Small mass plunging into a Kerr black hole: Anatomy of the inspiral-merger-ringdown waveforms

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1103/PhysRevD.90.084025">http://dx.doi.org/10.1103/PhysRevD.90.084025</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>American Physical Society</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/98061">http://hdl.handle.net/1721.1/98061</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Small mass plunging into a Kerr black hole: Anatomy of the inspiral-merger-ringdown waveforms

Andrea Taracchini,1 Alessandra Buonanno,1 Gaurav Khanna,2 and Scott A. Hughes3

1Maryland Center for Fundamental Physics and Joint Space-Science Institute, Department of Physics, University of Maryland, College Park, Maryland 20742, USA
2Department of Physics, University of Massachusetts Dartmouth, North Dartmouth, Massachusetts 02747, USA
3Department of Physics and MIT Kavli Institute, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

(Received 9 April 2014; published 15 October 2014)

We numerically solve the Teukolsky equation in the time domain to obtain the gravitational-wave emission of a small mass inspiraling and plunging into the equatorial plane of a Kerr black hole. We account for the dissipation of orbital energy using the Teukolsky frequency-domain gravitational-wave fluxes for circular, equatorial orbits, down to the light-ring. We consider Kerr spins $-0.99 \leq q \leq 0.99$, and compute the inspiral-merger-ringdown $(2,2)$, $(2,1)$, $(3,3)$, $(3,2)$, $(4,4)$, and $(5,5)$ modes. We study the large-spin regime, and find a great simplicity in the merger waveforms, thanks to the extremely circular character of the plunging orbits. We also quantitatively examine the mixing of quasinormal modes during the ringdown, which induces complicated amplitude and frequency modulations in the waveforms. Finally, we explain how the study of small mass-ratio black-hole binaries helps extending effective-one-body models for comparable-mass, spinning black-hole binaries to any mass ratio and spin magnitude.

DOI: 10.1103/PhysRevD.90.084025 PACS numbers: 04.25.D-, 04.25.dg, 04.25.Nx, 04.30.-w

I. INTRODUCTION

Over the last few years, analytical and numerical studies have revealed interesting features of the dynamics and gravitational radiation of extreme mass-ratio black-hole binaries, especially during ringdown (RD) and when the spin of the central black hole (BH) is close to maximal, and the orbits approach the horizon. References [1–4] pointed out the possibility of describing analytically various processes of the dynamics and radiation in the near-horizon region of a nearly extremal black hole by exploiting an infinite-dimensional conformal symmetry that the Kerr metric satisfies in this particular limit. Applying the WKB method to the Teukolsky equation in the eikonal approximation, Ref. [5] found a geometric interpretation of the black-hole quasinormal modes (QNMs) through spherical light-ring orbits, extending to generic orbits what was previously derived for equatorial [6,7] and polar orbits [8]. Moreover, an interesting bifurcation leading to a splitting of zero and nonzero damped QNMs was found as one approaches nearly extremal spins [9,10]. Quite interestingly, Refs. [11,12] found that damped modes different from the usual QNMs are present in the gravitational-radiation spectrum close to the black-hole horizon. It remains an open question whether those damped modes are excited as a test body plunges into the central black hole.

Furthermore, gravitational waveforms emitted during the inspiral, plunge and merger stages of a test body orbiting a Kerr black hole have been exploited to grasp unique, physical information on the merger phase and they have been employed to extend analytical models, notably the effective-one-body (EOB) model [13,14], from the comparable-mass to the test-particle limit case [15–24]. Solving the time-domain Regge-Wheeler or Teukolsky equations is significantly less expensive than evolving a black-hole binary in full numerical relativity. The possibility of using the test-particle limit to infer crucial information about the merger waveform of bodies of comparable masses follows from the universality of the merger process throughout the binary parameter space.

In Ref. [20], some of us investigated the inspiral-merger-ringdown waveforms produced by the time-domain Teukolsky equation where the source term is evaluated along the quasicircular plunging trajectory of a nonspinning test particle inspiraling in the equatorial plane. The trajectory was computed by solving Hamilton’s equations in the Kerr spacetime, augmented by a suitable radiation-reaction force, notably the one constructed from the factorized energy flux of the EOB formalism [25,26]. The Teukolsky waveforms were then used to improve spinning EOB waveforms during the transition from plunge merger to ringdown. However, the study of Ref. [20] was limited to moderate spins of the Kerr black hole, i.e., $a/M \leq 0.8$. Here, we build on Ref. [20], and extend the analysis in a few directions. First, the analytical energy flux based on spinning, factorized multipolar waveforms [25,26] can differ from the Teukolsky flux; for instance, even for a moderate spin value of 0.7, the modeling error at the innermost stable circular orbit is as large as 10%. This error comes from a combination of insufficient
knowledge of high-order post-Newtonian (PN) terms, and from the truncation at modes with $\ell = 8$. As the spin increases, the motion becomes more relativistic and a growing number of modes are excited. Therefore, to overcome this problem, in the equations of motion for the orbital dynamics of the plunging particle, we employ the energy flux computed by a highly accurate frequency-domain Teukolsky code [27,28]. Second, we consider spins in the range $-0.99 \leq a/M \leq 0.99$, but investigate in greater detail spins close to extremal, for prograde and retrograde orbits. In fact, those almost-extremal cases display peculiar features in the dynamics and waveforms. When the spin is close to 1, the merger waveforms are particularly simple, with a remarkably flat amplitude, as a consequence of the circular nature of the plunge. When the spin is close to $-1$, instead, the phenomenon of QNM mixing dominates the ringdown waveforms. Third, we use those findings to suggest a new procedure for modeling the transition from merger to ringdown in the EOB waveforms for spins larger than 0.8 and mass ratios smaller than $\sim 1/100$. Preliminary results of this paper were employed in Ref. [29] to build a spinning EOB model that is valid for any mass ratio and spin magnitude.

This paper is organized as follows. In Sec. II we describe how we build the orbital dynamics to compute the quasicircular plunging trajectory that is used in the source term of the Teukolsky equation. In Sec. III we review the time-domain Teukolsky code which computes the waveforms. In Sec. IV we describe interesting features characterizing the dynamics and the merger waveforms for spins close to extremal. In Sec. V we carry out a detailed study to understand and model the mixing of QNMs for the dominant (2,2), (2,1), (3,3), (3,2), (4,4), and (5,5) waveforms. In Sec. VI we explain how the information obtained from the Teukolsky waveforms has been used to design a new way of generating the EOB merger-ringdown waveform for spins larger than 0.8 and mass ratios smaller than $\sim 1/100$. In Sec. VII we compare spinning EOB waveforms developed in the comparable-mass regime [29] to the Teukolsky waveforms. Section VIII summarizes our main conclusions and discusses future directions. Appendix A provides numerical information about the Teukolsky merger waveforms that can be incorporated in generic spinning EOB models.

Henceforth, we use geometric units with $G = c = 1$.

II. ORBITAL DYNAMICS TO GENERATE INSPIRAL-MERGER-RINGDOWN TEUKOLSKY WAVEFORMS

In this section we review how the trajectory entering the source term of the Teukolsky equation is computed. We restrict our attention to systems where the smaller BH is nonspinning, and the orbits are confined to the equatorial plane of the larger, spinning BH. Let $\mu$ be the mass of the smaller object, and let $M$ and $J \equiv aM \equiv qM^2$ (with $-1 \leq q \leq 1$) be the mass and spin of the larger one. In this paper we consider systems with $\mu/M = 10^{-5}$. In the spirit of the EOB formalism, and as in Ref. [20], we model the orbital dynamics using the Hamiltonian of a nonspinning test particle of mass $\mu$ in the Kerr spacetime

\[
H = \beta^i p_i + \alpha \sqrt{\mu^2 + \gamma^{ij} p_i p_j},
\]

where $\alpha \equiv (-g^{tt})^{-1/2}$, $\beta^i \equiv g^{tt} g^{ij}$ and $\gamma^{ij} \equiv g^{ij} - g^{tt} g^{ij}/g^{tt}$, $i, j$ are spatial indices, $t$ is the time index, $g_{ij}$ is the Kerr metric in Boyer-Lindquist coordinates, and the $p_i$’s are the conjugate momenta to the spatial coordinates. We numerically solve Hamilton’s equations for $H$ subject to a radiation-reaction force $F$ which describes the dissipation of energy into gravitational waves (GWs); the radiation-reaction force is proportional to the sum of the GW energy flux at infinity, $F^\infty$, and through the horizon, $F^H$. It reads [30]

\[
F = \frac{F^\infty + F^H}{\Omega |r \times \mathbf{p}|} \mathbf{p},
\]

where $F \equiv F^\infty + F^H$, $r$ is the separation vector, and $\Omega \equiv J \cdot (r \times \dot{r})/r^2$ is the orbital frequency, where $\mathbf{J}$ is the unit vector along the spin of the Kerr BH. We indicate with an over-dot the derivative with respect to time $t$.

Some of us, in Ref. [20], employed the outgoing factorized energy flux of Ref. [26] for the term $F^\infty$, while setting $F^H = 0$; that choice was motivated partly by the focus on understanding the effect of the model flux, and partly by the availability of numerical Teukolsky energy fluxes only down to the innermost stable circular orbit (ISCO). Here, instead, we are mainly interested in the characterization of the Teukolsky waveforms, and we want to remove any modeling error from the orbital motion. Similarly to what is done in Ref. [31], we source our equations of motion with GW energy fluxes computed in perturbation theory; in particular, we use the Teukolsky fluxes of Ref. [24], where we numerically solved the Teukolsky equation in frequency domain [27,28] for circular, equatorial orbits all the way down to a radial separation of $r_{\min} = r_{LR} + 0.01M$, where $r_{LR}/M \equiv 2 + 2 \cos \left[\frac{3}{2} \arccos (-q)\right]$ is the position of the photon orbit, or light-ring (LR) [32]. The GW fluxes were computed for spins from $q = -0.9$ up to $+0.9$ in steps of 0.1, and also for $q = \pm 0.95, \pm 0.99$. Those computations assumed circular orbits, for which a precise relation

\[1\text{Positive (negative) values of } q \text{ indicate that the spin of the Kerr BH is aligned (antialigned) with the inspiral orbital angular momentum, i.e., the motion is prograde (retrograde) during the inspiral. At the end of the plunge, because of frame dragging, the trajectory always becomes prograde.}

\[2\text{The GW energy flux falling into the horizon is also referred to as “ingoing flux,” “absorption flux,” or “horizon flux.”} \]
between radius $r$ and orbital frequency $\Omega_{\text{circ}}$ holds, namely $M \Omega_{\text{circ}} = [(r/M)^{3/2} + q]^{-1}$.

