circular, nonequatorial orbits of Kerr black holes due to gravitational wave emission. *Phys.Rev.*, D61:084004, 2000.
[42] Scott A. Hughes. Evolution of circular, nonequatorial orbits of Kerr black holes due to gravitational wave emission. 2. Inspiral trajectories and gravitational wave forms. *Phys.Rev.*, D64:064004, 2001.
[43] Kostas Glampedakis, Scott A. Hughes, and Daniel Kennefick. Approximating the inspiral of test bodies into Kerr black holes. *Phys. Rev.*, D66:064005, 2002.
[44] Scott A. Hughes, Steve Drasco, Eanna E. Flanagan, and Joel Franklin. Gravitational radiation reaction and inspiral waveforms in the adiabatic limit. *Phys. Rev. Lett.*, 94:221101, 2005.
[45] Ryuichi Fujita, Wataru Hikida, and Hideyuki 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.*, 121:843–874, 2009.
[46] Pranesh A. Sundararajan. The transition from adiabatic inspiral to geodesic plunge for a compact object around a massive Kerr black hole: Generic orbits. *Phys. Rev.*, D77:124050, 2008.
[47] Tullio Regge and John A. Wheeler. Stability of a Schwarzschild singularity. *Phys. Rev.*, 108:1063–1069, 1957.
[48] Frank J. Zerilli. Effective potential for even parity Regge-Wheeler gravitational perturbation equations. *Phys. Rev. Lett.*, 24:737–738, 1970.
[49] Alessandro Nagar, Thibault Damour, and Angelo Tartaglia. Binary black hole merger in the extreme mass ratio limit. *Class. Quant. Grav.*, 24:S109–S124, 2007.
[50] Thibault Damour and Alessandro Nagar. Faithful Effective-One-Body waveforms of small-mass-ratio coalescing black-hole binaries. *Phys. Rev.*, D76:064028, 2007.
[51] Thibault Damour, Bala R. Iyer, and Alessandro Nagar. Improved resummation of post-Newtonian multipolar waveforms from circularized compact binaries. *Phys. Rev.*, D79:064004, 2009.
[52] Sebastiano Bernuzzi and Alessandro Nagar. Binary black hole merger in the extreme-mass-ratio limit: a multipolar analysis. *Phys. Rev.*, D81:084056, 2010.
[53] Sebastiano Bernuzzi, Alessandro Nagar, and Anil Zenginoglu. Horizon-absorption effects in coalescing black-hole binaries: An effective-one-body study of the non-spinning case. *Phys.Rev.*, D86:104038, 2012.
[54] Sebastiano Bernuzzi, Alessandro Nagar, and Anil Zenginoglu. Binary black hole coalescence in the extreme-mass-ratio limit: testing and improving the effective-one-body multipolar waveform. *Phys.Rev.*, D83:064010, 2011.
[55] Thibault Damour, Alessandro Nagar, and Sebastiano Bernuzzi. Improved effective-one-body description of coalescing nonspinning black-hole binaries and its numerical-relativity completion. *Phys.Rev.*, D87:084035, 2013.
[56] Sebastiano Bernuzzi, Alessandro Nagar, and Anil Zenginoglu. Binary black hole coalescence in the large-mass-ratio limit: the hyperboloidal layer method and waveforms at null infinity. *Phys.Rev.*, D84:084026, 2011.
[57] Alessandro Nagar. Gravitational recoil in nonspinning black hole binaries: the span of test-mass results. *Phys.Rev.*, D88:121501, 2013.
[58] Anil Zenginoğlu. Hyperboloidal layers for hyperbolic equations on unbounded domains. *J.Comput.Phys.*, 230:2286–2302, 2011.
[59] Anil Zenginoğlu. Asymptotics of Schwarzschild black hole perturbations. *Class. Quant. Grav.*, 27:045015, 2010.
[60] Enrico Barausse, Alessandra Buonanno, Scott A. Hughes, Gaurav Khanna, Stephen O'Sullivan, et al. Modeling multipolar gravitational-wave emission from small mass-ratio mergers. *Phys.Rev.*, D85:024046, 2012.
[61] Andrea Taracchini, Alessandra Buonanno, Gaurav Khanna, and Scott A. Hughes. Small mass plunging into a Kerr black hole: Anatomy of the inspiral-merger-ringdown waveforms. 2014.
[62] Eric Poisson. Absorption of mass and angular momentum by a black hole: Time-domain formalisms for gravitational perturbations, and the small-hole / slow-motion approximation.

Phys.Rev., D70:084044, 2004.

[63] Istvan Rácz and Gabor Zsolt Tóth. Numerical investigation of the late-time Kerr tails. *Class.Quant.Grav.*, 28:195003, 2011.

[64] Andrea Taracchini, Alessandra Buonanno, Yi Pan, Tanja Hinderer, Michael Boyle, et al. Effective-one-body model for black-hole binaries with generic mass ratios and spins. 2013.

[65] Ian Hinder, Alessandra Buonanno, Michael Boyle, Zachariah B. Etienne, James Healy, et al. Error-analysis and comparison to analytical models of numerical waveforms produced by the NRAR Collaboration. *Class.Quant.Grav.*, 31:025012, 2013.

[66] Douglas M. Eardley and Larry Smarr. Time function in numerical relativity. Marginally bound dust collapse. *Phys.Rev.*, D19:2239–2259, 1979.

[67] Vincent Moncrief. Conformally regular ADM evolution equations, 2000. Talk at Santa Barbara, `http://online.itp.ucsb.edu/online/numrel00/moncrief`.

[68] Sascha Husa. Numerical relativity with the conformal field equations. *Lect.Notes Phys.*, 617:159–192, 2003.