To accurately describe the transition from inspiral to plunge, we adopt here the same strategy used in the EOB models of comparable-mass BH binaries [33–35]. First, if we introduce the velocity parameter $v_\Omega \equiv (M \Omega)^{1/3}$, then the total GW flux for circular orbits can be written as $F = 32 \mu^2 v_\Omega^3 \bar{F}(v_\Omega)/(5M^2)$, where $\bar{F}(v_\Omega) = 1 + \mathcal{O}(v_\Omega^2)$. Second, we replace $v_\Omega$ in the leading term of $F$ with the non-Keplerian velocity for a circular orbit defined by $v_\phi = \Omega R$, where $\Omega M = (M \Omega_{\text{circ}})^{-2/3}$ [see also Eq. (32) in Ref. [35]]; note that since we work with nonadiabatic\(^3\) orbital evolutions $\Omega \neq \Omega_{\text{circ}}$. This replacement moderates the growth of the GW frequency close to merger [33], and allows a more accurate modeling of numerical-relativity waveforms in the comparable-mass regime, also when spins are present [35].

We need to integrate the equations of motion to the event horizon, $r_+ / M = 1 + \sqrt{1 - q^2}$. Teukolsky fluxes are only available, however, down to the radius $r_{\text{min}} = r_{\text{LR}} + 0.01$: Circular orbits do not exist at radii $r < r_{\text{LR}}$, and the growing number of significant multipolar contributions force us to terminate our flux calculations slightly outside $r_{\text{LR}}$ (see Sec. III of Ref. [24] for detailed discussion). Note that the radial distance between $r_{\text{min}}$ and $r_+$ decreases from $3M$ when $q = 1$ down to $1M$ when $q = 0$, and vanishes for $q = 0$. Therefore, we have to provide a prescription for $\bar{F}$ in the interval $r_{\text{min}} < r < r_+$. Since these values of $r$ are well within the plunge phase, where the conservative part of the dynamics is known to dominate, we decide to smoothly switch off the GW flux at $r_{\text{end}} = r_{\text{min}}$. Let $v_{\text{end}}$ be the velocity of a circular orbit of radius $r_{\text{end}}$. Explicitly, if $r < r_{\text{end}}$ but $v_\Omega \leq v_{\text{end}}$, then we suppress $\bar{F}(v_\Omega)$ by a factor $1/[1 + \exp(-(r - r_{\text{end}})/\sigma)]$; if $r < r_{\text{end}}$ and $v_\Omega > v_{\text{end}}$, then we set $\bar{F}(v_\Omega) = \bar{F}(v_{\text{end}})/[1 + \exp(-(r - r_{\text{end}})/\sigma)]$. We find that, as long as $\sigma \lesssim 0.01M$, the trajectories are insensitive to the specific value of $\sigma$. We test the effect of the switch-off point by changing its position to $r_{\text{end}} = r_{\text{LR}} + b(r_{\text{ISCO}} - r_{\text{LR}})$, where $r_{\text{ISCO}}$ is the position of the ISCO, and $b = 0.25, 0.5, 0.75$, for spins $q = 0.5, 0.9$; the difference in the orbital phase is always negligible, within 0.003 (0.006) rads for $q = 0.5$ (0.9) when $b = 0.75$ with respect to the fiducial case $r_{\text{end}} = r_{\text{min}}$ (i.e., $b \approx 0$), since the plunging motion is indeed geodetic to a good approximation, and is not affected by the details of the GW fluxes.

As in Ref. [20], we compute the trajectory from the equations of motion down to a point slightly outside the horizon (at $\sim 1.05 r_+$). Then, to model the locking of the plunging particle to the rotating horizon, we smoothly connect the trajectory obtained by solving Hamilton’s equations to several orbital cycles at $r = r_+$ with frequency

\(^3\)This means that the orbital motion includes not only tangential, but also radial velocities.

### III. Numerical Solution of the Time-Domain Teukolsky Equation

In this section we review the numerical method used to solve the Teukolsky equation in the time domain. The approach we follow to solve this linear partial differential equation (PDE) is the same as presented in our earlier work (see Ref. [20] and references therein). The main points of this technique are as follows: (i) We first rewrite the Teukolsky equation using suitable coordinates—the tortoise radius $r^\tau$ and Kerr azimuthal angle $\varphi$, defined precisely in [20]. (ii) Taking advantage of axisymmetry, we separate the dependence on azimuthal coordinate $\varphi$. We thus obtain a set of $2 + 1$ dimensional equations. (iii) We recast these equations into a first-order, hyperbolic PDE form. (iv) Finally, we implement a two-step, second-order Lax-Wendroff, time-explicit, finite-difference numerical scheme. The particle-source term on the right-hand side of the Teukolsky equation requires some care for such a numerical implementation. All relevant details can be found in our earlier work [20] and the associated references.

Since Ref. [20] was published, two technical advances have been introduced into the solver code aimed at improving results for the present paper. First, a compactified hyperboloidal layer has been added to the outer portion of the computational domain [36]. This advancement allows us to map null infinity onto the computational grid and also completely solves the so-called “outer-boundary problem” (i.e., it eliminates unphysical reflections from the artificial boundary of the domain). Therefore, differently from Ref. [20], we are now able to extract gravitational waveforms directly at null infinity, completely eliminating the “extraction error,” as discussed in Ref. [20]. Secondly, we have taken advantage of advances made in parallel computing hardware, and we have developed a very high-performing OpenCL implementation of the Teukolsky code that takes full benefit of GPGPU-acceleration and cluster computing. Details on this parallel implementation and careful measurements of gains in overall performance can be found in Ref. [37].

These advances have helped improve the performance and accuracy of the time-domain Teukolsky code by several orders of magnitude over previous versions. In particular, Ref. [37] demonstrated that errors with the improved code are typically at the level of 0.01\%, an order of magnitude better than earlier versions [20], while performing faster. For long evolutions, these improvements yield a several thousand-fold speedup [36]. Consider the impact of such improvements on modeling the evolution of a system for $20,000 M$, a typical span for our studies. With our previous Cauchy-evolution-based Teukolsky code, we would need to place the outer boundary at $r \gtrsim 10,000 M$ to

---

084025-3
avoid impact of boundary effects—outside the domain of causal influence for the location and duration of interest. Using hyperboloidal slicing, the outer boundary can be placed as close as 50M [36]. This immediately gains 2 orders of magnitude in performance, while generating waveforms directly at null infinity as desired. In addition, the use of GPGPU computer hardware acceleration typically yields another order of magnitude gain in performance through many-core parallelism [37].

Since we now compute the waveforms exactly at null infinity (eliminating the extraction and outer-boundary related errors entirely), the only remaining source of numerical error is the "discretization error" introduced by the finite-difference numerical scheme. It is relatively straightforward to estimate this discretization error: We first compute the waveforms at multiple grid resolutions, in particular we choose \((dr, \theta) = (M/80, \pi/128), (M/40, \pi/64)\) and \((M/20, \pi/32)\). Second, we derive the Richardson extrapolant using this data. Then, we simply use this extrapolant as a reference to estimate the discretization error in the original waveforms computed by our code. In other words, we take the relative difference between the highest resolution data and the Richardson extrapolant as a measure of the discretization error. As done typically in the literature, we decompose the waveforms in \(-2\)-spin-weighted spherical-harmonic modes, labeled by \((\ell, m)\). In Fig. 1 we depict the discretization errors for the phase and the amplitude for one particular choice of the spin. These results should be considered representative of all the other cases that we present in this work. Figure 1 demonstrates that the numerical error in our waveform data is at a level of a few \(\times 0.1\%\). As expected, the relative error is generally lower for the dominant modes such as \(h_{32}\) and \(h_{33}\), and higher for the weaker ones. In addition, the error levels stay very uniform during the long inspiral phase of the binary evolution and only begin to vary significantly during the plunge. This happens due to the fact that the numerical computation shifts from being dominated by the particle-source term during inspiral, to a nearly source-free evolution during and after the plunge phase. It should be noted that the numerical errors can be further reduced by an order of magnitude as demonstrated in Ref. [37], through an increase in grid resolution. However, given the large number and long duration of the evolutions presented in this work, reducing the numerical error further was neither very practical nor needed.

Besides the discretization error discussed above, other (controlled) sources of error are (i) the adiabatic treatment of the energy fluxes, i.e., the use of circular-orbit fluxes for inspiraling and plunging orbits, and (ii) the integration error on the equations of motion describing the particle’s trajectory. We found that little changes in the fluxes during the plunge (i.e., where nonadiabatic effects are the largest) do not affect the waveform to any significant extent, because below the ISCO the motion is well approximated by a geodesic. So we expect that including nonadiabatic effects in the energy flux is not going to affect our

---

4Note that representing the singular particle-source terms using a narrow Gaussian distribution or a discrete delta [38,39] is potentially an additional source of error. However, that error is included in our measured discretization error, because that approximation is dependent on the discretization scale itself. In other words, our numerical representation of the singular particle-source terms converges to the true expression with the increase in grid resolution.
conclusions. For the numerical integration of the equations of motion, we used a 4th order Runge-Kutta integrator with adaptive step size, with a tolerance set to $10^{-8}$. The evolutions were typically from 10,000 to 60,000 integrator steps in length. We checked that a smaller tolerance ($10^{-9}$) would yield waveforms with a dephasing within 0.01 rads at merger with respect to our choice of $10^{-8}$.

IV. SIMPLICITY OF INSPIRAL-PLUNGE TEUKOLSKY WAVEFORMS FOR LARGE SPINS

In this section we characterize the salient features displayed by the Teukolsky waveforms during late inspiral and plunge. For spins $q = 0, \pm 0.5, \pm 0.7, \pm 0.8, \pm 0.9, \pm 0.95$, and $\pm 0.99$, we compute the Teukolsky (2,2), (2,1), (3,3), (3,2), (4,4), and (5,5) modes as explained in Secs. II and III. In the test-particle limit, as the spin increases, more and more $(\ell, m)$ modes become important at merger [with respect to the (2,2) mode]; however, for the modeling of comparable-mass BH binaries, only the few modes above give significant contribution to the energy flux. Eventually, we are interested in exploiting the results of this paper in the comparable-mass limit, therefore we restrict the discussion to those modes.

In the large-spin regime, a prograde inspiraling particle reaches very relativistic speeds before getting to the horizon; for instance, when $q = 0.99$, the peak speed (attained at the peak of the orbital frequency) is around 0.75. At such speeds, the PN expansion is inadequate for analytically describing such systems. However, the Teukolsky inspiral-merger waveforms turn out to be extremely simple. Consider, for example, the (2,2) mode emitted when $q = 0.99$, shown in Fig. 2. The prominent feature that we recognize is the extreme flatness of the amplitude versus time, across hundreds of $M$, well before the plunge starts at the ISCO. The GW frequency $\omega_{T22}^{\text{Teuk}}$, defined as $-3(\dot{\mathcal{H}}_{T22}^{\text{Teuk}}/\mathcal{H}_{T22}^{\text{Teuk}})$, does not display any particular characteristic, and we notice that it is well approximated by twice the orbital frequency even during ringdown, thanks to the fact that $2\Omega_{H}$ is very close to the least-damped quasinormal mode. We find that the flattening of the amplitudes $|h_{T22}^{\ell m}|$ around their respective peaks is more and more apparent as $q$ approaches 1. This aspect of the numerical waveforms does not depend on minute details of the flux used to generate the underlying orbital dynamics. In Fig. 3 we show the amplitudes of the Teukolsky (2,2) modes for $q = 0.7, 0.8, 0.85, 0.9, 0.95$ aligned at $t_{\text{peak}}^{22}$, with $t_{\text{peak}}^{22}$ being the time when the $(\ell, m)$ mode reaches its maximum amplitude. The almost-extremal case $q = 0.99$ was not included in Fig. 3 since its (2,2) amplitude is so flat that it is quite difficult to localize $t_{\text{peak}}^{22}$. In fact, across the (2,2) peak, over a large time interval, its $|\partial_{t}|h_{22}^{\text{Teuk}}|$ is so small that it is dominated by numerical noise, making it difficult to clearly locate its zero. The curvature $(|\partial_{t}^{2}|h_{22}^{\text{Teuk}}|)_{\text{peak}}$ becomes vanishingly small as $q \rightarrow 1$; see also Fig. 14. Although we have shown only (2,2) mode waveforms, the same holds true for higher harmonics.

We can find a physical explanation for why this happens considering the underlying orbital dynamics. As the spin grows larger, the ISCO moves to smaller separations and gets closer to the horizon, so that the plunging phase becomes shorter (in the radial coordinate), and moves to higher frequencies. This is equivalent to saying that Kerr BHs with larger spins support longer quasicircular inspirals given the same initial frequency. For instance, let us consider

$$a / M = 0.99, (2,2) \text{ mode}$$

![Fig. 2 (color online). Late inspiral, plunge, merger and ringdown of the Teukolsky $h_{T22}^{\text{Teuk}}$ waveform (upper panel), its GW frequency $\omega_{T22}^{\text{Teuk}}$ and orbital frequency $\Omega$ of the underlying dynamics (lower panel) for spin $q = 0.99$. We note the simplicity of the amplitude during the last phase of the evolution. The plot spans a radial range from $r = 2.21M$ to the horizon, located at $r = 1.14M$; here, $r_{\text{ISCO}} = 1.45M$ and $r_{\text{LR}} = 1.17M$. Vertical dashed lines mark the position of the ISCO and the light-ring. A vertical green line marks the position $r_{\text{match}}^{22} = r_{\text{peak}}^{22} + \Delta r_{\text{peak}}^{22}$ of the ringdown matching as prescribed in the EOB model of Ref. [29] (see discussion in Sec. VI). $R$ is the distance to the source.](084025-5)
spins 0.5 and 0.99. Their dimensionless horizon frequencies are 0.13 and 0.43 respectively. An initial orbital frequency of 0.1 corresponds to radial separations, 4.5M and 4.3M, respectively, which are quite close to each other; while for spin 0.5 we are sitting just outside the ISCO \( [r_{\text{ISCO}}(q = 0.5) = 4.2M] \), for spin 0.99 we are still far from it \( [r_{\text{ISCO}}(q = 0.99) = 1.5M] \). Furthermore, for very large spins the orbital timescale \( T_{\text{orb}} \) is much shorter than the radiation-reaction timescale \( T_{\text{rad}} \). We can estimate these characteristic timescales for different values of \( q \) as \( T_{\text{orb}} = 2\pi / \Omega \) and \( T_{\text{rad}} = -r / i \). The orbital frequency grows during the inspiral, reaches a peak value \( \Omega_{\text{peak}} \) at time \( t^\Omega_{\text{peak}} \), and eventually converges to the horizon frequency \( \Omega_{\text{H}} \) at late times. One can show that, for all practical purposes, the peak of \( \Omega \) occurs at a radius \( r^\Omega_{\text{peak}} \) which nearly coincides with \( r_{\text{LR}} \), the coincidence being exact for \( q = 0, 1 \). In Fig. 4 we plot the ratio \( T_{\text{rad}} / T_{\text{orb}} \) as a function of the radial separation \( r \). The solid lines are computed along nonadiabatic trajectories from the numerical integration of the equations of motion, up to the peak of the orbital frequency \( \Omega \). At fixed \( r \), the orbital timescale \( T_{\text{orb}} \) does not vary much with \( q \); for example, when \( r = 4M \), \( T_{\text{orb}} = 53M \) for \( q = 0.5 \), while \( T_{\text{orb}} = 56M \) for \( q = 0.99 \); but the ratio \( T_{\text{rad}} / T_{\text{orb}} \) for spin 0.99 is 55 times larger than for spin 0.5. Hence, the plot demonstrates that there is a clear hierarchy in the radiation-reaction timescales: the larger the spin, the larger \( T_{\text{rad}} \). As a result, the secular evolution is much slower for large spins, given the same initial separation. This hierarchy can be easily understood using analytical considerations at leading order. During the quasicircular inspiral we have \( T_{\text{orb}} \approx 2\pi / \Omega_{\text{circ}} \) and the orbital energy \( E \) can be approximated by the energy of a circular orbit in Kerr spacetime \( [32] \):

\[
\frac{E_{\text{circ}}}{\mu} = \frac{1 - 2M/r + q(M/r)^{3/2}}{\sqrt{1 - 3M/r + 2q(M/r)^{3/2}}}.
\]

Note that \( E_{\text{circ}} \) diverges at \( r = r_{\text{LR}} \). Moreover, assuming mainly leading quadrupolar energy loss \([40]\) and circularity, we get \( F = -\dot{E} \approx 32\mu^2 r^4 \Omega_{\text{circ}}^6 / 5 \); thus, we find that

\[
T_{\text{rad}} = -r \frac{dE}{dr} \approx \frac{dE_{\text{circ}}}{dr} \approx \frac{32}{5} \mu^2 r^3 \Omega_{\text{circ}}^4.
\]

In Fig. 4 we plot the analytical estimate (4) with dashed lines, and find that it captures the numerical result (solid lines) fairly well at large \( r \), and, most importantly, can account for the hierarchy of the curves due to the presence of spin. We can now understand why large-spin waveforms are so flat. For large \( q \) the radiation-reaction timescale is much larger than the orbital timescale, which means that the particle performs many orbits while sweeping very slowly through the frequency range up to the horizon, so that the secular evolution of the emitted GW signal is much slower as compared to systems with smaller \( q \). This is consistent with the behavior of the frequency-domain Teukolsky fluxes that we employ in the equations of motion, whose (2,2) component is plotted in Fig. 5 versus radius; at fixed \( r \), the dissipation of energy is smaller for larger spins.

Notice that, only for this plot, we include spins as large as \( q = -0.9999, 0.99999 \). Interestingly, as \( q \to 1 \) the
fluxes become small even outside the ISCO and approach vanishingly small values beyond the ISCO, which accounts for the behavior of the ratio $T_{\text{rad}} / T_{\text{orb}}$ in the late inspiral and plunge. It is well known [41–44] that $F_{22}/\mu^2$ should diverge at the light-ring as \((E_{\text{circ}}/\mu)^2 \sim (r - r_{LR})^{-1}\). We notice that, starting around spin 0.99, $F_{22}/\mu^2$ does not diverge monotonically with $r \to r_{LR}$, but instead it displays a peak, and the divergence is confined very close to the light-ring. A similar behavior is seen when considering the total flux $F/\mu^2$. For even larger spins, $q = 0.99999$, $F_{22}/\mu^2$ tends to decrease towards 0, and becomes linear in $(r - r_+)$ when $r_{LR} \sim r_+$. We note that analytical work on the gravitational radiation from a particle plunging into a nearly extremal Kerr BHs [3] predicted such behavior. However, we still expect the divergence to be present, although confined so close to the light-ring that we cannot resolve it.

Furthermore, during the plunge (which is governed mostly by conservative effects), the radial velocity $\dot{r}$ reaches maximum values that decrease with $q$, meaning that for large spins even the plunge is not too far from being circular. Figure 6 plots the $r$ dependence of the radial velocity in the region inside the ISCO for several different spin configurations, using orbital evolutions obtained by solving Hamilton’s equations. The peak radial velocity differs by more than 1 order of magnitude between $q = -0.99$ and $q = 0.99$.

V. QUASINORMAL-MODE MIXING IN RINGDOWN TEUKOLSKY WAVEFORMS AND ITS MODELING

The merger of a BH binary (of any mass ratio) eventually leads to the formation of a remnant Kerr BH of mass $M_f$ and dimensionless spin $q_f$. In this work, since we are dealing with an extreme mass-ratio system, we have $M_f = M$ and $q_f = q$. In the process of settling down to its final, stable state, the binary emits GWs. Those waves can be modeled as a linear superposition of QNMs [45,46] with complex frequencies $\sigma_{\ell mn}$, which depend only on $M_f$ and $q_f$, and are labeled by the spheroidal-harmonic indices $(\ell',m)$ and by an overtone index $n = 0, 1, \ldots$. For future convenience, we define $\omega_{\ell mn} \equiv \Re(\sigma_{\ell mn})$ and $\tau_{\ell mn} \equiv -1/3(\sigma_{\ell mn})$. We adopt the convention that $\omega_{\ell mn} > 0$ and $\tau_{\ell mn} > 0$ for any choice of the indices $(\ell', m, n)$.

In general, the strain waveform $h$ during the RD contains QNMs with all possible values of $(\ell', m, n)$. Additionally, given a spin $q$ and indices $(\ell', m)$, the angular differential equation which stems from the separation of the Teukolsky equation in spheroidal coordinates admits a pair of solutions characterized by frequencies $\sigma_{\ell \pm m}$. This implies that, whenever considering a specific component $(\ell', m)$ of $h$, even in principle, we get contributions from both positive- and negative-$m$ modes. As argued by Ref. [45], restricting to only positive-$m$ modes would enforce the assumption of circular polarization of the radiation. Of course, the actual importance of the modes depends on the details of how they are excited by the perturbing source, and by their decay times.