[69] Gyula Fodor and Istvan Racz. What does a strongly excited 't Hooft-Polyakov magnetic monopole do? *Phys. Rev. Lett.*, 92:151801, 2004.

[70] Huan Yang, Aaron Zimmerman, Anl Zenginolu, Fan Zhang, Emanuele Berti, et al. Quasinormal modes of nearly extremal Kerr spacetimes: spectrum bifurcation and power-law ringdown. *Phys.Rev.*, D88(4):044047, 2013.

[71] Ezra T. Newman and Penrose R. *J. Math. Phys.*, 3:566, 1962.

[72] Anil Zenginoğlu. A Geometric framework for black hole perturbations. *Phys. Rev.*, D83:127502, 2011.

[73] Manuela Campanelli, Gaurav Khanna, Pablo Laguna, Jorge Pullin, and Michael P. Ryan. Perturbations of the Kerr space-time in horizon penetrating coordinates. *Class.Quant.Grav.*, 18:1543–1554, 2001.

[74] Anil Zenginoğlu and Manuel Tiglio. Spacelike matching to null infinity. *Phys. Rev.*, D80:024044, 2009.

[75] Karl Martel and Eric Poisson. Gravitational perturbations of the schwarzschild spacetime: A practical covariant and gauge-invariant formalism. *Physical Review D (Particles, Fields, Gravitation, and Cosmology)*, 71(10):104003, 2005.

[76] Alessandro Nagar and Luciano Rezzolla. Gauge-invariant non-spherical metric perturbations of Schwarzschild black-hole spacetimes. *Class. Quant. Grav.*, 22:R167, 2005.

[77] S.W. Hawking and J.B. Hartle. Energy and angular momentum flow into a black hole. *Commun.Math.Phys.*, 27:283–290, 1972.

[78] James M. Bardeen, B. Carter, and S.W. Hawking. The Four laws of black hole mechanics. *Commun.Math.Phys.*, 31:161–170, 1973.

[79] S.A. Teukolsky and W.H. Press. Perturbations of a rotating black hole. III - Interaction of the hole with gravitational and electromagnetic radiation. *Astrophys.J.*, 193:443–461, 1974.

[80] R.H. Price and K.S. Thorne. Membrane Viewpoint on Black Holes: Properties and Evolution of the Stretched Horizon. *Phys.Rev.*, D33:915–941, 1986.

[81] A.-K. Tornberg and B. Engquist. Numerical approximations of singular source terms in differential equations. *Journal of Computational Physics*, 200:462–488, November 2004.

[82] B. Engquist, A.-K. Tornberg, and R. Tsai. Discretization of Dirac delta functions in level set methods. *Journal of Computational Physics*, 207:28–51, July 2005.

[83] Donato Bini and Thibault Damour. Gravitational radiation reaction along general orbits in the effective one-body formalism. *Phys.Rev.*, D86:124012, 2012.

[84] Thibault Damour and Achamveedu Gopakumar. Gravitational recoil during binary black hole coalescence using the effective one body approach. *Phys. Rev.*, D73:124006, 2006.

[85] Thibault Damour and Alessandro Nagar. Binary black hole merger waveforms in the extreme mass ratio limit. pages 1636–1640, 2006.

[86] Thibault Damour and Alessandro Nagar. An improved analytical description of inspiralling and coalescing black-hole binaries. *Phys. Rev.*, D79:081503, 2009.

[87] Yi Pan, Alessandra Buonanno, Michael Boyle, Luisa T. Buchman, Lawrence E. Kidder, et al. Inspiral-merger-ringdown multipolar waveforms of nonspinning black-hole binaries using the effective-one-body formalism. *Phys.Rev.*, D84:124052, 2011.

[88] Yi Pan, Alessandra Buonanno, Ryuichi Fujita, Etienne Racine, and Hideyuki Tagoshi. Post-Newtonian factorized multipolar waveforms for spinning, non-precessing black-hole binaries. *Phys.Rev.*, D83:064003, 2011.

[89] Ryuichi Fujita and Bala R. Iyer. Spherical harmonic modes of 5.5 post-Newtonian gravitational wave polarizations and associated factorized resummed waveforms for a particle in circular orbit around a Schwarzschild black hol. *Phys. Rev.*, D82:044051, 2010.

[90] Emanuele Berti, Vitor Cardoso, and Andrei O. Starinets. Quasinormal modes of black holes and black branes. *Class. Quant. Grav.*, 26:163001, 2009.

[91] Nicolas Yunes, Alessandra Buonanno, Scott A. Hughes, Yi Pan, Enrico Barausse, et al. Extreme Mass-Ratio Inspirals in the Effective-One-Body Approach: Quasi-Circular, Equatorial Orbits around a Spinning Black Hole. *Phys.Rev.*, D83:044044, 2011.

[92] Thibault Damour and Alessandro Nagar. TbA. *In preparation*, 2014.

[93] Nils Andersson and Kostas Glampedakis. A Superradiance resonance cavity outside rapidly rotating black holes. *Phys.Rev.Lett.*, 84:4537–4540, 2000.

[94] Kostas Glampedakis and Nils Andersson. Late time dynamics of rapidly rotating black holes. *Phys.Rev.*, D64:104021, 2001.

[95] Alessandro Nagar and Sarp Akcay. Horizon-absorbed energy flux in circularized, nonspinning black-hole binaries and its effective-one-body representation. *Phys.Rev.*, D85:044025, 2012.

[96] Anil Zenginoglu and Gaurav Khanna. Null infinity waveforms from extreme-mass-ratio inspirals in Kerr spacetime. *Phys.Rev.*, X1:021017, 2011.

[97] Jeremy D. Schnittman and Edmund Bertschinger. A model for high frequency quasi-periodic oscillations from accreting black holes. *Astron.Astrophys.*, 606:1098, 2004.