As already found by numerical investigations of the extreme and small mass-ratio limits [15–17,20,47,48], the dominant and leading subdominant ringdown Teukolsky modes can display a rich amplitude and frequency structure that hints at the interference of different QNMs besides the overtones of the least-damped mode, a phenomenon known as mode mixing. On the contrary, in the case of
comparable-mass BH binaries, mode mixing seems less ubiquitous, and so far it has only been seen in the (3,2) mode [49–55]. For this reason, in the past, when modeling the ringdown of the \((\ell', m)\) mode in the EOB approach, one could simply use the \((\ell', m, n)\) QNMs. However, the lack of mode mixing during ringdown in the comparable-mass case is inferred by the analysis of nonspinning, nonprecessing or mildly precessing configurations. We do not know yet whether this conclusion will hold when strongly precessing systems with mass ratios \(\geq 1/10\) will be considered.

QNM mixing manifests itself through striking features in the Teukolsky ringdown waveforms, which are modulated both in amplitude and frequency. To understand the composition of the QNM spectrum of the Teukolsky data, we will study in particular the GW frequency of each mode, defined as 

\[
\omega_{\text{Teuk}}(l, m) = -3 \left( \frac{h_{l}^{e}(\ell, m)}{h_{l}^{i}(\ell, m)} \right),
\]

since this quantity is directly related to the frequencies of the most excited QNMs, and is numerically well determined. As an example, Fig. 7 displays the ringdown (3,2) mode frequencies for several positive spins, with a common time axis rescaled by \(2\pi/\omega_{2320}\). We observe that different spins have completely different ringdown frequencies; each case has distinct features (spikes, oscillations), occurring with specific periodicities. The averages of the oscillatory features are closer either to \(\omega_{2320}\) (as one would naively expect) or to \(\omega_{2320}\), according to the value of \(q\). Examples of amplitude modulations can be found in Figs. 8, 9, and 10, which show a few Teukolsky merger-ringdown waveforms (solid blue lines), chosen within the large set that we computed for this paper. Among them, the most modulated case is spin \(-0.99\) [its (2,2) mode is shown in the right panel of Fig. 8; a similar behavior is also present in its higher-order modes].

In extreme and small mass-ratio binaries, two instances may enhance the excitation and/or mixing of modes other than the \((\ell', m, n)'s\) in the ringdown of \((\ell, m)\). On the one hand, for modeling purposes, the strain waveform \(h\) is typically decomposed onto \(-2\)-spin-weighted spherical harmonics \(-2Y_{\ell m}\), while the Teukolsky equation is separated using \(-2\)-spin-weighted spheroidal harmonics \(-2S_{\ell m}^{q}\), which depend on the Kerr spin \(q\) and the (possibly complex) frequency \(\omega\) of the gravitational perturbation. The expansion of the \(-2S_{\ell m}^{q}'s\) in terms of the \(-2Y_{\ell m}'s\) can be found [to order \((qM\omega)^{2}\)] in Appendix F of Ref. [56]. Using this result, one can derive a formula relating the spherical to the spheroidal waveforms (see, e.g., Eq. (19) of Ref. [26]) and one finds that the spherical mode \(h_{\ell m}\) receives contributions from all spheroidal modes with the same \(m\), but different \(\ell\) (see also Eq. (38) in Ref. [49]). Another source of mixing is the orbital motion of the perturbing particle: Whenever \(q < 0\), the orbital frequency switches sign during the plunge, because of frame dragging exerted by the spinning BH; this results in a significant excitation of modes with opposite \(m\), but with the same \(\ell\).

Reference [55] investigated in detail the origin of the mixing in the (3,2) mode of several comparable-mass, nonprecessing BH binaries, and attributed it mostly to angular-basis effects, using \(\omega = q\sqrt{M\sigma_{2320}}\).

To understand quantitatively the QNM mixing in our Teukolsky waveforms, we model the ringdown as done in
usual the mixing, starting at a time where $\frac{t}{r^*} \approx 0$, ringdown contains the modes $\ell = 2$ (2, 2, 0) besides the usual $(2, 2, n)$ $(n = 0, 1, \ldots)$. The vertical dashed lines mark $t = t^\text{match}$, note that the amplitudes have been rescaled by a factor of 5. $R$ is the distance to the source.

FIG. 8 (color online). Teukolsky $(2,2)$ mode waveforms for spin $q = -0.7$ (left panel) and $-0.99$ (right panel), displaying mode mixing during the ringdown phase. For $q = -0.7$ we plot a ringdown waveform which contains the mode $(2, -2, 0)$ besides the usual $(2, 2, n)$ $(n = 0, 1, \ldots)$. The vertical dashed lines mark $t = t^\text{match}$. Note that the amplitudes have been rescaled by a factor of 5.

![Waveform plots](waveform_plots.png)

FIG. 9 (color online). Teukolsky $(2,1)$ mode for spin $q = -0.8$ (left panel) and $(3,2)$ mode for spin $q = 0.9$ (right panel). For the $q = -0.8$ waveform, the modulations come from the interference of $(2,1,0)$ and $(2, -1, 0)$; note how the amplitude peak is affected by the mixing, starting at a time where $\Omega = 0$, i.e., the turning point of the azimuthal motion. For the $q = 0.9$ waveform, instead, the ringdown contains the modes $(3, 2, n)$’s and $(2, 2, 0)$. The vertical dashed lines mark $t = t^\text{match}$. $R$ is the distance to the source.

EOB models (i.e., as a linear superposition of overtones of the least-damped QNM), but with the addition of up to 2 further QNMs. While the least-damped mode and its overtones are going to account for the overall shape of the ringdown waveform, the additional QNMs are going to induce the modulations. More explicitly, [except for the $(3, 2)$ mode of systems with $q = 0.99$] we model the $(\ell', m)$ mode of the ringdown waveforms as

$$h^{\text{RD}}_{\ell' m} = \sum_{n=0}^{N-1} A_{\ell' m n} e^{-i \omega_{\text{eom}} (t - t^\text{match}_{\ell' m})} + S(t) [A_{\ell' 0 n} e^{-i \omega_{\text{eom}} (t - t^\text{match}_{\ell' 0})} + A_{\ell' \ell 0} e^{i \omega_{\text{eom}} (t - t^\text{match}_{\ell' \ell})}],$$

where $t^\text{match}_{\ell' m}$ is the time of merger, $N$ is the number of overtones included, the $A_{\ell' m n}$’s are the (constant) coefficients of the overtones, $S(t) \equiv [1 + \tanh [(t - t_s)/\tau_3]]/2$ is
FIG. 10 (color online). Teukolsky (3,2) waveforms for spin $q = -0.5$ (left panel) and $-0.95$ (right panel). For $q = -0.5$, the ringdown waveform contains the modes $(3, -2, 0)$ and $(2,2,0)$ besides the usual $(3,2,n)$'s. For the $q = -0.95$ waveform, the GW frequency modulations in the late ringdown are not centered neither about $-\omega_{3-20}$ nor about $-\omega_{2-20}$, and we cannot apply the simple model of Eq. (5). The vertical dashed lines mark $t = t_{\text{match}}^m$, $R$ is the distance to the source.

a factor introduced to have a smooth switch-on of the interfering QNMs (with $t_s$ and $r_s$ optimized mode by mode), and $A_{\ell m0}$ and $A_{\ell m0}$ are constants computed from a fit (see below). $A_{\ell m0}$ and $A_{\ell m0}$ quantitatively describe the strength of the QNM mixing. Note that $\sigma_{\ell m}(M_f,q_f) = \sigma_{\ell m}(M_f, -q_f)$. Since overtones with $n > 0$ have short decay times with respect to those with $n = 0$, Eq. (5) is actually dominated by terms with $n = 0$ when $t \gg t_{\text{match}}^m$.

The coefficients $A_{\ell m0}$, $A_{\ell m0}$, and $A_{\ell m0}$ can be determined from the Teukolsky data as follows. Whenever mode mixing is resolved, $A_{\ell m0}$ and $A_{\ell m0}$ are obtained by fitting the GW frequency $\omega_{\ell m}^{\text{RD}} = -3(h_{\ell m}^{\text{RD}}/h_{\ell m})$ to the ringdown Teukolsky GW frequency $\omega_{\ell m}^{\text{Teuk}}$, while setting $A_{\ell m0} = 0$ for $n > 0$; we choose a fitting window as wide as possible, but still avoiding any numerical noise. Once $A_{\ell m0}$ and $A_{\ell m0}$ are fixed by the fit, the $A_{\ell m0}$'s are calculated via the hybrid matching procedure detailed in Ref. [53], which consists in a smooth stitching of the ringdown waveforms $h_{\ell m}^{\text{RD}}$ to the Teukolsky waveform $h_{\ell m}^{\text{Teuk}}$ at a time $t_{\text{match}}^m$.

As in Ref. [20], we find that, in the test-particle limit, and when the spin is $q \leq 0$, some of the physical overtones included in Eq. (5) have frequencies smaller than $\omega_{\ell m}^{\text{Teuk}}(t_{\text{match}}^m)$, causing the slope of $\omega_{\ell m}^{\text{RD}}$ to be too steep. Therefore, we introduce a pseudo-QNM (i.e., a mode not belonging to the physical QNM spectrum). In the past, pseudo-QNMs were exploited in comparable-mass EOB models [23,29,53,57] to reduce the slope of the GW frequency in the transition from plunge to ringdown.

To summarize, the matching procedure has the following tuning parameters: the matching point $t_{\text{match}}^m$, the size of the time interval over which one carries out the matching $\Delta t_{\text{match}}^m$, a pseudo-QNM mode with frequency and decay time $\omega_{\ell m}^{\text{QNM}}$ and $\tau_{\text{match}}^m$, $t_s$ and $r_s$. These tuning parameters are chosen with the goal of minimizing the phase and relative amplitude difference between $h_{\ell m}^{\text{RD}}$ and $h_{\ell m}^{\text{Teuk}}$ when $t > t_{\text{match}}^m$.

Before modeling the entire ringdown waveforms, to better understand how the mixing works, let us consider the simple case of just 2 QNMs interfering: let $A_{\ell m0} = 0$ and $A_{\ell m0} = 0$ for $n > 0$ i.e., a waveform dominated by the $(\ell, \pm m, 0)$ modes. This is similar to what was done in Refs. [17,20], where the modulations in the ringdown frequency of the numerical modes were fitted with a simple analytical formula that accounted for the interference between the $(\ell, \pm m, 0)$ QNMs. The GW frequency is $\omega_{\ell m}^{\text{RD}} = -3(h_{\ell m}^{\text{RD}}/h_{\ell m})$, thus we have [leaving out the factor $S(t)$ for simplicity]

$$\omega_{\ell m}^{\text{RD}} = \omega_{\ell m} - \omega_{\ell m}^0 \frac{\Delta \alpha}{1 + |A|^2 e^{2(t-t_{\text{match}}^m)} \Delta \alpha} \cos (\tilde{\omega}(t - t_{\text{match}}^m) + \tilde{\theta}) - \Delta \alpha \sin (\tilde{\omega}(t - t_{\text{match}}^m) + \tilde{\theta})$$

where $\omega_{\ell m}^0 \equiv \omega_{\ell m0}$, $\alpha_{\ell m}^0 \equiv 1/\tau_{\pm m0}^0$, $\Delta \omega \equiv \alpha_{\ell m}^0 - \alpha_{\ell m0}$, $\Delta \alpha \equiv \alpha_{\ell m0} - \alpha_{\ell m0}$, $\tilde{\omega} \equiv \omega_{\ell m0} + \omega_{\ell m}^0$, and $A_{\ell m0}/A_{\ell m0} \equiv |A| \exp (i\tilde{\theta})$. Typically, $|A| < 1$. Note that Eq. (19) in Ref. [17] is simpler than our Eq. (6) above since that paper considered the
Schwarzschild case, for which $\sigma_{mn} = \sigma_{-mn}$. Equation (6) describes a function with exponentially growing oscillations about $\omega_+ \approx 0$ when $\rho_{\text{match}}^m < t < \rho_{\text{match}}^m - \log |A|/|\Delta \alpha| \approx t_p$, and with exponentially decreasing oscillations about $-\omega_+$ when $t > t_p$, the frequency of the oscillations is $\tilde{\omega}$. The point $t_p$ marks the transition from oscillations about $\omega_+$ to oscillations about $-\omega_+$; note that if $t_p - \rho_{\text{match}}^m \ll 1/\alpha$, then the transition occurs in a region where the amplitude is absolutely negligible. Given the size of the numerical errors discussed in Sec. III, we consider that the ringdown has ended whenever the amplitude drops below $10^{-3} \mu / R$, where $R$ is the distance to the source.

In the next two sections we shall discuss how we apply Eq. (5) to model the $\ell = m$ and $\ell \neq m$ numerical modes, respectively. The main conclusions can be summarized as follows. We are able to model the $\ell = m$ modes for any spin $q \geq -0.9$. The reason why we cannot model spins smaller than $-0.9$ is the conjectured presence of one or more QNMs that we are unable to recognize, which manifest themselves in a drift of the Teukolsky GW frequency at late times (see discussion in Sec. V A). We can model the (2,1) mode for any spin except $q = -0.95, -0.99$ because of large inaccuracies in capturing modulations in the early ringdown. Finally, we can model the (3,2) mode for any spin except $q = -0.9, -0.95, -0.99$ because of these very negative spins the oscillations in the amplitude and GW frequency become very dramatic, preventing us from reliably fitting the amplitude and phase of all the modes (see the right panel of Fig. 10).

Note that we use a unique tuning for the pseudo-QNM, namely $\omega_{\rho_{\text{match}}^m} = [\omega_{\rho_{\text{match}}^m} + \omega_{\text{Teuk}}(\rho_{\text{match}}^m)]/2$ and $\rho_{\text{match}}^m = 0.2 \rho_{\text{match}}^m$. When no mode mixing is present, the pseudo-QNM replaces the $8$th physical overtone ($\ell, m, 7$), otherwise it is added to the rest of the mode spectrum. For all the spins that we have been able to model, the matching intervals $\Delta \rho_{\text{match}}^m$ are listed in Table III.

### A. \(\ell = m\) modes

For all modes with $\ell = m$, we choose $\rho_{\text{match}}^m = \rho_{\text{peak}}$, which is the time when the orbital frequency $\Omega$ peaks, very close to the light-ring; this choice has the advantage of avoiding the ambiguity of locating the amplitude peak when $\rho \sim 1$ (see Sec. IV). For the (2,2) and (3,3) modes we choose $\rho_{\text{match}} = \rho_{\text{match}}^m + 20M$ and $\tau_s = 7.5M$; when $(\ell, m) = (4, 4)$, we choose instead $\rho_{\text{match}} = \rho_{\text{match}}^m + 25M$ and $\tau_s = 4.5M$.

Let us first consider the dominant (2,2) mode. For spins $q \geq 0.5$, we find that the ringdown is quite standard, as no appreciable mode mixing is present, and $h_{\text{Teuk}}^{22}$ is well described by a linear superposition of overtones of the least-damped mode [i.e., Eq. (5) with $A_{\ell' = 20, A_{\ell - 20} = 0}$]. The matching interval $\Delta \rho_{\text{match}}^{22}$ varies with $q$ as prescribed in Table III. We find that $\Delta \rho_{\text{match}}^{22}$ tends to grow towards large positive spins since the light-ring (i.e., the matching point) occurs progressively later, during the ringdown, well past the amplitude peak, in a region where the waveform is rapidly decaying. For spins $q \leq 0$, we find it necessary to include the $(2, -2, 0)$ mode in the QNM spectrum [i.e., Eq. (5) with $A_{\ell' = 20} = 0$]; this mode has an amplitude $|A_{2, 20}^{20}|$ that grows (relative to $|A_{220}^{220}|$) as the spin decreases, which can be understood based on the fact that the portion of orbit with $\Omega < 0$ (due to frame dragging during the plunge) becomes progressively longer. In Table I we provide magnitude and phase of $A_{220}$, i.e., the ratio of $(2, -2, 0)$ relative to the least-damped QNM. The numbers in the table are obtained from a fit of $\omega_{\text{Teuk}}^{22}$ using Eq. (6).

The typical performance of the model is illustrated in the left panel of Fig. 8, which shows the case with spin $-0.7$. The Teukolsky amplitude (frequency) is plotted in blue (cyan), while the model amplitude (frequency) is plotted in red (orange). We clearly recognize the growing oscillations of the GW frequency about $M \Omega_{220} \approx 0.31$; by fitting, we find that $|A_{2, 20}| / |A_{220}| \approx 0.12$, so that $t_p \approx \rho_{\text{match}}^m + 270M$ (i.e., in a region where $|R h_{\text{Teuk}}^{22} / \mu | \ll 10^{-4}$). The waveform $R h_{\text{Teuk}}^{22}$ does a good job at capturing the modulations everywhere, except in the early ringdown ($\rho_{\text{match}}^m < t \lesssim t_s$), where the oscillations in $\omega_{\text{Teuk}}^{22}$ occur at a frequency $\tilde{\omega} \neq \omega_{220} + \omega_{2, 20}$, and with an amplitude growth whose timescale does not clearly relate to either $\tau_{220}$ or $\tau_{2, 20}$ as one would expect from Eq. (6). One limitation inherent to our approach is the specific form $S(t)$ of the time dependence of the coefficients $A_{\ell' = 0}^{m}$ and $A_{\ell = 0}$, which may not correctly model the process of excitation (in spite of the two adjustable parameters $t_s$ and $\tau_s$). Note that in

### Table I

Relative amplitude and phase of the QNMs responsible for mixing in modes with $\ell = m$. No QNM mixing is present when the spins are positive. Spins $q = -0.95, -0.99$ cannot be modeled with Eq. (6) due to the presence of additional interfering QNMs that we are unable to extract, which results in a GW frequency drift at late times (see, for instance, the inset in the right panel of Fig. 8).

| $a/M$ | $|A_{2, 20}^{20} / A_{220}^{220}|$ | $\arg(A_{2, 20}^{20} / A_{220}^{220})$ | $|A_{3, 30}^{30} / A_{330}^{330}|$ | $\arg(A_{3, 30}^{30} / A_{330}^{330})$ | $|A_{4, 40}^{40} / A_{440}^{440}|$ | $\arg(A_{4, 40}^{40} / A_{440}^{440})$ | $|A_{5, 50}^{50} / A_{550}^{550}|$ | $\arg(A_{5, 50}^{50} / A_{550}^{550})$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.0036 | −5.70 | 0.0029 | −4.07 | 0.0035 | −5.80 | 0.0047 | −7.58 |
| −0.5 | 0.052 | −0.24 | 0.049 | −1.26 | 0.056 | 3.82 | 0.073 | 2.82 |
| −0.7 | 0.12 | −0.26 | 0.22 | 4.31 | 0.29 | 2.86 | 0.39 | 1.60 |
| −0.8 | 0.10 | −0.41 | 0.22 | 3.03 | 0.31 | 1.02 | 0.38 | −0.81 |
| −0.9 | 0.28 | −1.31 | 0.27 | 1.41 | 0.32 | −1.76 | 0.46 | −4.55 |
comparable-mass EOB models the coefficients in front of the QNMs in \( h^{RD}_{\ell m} \) have no time dependence.

For spins \( q < -0.8 \), the point \( t_p \) moves closer to \( t_{\text{match}}^{\ell m} \), and the performance of the model in the early ringdown (i.e., \( t_{\text{match}}^{\ell m} < t < t_p \)) becomes worse; however, Eq. (5) with \( A_{r,20} = 0 \) can still describe the region \( t > t_p \) quite accurately.

We find that the most difficult ringdown waveforms to model are the ones with spin \(-0.95 \) and \(-0.99 \); the case \( q = -0.99 \) is shown in the right panel of Fig. 8. These are the cases with the longest inversion of the trajectory due to frame dragging. We have verified that the numerical errors during the ringdown are not responsible for creating any of the modulations. Note that for such extreme (negative) spins we have \( \Delta \Omega \approx 10^{-3} - 10^{-2} \).

We observe that the amplitude oscillations begin at the turning point of the particle’s azimuthal motion (i.e., when \( \Omega \) vanishes); thus, we choose this as our \( \tau_{\text{match}}^{\ell m} \), which we are unable to identify; their presence is hinted by the (physical) drift in the GW frequency at late times. This is exemplified in the right panel of Fig. 8, which refers to the \((2,2)\) mode of spin \(-0.99 \). The inset therein zooms into the late ringdown, past the point where transitions to oscillations about a negative frequency. One can see that the average of the oscillations is not \(-\omega_{2-20}\), but instead it slowly asymptotes to that value from above.

As to the other modes with \( \ell = m \), they behave similarly to the \((2,2)\) mode, namely for spins \( q > 0 \) no significant mode mixing is present, while for \( q \leq 0 \) the mode \((\ell, -m, 0)\) is excited. In Table I we list the extracted coefficients (relative to the coefficient of the dominant QNM) of those QNMs that cause amplitude and frequency modulations. Again, the simple ringdown model of Eq. (5) with \( A_{r,00} = 0 \) fails to accurately describe the early ringdown for spins \( q < -0.8 \), so that we cannot model \( q = -0.95, -0.99 \).

We have also tried to look for contributions from the horizon modes suggested by Refs. [11,12], whose frequency is \( m\Omega_{H} \), but their decay time \( r_s / (2\sqrt{1 - q^2}) \) is not compatible with any of the timescales present in the Teukolsky data, and we did not observe their presence in the numerical waveforms.

### Table II

<table>
<thead>
<tr>
<th>( a/M )</th>
<th>( A_{2,10}/A_{2,10} )</th>
<th>( A_{2,20}/A_{3,20} )</th>
<th>( A_{3,20}/A_{3,20} )</th>
<th>( \text{arg}(A_{2,10}/A_{2,10}) )</th>
<th>( \text{arg}(A_{2,20}/A_{3,20}) )</th>
<th>( \text{arg}(A_{3,20}/A_{3,20}) )</th>
<th>( \text{arg}(A_{3,20}/A_{3,20}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>0.038</td>
<td>13.7</td>
<td>13.7</td>
<td>only ((2,2, n)) overtones</td>
<td>0.99</td>
<td>0.038</td>
<td>13.7</td>
</tr>
<tr>
<td>0.95</td>
<td>0.030</td>
<td>9.82</td>
<td>2.91</td>
<td>6.44</td>
<td>0.95</td>
<td>0.030</td>
<td>9.82</td>
</tr>
<tr>
<td>0.9</td>
<td>0.0025</td>
<td>5.61</td>
<td>2.11</td>
<td>5.99</td>
<td>0.9</td>
<td>0.0025</td>
<td>5.61</td>
</tr>
<tr>
<td>0.8</td>
<td>0.0024</td>
<td>4.24</td>
<td>1.13</td>
<td>2.86</td>
<td>0.8</td>
<td>0.0024</td>
<td>4.24</td>
</tr>
<tr>
<td>0.7</td>
<td>0.0051</td>
<td>3.73</td>
<td>0.71</td>
<td>2.10</td>
<td>0.7</td>
<td>0.0051</td>
<td>3.73</td>
</tr>
<tr>
<td>0.5</td>
<td>0.010</td>
<td>-0.16</td>
<td>0.34</td>
<td>-7.02</td>
<td>0.5</td>
<td>0.010</td>
<td>-0.16</td>
</tr>
<tr>
<td>0</td>
<td>0.069</td>
<td>-3.46</td>
<td>0.010</td>
<td>1.08</td>
<td>0</td>
<td>0.069</td>
<td>-3.46</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.21</td>
<td>-5.64</td>
<td>0.13</td>
<td>-2.46</td>
<td>0.093</td>
<td>5.26</td>
<td></td>
</tr>
<tr>
<td>-0.7</td>
<td>0.26</td>
<td>-6.58</td>
<td>0.15</td>
<td>3.47</td>
<td>0.22</td>
<td>10.6</td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>0.30</td>
<td>-7.37</td>
<td>0.16</td>
<td>3.91</td>
<td>0.23</td>
<td>12.4</td>
<td></td>
</tr>
<tr>
<td>-0.9</td>
<td>0.32</td>
<td>-8.30</td>
<td>0.16</td>
<td>3.91</td>
<td>0.23</td>
<td>12.4</td>
<td></td>
</tr>
</tbody>
</table>
mixing, as can be seen in the left panel of Fig. 9, mainly due to $S(t)$. In spite of the different ringdown prescriptions used in the positive versus negative spin regime, we can see from Table II that $|A_{20}/A_{220}|$ and arg $(A_{20}/A_{220})$ are well-behaved functions of spin. Spins $-0.95$ and $-0.99$ cannot be modeled accurately, the issue being the early ringdown (i.e., $t_{\text{match}}^1 < t < t_p$), whose modulations become rather extreme, and are not captured by $S(t)$. The late ringdown (i.e., $t > t_p$) follows instead well our model.

The more challenging mode to model is the $(3,2)$. For $q > 0$ we use $t_{\text{match}}^m = t_\text{peak}^\Omega$, $t_s = t_{\text{match}}^2 + 2.5M$, and $\tau_s = 10M$. As already seen in Fig. 7, when $q \gtrsim 0.7$ the QNM mixing induces a transition of the average (final) ringdown frequency from the expected least-damped mode frequency $\omega_{320}$ to $\omega_{220}$. Note how the case $q = 0.7$ (fifth panel of Fig. 7) sits at the transition between the two regimes, featuring wide frequency oscillations around both $\omega_{320}$ and $\omega_{220}$. The case $q = 0.99$ stands out, since its ringdown can be described by the $(2,2)$-mode spectrum [i.e., $h_{32}^{2D} = \sum_n b_{22n} e^{-i\sigma_{22n}(t - t_{\text{match}}^m)}$ is a good model for $h_{32}^{\text{Teuk}}$]. This happens because there are no significant mode-mixing modulations (see the first panel of Fig. 7) and the asymptotic GW frequency is $\omega_{32}^{\text{Teuk}}(t \to \infty) = \omega_{220}$. In the range $0 < q \lesssim 0.95$, instead, we model the ringdown via Eq. (5), setting $A_{32-20} = 0$ and $\ell'' = 2$, i.e., the QNM spectrum is that of the $(3,2)$ mode with interference from $(2,2,0)$. When $0.8 \lesssim q \lesssim 0.95$, the ringdown displays large features, with a GW frequency oscillating about $\omega_{220}$ (see second to fourth panel of Fig. 7); this means that $(2,2,0)$ is more excited than the least-damped mode $(3,2,0)$, which is confirmed by our fits, as $|A_{220}/A_{220}| > 1$ (see Table II). The right panel of Fig. 9 shows the good agreement of the model to the Teukolsky data for $q = 0.9$. Notice how the matching point lies in a region where the amplitude has already started to drop, quite a bit later than the peak. As already discussed, the case with $q = 0.7$ represents a sort of threshold, in that its GW frequency oscillates about $\omega_{320}$ in the early ringdown and then about $\omega_{220}$ in the late ringdown (see the fifth panel of Fig. 7).

Similarly to the $(2,1)$ mode, when $q \leq 0$, $t_{\text{peak}}^\Omega$ occurs quite early; when $q < 0$, the $(3,2)$ amplitude peak is modulated by the mode mixing, but now the turning point of the particle happens somewhat earlier relative to it. Therefore, when $q \leq 0$, we choose the matching point in the “middle” of the amplitude peak, where $\partial^2 t_{\text{match}} h_{32}^{\text{Teuk}} = 0$; we also choose $t_s = t_{\text{match}}^2 + 10M$ and $\tau_s = 7.5M$. In terms of QNM spectrum, as happens for all the modes we studied, for $q \leq 0$ the mode with opposite $m$ is excited [i.e., $(3, -2, 0)$]. However, the $(2,2,0)$ mode can still be extracted from the $(3,2)$ waveforms for spins as small as $-0.8 \lesssim q < 0$. Here we use Eq. (5) with $A_{3-20}, A_{220} \neq 0$. The function that we fit to the Teukolsky data is simply the generalization of Eq. (6) to three interfering QNMs. The extracted coefficients are found in Table II. An example of this regime is shown in the left panel of Fig. 10, for spin $-0.7$; one can notice two effects in $h_{32}^{\text{Teuk}}$, the high-frequency modulations due to the interference of $(3, -2, 0)$, and the low-frequency ones due to the interference of $(2,2,0)$. It is also possible to appreciate how well the model (red and orange lines) can capture all these features. For spins $q \lesssim -0.9$, the waveforms asymptote to a frequency lying between $-\omega_{220}$ and $-\omega_{320}$, and we cannot extract the coefficients because $h_{32}^{\text{Teuk}}$ has a very irregular behavior and we find it hard to determine the appropriate fitting window. This problematic regime is depicted in the right panel of Fig. 10, where we plot the $(3,2)$ mode of spin $-0.95$.

VI. CONSIDERATIONS ON THE MODELING OF COMPARABLE-MASS BINARY SYSTEMS

In this section we explain how the findings of Secs. IV and V can help EOB waveform modeling.

The EOB approach employs factorized analytical (multiorbital) waveforms that resum the circular PN formulas, while incorporating strong-field and noncircular effects [16,24–26,58]. An example of strong-field feature is the divergence of the factorized modes at the light-ring for circular orbits through the “source” term proportional to the binding energy (angular momentum) for $\ell = m (\ell \neq m)$ modes. Deviations from circularity are modeled in the EOB waveforms through a phenomenological non-quasicircular (NQC) factor that reshapes the EOB factorized waveforms during plunge and around merger in order to better match the numerical waveforms (computed either with numerical relativity or Teukolsky-equation codes).

---

5 Reference [24] (see Appendices C and D) computed mode-by-mode amplitude fits of the Teukolsky modes generated by a frequency-domain code, which assumed circular orbits in Kerr. High, unknown PN terms in the factorized waveforms were fitted up to the ISCO, for both ingoing and outgoing radiation.
The NQC factor is determined once the numerical “input values” (i.e., the amplitude $|h_{num}^{jm}|$, the slope $\partial_j|h_{num}^{jm}|$, the curvature $\partial^2_j|h_{num}^{jm}|$, the frequency $\omega_{num}^{jm}$, the slope of the frequency $\partial_j\omega_{num}^{jm}$) are prescribed. Typically, the input values are read off at the peak of the numerical waveforms, and, on the EOB side, they are enforced at a specific time relative to the peak of the orbital frequency (which occurs at $t = \Omega_{peak}^0$). That same time is used as the attachment point for the ringdown waveform (for more details, see Sec. IV of Ref. [20]).

As discussed in Sec. IV, the very circular character of the Teukolsky waveforms when $q \to 1$ is very appealing from the point of view of the modeling, since the EOB factorized modes (without NQC corrections) are built under the assumption of quasicircular adiabatic motion. However, highly spinning systems are also very relativistic, and current PN waveforms (on which the factorized ones are based) are not accurate enough for such regimes already hundreds of cycles before merger. As already pointed out in Refs. [20,26] by comparisons with frequency-domain Teukolsky waveforms, due to the lack of enough PN knowledge in the test-particle limit, the amplitude of the factorized waveforms performs poorly even before the ISCO for large spins, implying also inaccurate multipolar fluxes. While we were finalizing this paper, Ref. [59] was posted; the author computed the energy fluxes for a particle in circular, equatorial orbit in Kerr spacetime up to 20PN order. In spite of the high PN order of the calculation, the relative accuracy of the analytical flux (when compared to numerical Teukolsky data) is within $10^{-3}$ only down to 2.97rISCO for spin 0.9, i.e., for an orbital speed around 0.37 (to be compared with $r_{ISCO} \approx 0.61$). As we shall see in Sec. VII, even modeling errors as small as $10^{-3}$ at the ISCO may result in large dephasings once the analytical fluxes are employed in time evolutions.

Moreover, as originally found in Ref. [20], the larger the spin, the earlier the (2,2) mode peaks with respect to $\Omega_{peak}^0$: When $q \geq 0.9$, the peak occurs before the ISCO, during the inspiral phase, where the radial motion is absolutely negligible, as we discussed in Sec. IV. As a consequence, when calculated at the amplitude peak, the NQC functions are heavily suppressed for large and positive spins, because they are proportional to $p_r \propto \dot{r}$ (see Fig. 6), and cannot help correcting the waveform. One could see what can be gained by applying the factorized resummation procedure to the PN-expanded fluxes of Ref. [59], or keep the current factorized flux while including the fits of Ref. [24], and obtain EOB amplitudes in greater agreement with the numerical ones without any need for NQC corrections. Note that the fits of Ref. [24] were computed up to the ISCO. Hence, after the peak, when the amplitude is falling off, the EOB waveform with fits can still differ from the Teukolsky one. However, applying an NQC correction at that late stage could be a viable option.

Furthermore, if we followed the standard EOB prescription of attaching the ringdown waveform at $t_{peak}^m$, we would not be able to successfully model the Teukolsky waveform, because its ringdown sets in at times which are rather close to the peak of the orbital frequency at time $\Omega_{peak}^0$, while $t_{peak}^m \ll t_{peak}^0$. As we shall see below, to overcome this issue, we suggest a new prescription for the matching point of the ringdown in the EOB approach for small mass ratios and large spins.

These findings for large spins were effectively exploited in the construction of the EOB model of Ref. [29], which extended the model of Ref. [22] to generic mass ratios and spins; only the dominant (2,2) mode was considered. The model was calibrated to 38 numerical-relativity nonprecessing waveforms produced by the SXS Collaboration [60–63], spanning mass ratios from 1 to 8, spin magnitudes up to 0.98, and with 40 to 60 GW cycles. By construction, any EOB model incorporates the test-particle limit, since the whole formalism is based on a deformation of the Kerr spacetime. As explained above, the merger waveform critically depends on the information from numerical-relativity waveforms, in the form of input values. Since numerical-relativity simulations are still unable to explore the small mass-ratio limit, the Teukolsky waveforms are extremely valuable in bridging the gap between mass ratio $\sim 1/10$ and $\sim 1/1000$.

The prototype nonprecessing, spinning EOB model of Ref. [22] (which could cover spins only up to 0.6) introduced, for the first time, a spin-dependent (negative) time delay $\Delta t_{22}^{peak}$ between $\Omega_{peak}^0$ and the peak of $|h_{22}|$, which was inspired by the time delay seen in the Teukolsky data of Ref. [20]. Such time delay had already been found in Ref. [18] for the (2,2) mode in nonspinning binaries with small mass ratio, but because the time delay in the nonspinning case is quite small, it was not needed when modeling the (2,2) mode of nonspinning, comparable-mass systems [53]. Furthermore, Ref. [22] fixed the small mass-ratio limit of $\omega_{22}^{num}$ and $\partial_j\omega_{22}^{num}$ based on the Teukolsky waveforms of Ref. [20]. In the same spirit of Ref. [22], some of us used the additional information on the test-particle limit provided in this paper (in particular, the behavior of the Teukolsky waveforms beyond spin 0.8) to extend the nonprecessing EOB model to any spin and mass ratio [29].

First, we built a time-delay function $\Delta t_{22}^{peak}$ that, in the small mass-ratio limit, decreases with spin beyond 0.8; this guarantees that the ringdown starts close to $\Omega_{peak}^0$ and that the NQC equations are always enforced in a region with significant radial motion (at time $\Omega_{peak}^0 + \Delta t_{22}^{peak}$), as opposed to the extremely circular region around the

\[ m_1 m_2 / (m_1 + m_2)^2, \ m_1 \text{ and } m_2 \text{ being the BH masses.} \]
amplitude peak (at time $t^2_{\text{peak}}$). As an example, in Fig. 2 we indicate with a vertical green line where the point $t^2_{\text{peak}} + \Delta t^2_{\text{peak}}$ occurs for such time-delay function when the mass ratio is 1/1000 and the spin is 0.99: the point safely lies well after the ISCO, close to the light-ring. Remember that the analysis of Sec. VA has shown that for (2,2) modes and large spin one can reliably attach the ringdown waveform at the light-ring.

Second, we built piecewise continuous fitting functions for the input values along the spin dimension\(^8\) such that, beyond spin 0.8 and for mass ratio smaller than ~1/100, they approach spin 0.8 and for mass ratio 1/1000, the EOB model will not agree too well with the Teukolsky waveforms produced with the numerical flux, as in this paper. This is mainly a consequence of the limitation of the current factorized waveforms that we discussed above (especially as far as the amplitude is concerned). Imposing the exact Teukolsky input values beyond spin 0.8 at mass ratio 1/1000 would result in unwanted features (such as bumps in the inspiral amplitude), because the NQC corrections act only over short time intervals, while the factorized waveforms are discrepant over much longer corrections act only over short time intervals, while the bumps in the inspiral amplitude), because the NQC correction), evaluated at time $t^2_{\text{peak}} + \Delta t^2_{\text{peak}}$. This entails that, beyond spin 0.8 and for mass ratio 1/1000, the EOB factorized waveform itself (without any spinning NQC correction), evaluated at time $t^2_{\text{peak}} + \Delta t^2_{\text{peak}}$. This data can be used for future, improved versions of the EOB model.

Third, in the model of Ref. [29], the Teukolsky waveforms were also exploited to establish robust ringdown prescriptions in the small mass-ratio limit, especially for binaries with large spins. Indeed, we found it necessary to introduce mass-ratio and spin dependence in the ringdown tuning parameters (i.e., the size of the matching interval, frequency and decay time of the pseudo-QNMs).

Finally, in Appendix A we provide input values measured from the Teukolsky waveforms of this paper, as well as the measured time delay $\Delta t^2_{\text{peak}}$ as functions of the spin. This data can be used for future, improved versions of the EOB model.

### VII. THE COMPARABLE-MASS EFFECTIVE-ONE-BODY MODEL IN THE TEST-PARTICLE LIMIT

In this section we compare the comparable-mass EOB model of Ref. [22] to the numerical waveforms computed via the Teukolsky formalism in the test-particle limit. Before discussing the waveforms, we have to point out that the orbital dynamics generated by the EOB model in this section is quite different from that generated following the prescriptions of Sec. II. In fact, as already discussed, the EOB energy flux used in Ref. [22], which was based on Refs. [25,26] and used all the PN corrections available at the time of publication, has several shortcomings in the test-particle limit.

First, the EOB energy flux used in Ref. [22] does not account for the ingoing portion of the GW flux. Horizon absorption has the largest effect for nearly extremal positive spins, thanks to the slower rate of energy loss, due to superradiance. Note that the relative sign between ingoing and outgoing fluxes changes when the orbital frequency crosses the horizon frequency. When $\Omega \leq \Omega^\ast_H$ and $q > 0$, the ingoing flux fraction subtracts from the outgoing flux; otherwise, the absorption flux adds to the outgoing flux. For instance, when the spin is 0 (0.99), the absorption flux increases (decreases) dissipation by ~0.3% (~9%) for a particle orbiting at the ISCO (see Fig. 1 of Ref. [24]). References [27,65] found that in the nearly extremal case $q = 0.99$ the inspiral up to the ISCO can be longer by ~5% at low inclinations, depending on whether the ingoing flux is included or not. A study extending up to merger was done in the Schwarzschild case by Ref. [21], which considered an EOB evolution including the model absorption flux of Ref. [58]; when the symmetric mass ratio is $10^{-3}$, they found a dephasing of 1.6 rads for the (2,2) mode waveform at merger over an entire evolution of about 41 orbital cycles. As to the spinning case, Ref. [66] included the spinning horizon flux in an EOB model, using the Taylor-expanded expressions of Refs. [67,68]; the inclusion of absorption turned out to be important to obtain good agreement with the full Teukolsky flux, at least up to the ISCO. When modeling spinning binaries, one should bear in mind that the spin changes the PN order (with respect to the leading order flux at infinity) at which absorption enters in the energy flux: while this effect enters at 4PN order for Schwarzschild BHs, it enters at 2.5PN order for non-zero spin.

To confirm the impact of neglecting the ingoing flux, we evolve trajectories with either the total or only the outgoing Teukolsky flux, relying again on the data of Ref. [24]. We consider (2,2) waveforms that begin 100 GW cycles before the ISCO. For comparison, we align their phases both at low frequency (over the first 10 GW cycles) and at high frequency\(^9\) (over the 10 GW cycles following the ISCO), and then measure the phase difference either during ringdown (for the low frequency alignment) or at the beginning of the waveform (for the high-frequency alignment), using the case with the total flux as fiducial. After the low frequency alignment, we find that for spin 0 (0.99) the horizon absorption induces a dephasing of about $-2$ (+23) rads. After the high-frequency alignment, we find that for spin 0 (0.99) the

\(^8\)Note that, in both EOB models of Refs. [22] and [29], the input values are functions of only two parameters: the symmetric mass ratio and an effective spin [see definition in Eq. (32) of Ref. [22]].

\(^9\)Note that aligning the waveforms at the amplitude peak is not an option, given their extreme flatness when $q = 0.99$. 
horizon absorption induces a dephasing of about $-0.1$ ($+8$) rads. The different sign in the dephasings for spin 0 and 0.99 reflects the fact that for $q \leq 0$ the ingoing flux increases the rate of dissipation (thus hastening the coalescence), while for $q > 0$ super-radiance extracts energy from the rotation of the massive BH and transfers it into the orbital motion (thus delaying the coalescence). These effects can play a major role for space-based GW detectors, whose integration time will have to be of the order of $10^6$ GW cycles (or more) to achieve detection [69], hence requiring very long and accurate GW templates.

In principle, horizon absorption may also alter the merger waveform, which constitutes a numerical input for the EOB model via the NQC procedure outlined in Sec. VI. For $q = 0$ we compute the (2,2) mode input values $\frac{\partial}{\partial q} [H_{22}^{\text{Teuk}}]$, $\partial_l [H_{22}^{\text{Teuk}}]$, $\partial_{\text{sym}} [H_{22}^{\text{Teuk}}]$, $\partial_0 [H_{22}^{\text{Teuk}}]$ at $t_{\text{ISCO}}$ (here, of course, $\partial_j [H_{22}^{\text{Teuk}}] = 0$), while for $q = 0.99$, due to the flatness of the amplitude and the lack of an orbital frequency peak, we compute them at the ISCO.\(^\text{10}\) For spin 0 (0.99), the relative difference induced by horizon absorption on the four input variables is respectively: $0.0014\%$ (0.17\%), $0.50\%$ (5.8\%), $0.082\%$ (0.29\%), and $0.091\%$ (5.2\%). Similar results apply to higher-order modes. The larger discrepancies can be seen on the curvature and on the slope of the GW frequency, but the NQC procedure is only mildly sensitive to these two quantities, as the most important features to reproduce are the amplitude and the GW frequency, which means that the horizon absorption does not impact the merger waveform significantly.

Second, as compared to the total outgoing Teukolsky flux, the current EOB energy flux does not account for modes with $\ell > 8$. We can quantitatively assess this truncation error in the frequency domain by using the multipolar components of the Teukolsky fluxes computed in Ref. [24]. We find that, for a particle orbiting at the ISCO, the fractional contribution to the total outgoing flux coming from modes beyond $\ell = 8$ varies between $10^{-3}$ for $q = -0.99$ and $3 \times 10^{-3}$ for $q = 0.99$. The growing relevance of higher modes with spin is consistent with the trend that one sees when studying the amplitude hierarchy between the dominant (2,2) mode and higher modes [20]. For spins $q = 0, 0.99$, we compute the Teukolsky waveforms along trajectories sourced by Teukolsky flux modes only up to $\ell = 8$, and compare them to the waveforms generated using the total outgoing flux (taken as fiducial). We measure the dephasings with the same approach discussed above when studying the effect of horizon absorption. After the low frequency alignment, we find that for spin 0 (0.99) the higher-$\ell$ modes induce a dephasing of about $-0.3$ ($-7.5$) rads. After the high-frequency alignment, we find that for spin 0 (0.99) the higher-$\ell$ modes induce a dephasing of about $-0.015$ ($-3$) rads. The negative signs indicate that, obviously, whenever we neglect $\ell > 8$ modes the rate of dissipation is lower, hence the coalescence occurs later. These phase differences are less dramatic than those seen when neglecting the ingoing flux. Nonetheless, they are relevant for the purpose of generating templates for extreme and small mass-ratio inspirals.

Third, as discussed in Sec. VI, the modeling error on the amplitude of the individual factorized modes with $\ell \leq 8$ can be significant even before the ISCO for large spins: a more quantitative assessment of the disagreement with numerical amplitudes can be found in Ref. [26]. The origin of the poor performance lies in the limited PN knowledge, since for large spins the ISCO moves to a more relativistic regime: $v_{\text{ISCO}} \approx 0.41$ when $q = 0$, while $v_{\text{ISCO}} \approx 0.79$ when $q = 1$. Again, one could include the amplitude fits of Ref. [24] or apply the factorized resummation to the analytical energy flux of Ref. [59], and recalibrate the comparable-mass model to numerical-relativity simulations.

We now move on to discuss the waveforms. We evaluate the comparable-mass EOB model of Ref. [22] in the test-particle limit by setting the symmetric mass ratio $\mu/M$ to zero everywhere in the model, except in the leading term of the GW flux, where we set it to $10^{-3}$; this choice is consistent with the prescriptions of Sec. II for building orbital evolutions with the Teukolsky fluxes. The GW flux of the model is a sum of time derivatives of multipolar modes up to $\ell = 8$, according to Eq. (13) of Ref. [22]. All the modes are the $p$-resummed factorized ones of Ref. [26], except those with $\ell \leq 4$ and odd $m$, which instead follow the prescription given in Appendix A of Ref. [22]; test-particle limit nonspinning effects are included up to 5.5PN order (beyond the leading order), while spinning effects are included up to 4PN order (beyond the leading order). Here we are not interested in testing the EOB orbital dynamics, but we rather want to focus on the waveforms, therefore the Teukolsky waveforms are calculated along the EOB trajectories. The same approach was adopted in Ref. [20] for the case with spin 0. For spins as large as $q \sim 0.5$, the EOB waveforms are in good agreement with the numerical waveforms. In Fig. 11, for $q = 0.5$, we align EOB and Teukolsky (2,2) mode waveforms at the amplitude peak; we find a dephasing within 0.1 rads and a relative amplitude error which is negligible everywhere except during ringdown (where it is around 30\%). For larger spins, however, a large discrepancy in the amplitude shows up well before merger. In Fig. 12, for $q = 0.8$, we find an amplitude error around 5% during the late

\(^{10}\)For $q = 0.99$, the ISCO is only $0.3M$ away from the horizon in the radial coordinate. See also Fig. 2 for a more precise idea in the case of the (2,2) mode.
inspiral; the dephasing is quite large too, reaching about 0.8 rads 50 GW cycles before merger, and growing as one moves to lower frequencies.

VIII. CONCLUSIONS

Using the Teukolsky equation in the time domain, we have computed inspiral-merger-ringdown waveforms produced by the inspiraling motion of a nonspinning test particle in the equatorial plane of a Kerr BH with dimensionless spin $-0.99 \leq q \leq 0.99$, thus extending work done in Ref. [20]. The trajectory of the particle has been obtained from the geodesic equation, subject to a radiation-reaction force that is proportional to the total energy flux in GWs. We have used the GW fluxes computed for circular orbits down to the light-ring with a frequency-domain Teukolsky code [24]. We have computed the dominant and leading subdominant modes of the radiation: (2,2), (2,1), (3,3), (3,2), (4,4), and (5,5).
In Sec. IV, we have pointed out the simplicity of the waveforms emitted by systems with large, positive spins, in spite of the highly relativistic regime probed by the inspiraling orbital trajectories. The main feature of the mode amplitudes is their flattening towards the ISCO and during the plunge as the spin grows (see Fig. 3). We have given an explanation of this phenomenon in terms of the ratio between the orbital and the radiation-reaction timescales. On the one hand, as \( q \rightarrow 1 \) the total (i.e., ingoing + outgoing) GW flux tends to decrease, partly thanks to the extraction of energy from the rotation of the Kerr BH via super-radiance. On the other hand, as \( q \rightarrow 1 \) the horizon (i.e., the final point of the orbital evolution) moves to smaller radii, which implies higher orbital frequencies accessible to the inspiraling particle. This results in a significant increase in the number of orbits per unit frequency as \( q \rightarrow 1 \); the orbital motion becomes extremely circular, and highly relativistic.

In Sec. V, we have systematically studied the ringdown stage, whose waveforms display complicated amplitude and frequency modulations due to the interference of QNMs. In the comparable-mass range, with the notable exception of the (3,2) mode, the \((\ell, m)\) modes of nonprecessing BH binaries can be successfully modeled by the linear superposition of overtones of the least-damped QNM, i.e., \((\ell, m, n)\), with \( n = 0, 1, \ldots \) [14,49–55]. However, in the extreme and small mass-ratio regime, other QNMs can be excited [15–17,20,47,48]. We have found that, for \( \ell = m \) modes, the QNM mixing is present when \( q \lesssim 0 \) (see Fig. 8), and arises mainly due to modes with opposite \( m \), whose excitation grows as the spin decreases; for negative spins, the orbit changes direction during plunge (since the particle eventually locks to the rotating BH horizon), thus exciting \((\ell, -m, 0)\) modes. For \( \ell \neq m \) modes, instead, we have found QNM mixing across the entire spin range. For the \((2,1)\) mode, the main source of mixing is the \((2, -1, 0)\) QNM. For the \((3,2)\) mode (see Fig. 7), we have recognized 3 different behaviors: When \( q \gtrsim 0.8 \) the ringdown is dominated by \((2,2,0)\) with contamination from \((3,2,0)\); when \( 0 < q \lesssim 0.7 \), the ringdown is dominated by \((3,2,0)\) with contamination from \((2,2,0)\); when \( q \lesssim 0 \), the ringdown is dominated by \((3,2,0)\) with contamination from both \((3, -2, 0)\) and \((2,2,0)\). The excitation of QNMs with the same \( m \), but with different \( \ell \), is understood as a basis effect, since the QNMs are computed in a \(-2\)-spin-weighted spheroidal-harmonic separation of the Teukolsky equation, while the waveforms used in modeling are decomposed in \(-2\)-spin-weighted spheroidal-harmonic modes. We have fitted the relative amplitude between the main QNMs that are interfering for each mode (see Tables I and II), and have been able to model the ringdown Teukolsky waveforms using Eq. (5) for all spins except \( \tilde{q} = -0.95, -0.99 \) for all modes, and also \( \tilde{q} = -0.9 \) for the \((3,2)\) mode.

In Sec. VI, we have discussed how the inspiral-merger-ringdown Teukolsky waveforms helped the extension of the comparable-mass EOB model for nonprecessing, spinning BH binaries of Ref. [29] to small mass ratios and large spins. In particular, a time delay \( \Delta \tau_{\mathrm{22}} \) was introduced between the orbital frequency peak \( \tilde{\omega}_{\text{peak}} \) and the point \( \tilde{\omega}_{\text{peak}} + \Delta \tau_{\mathrm{22}} \), where no-quasicircular corrections are applied to the merger waveform. The specific dependence of the time-delay function on the spin takes into account the extreme circularity of the orbits encountered in the test-particle limit for large spins, and guarantees that \( \tilde{\omega}_{\text{peak}} + \Delta \tau_{\mathrm{22}} \) always lies in a region with significant radial motion. Older EOB models took \( \tilde{\omega}_{\text{peak}} + \Delta \tau_{\mathrm{22}} \) to coincide with the peak of the amplitude; however, in this paper, we have shown that such prescription is not adequate in the test-particle limit and, more generally when the mass ratio is smaller than \( \sim 1/100 \) if \( q > 0.8 \), since the peak occurs much before the ISCO and light-ring. The Teukolsky waveforms were also exploited to build fitting functions for the input values (i.e., \([h_{22}], [\partial_t h_{22}], \omega_{22}, \partial_t \omega_{22}\) at a point in time during merger) which are needed to impose nonquasicircular corrections to the merger EOB waveform.

Finally, in Sec. VII, we have evaluated the comparable-mass EOB model of Ref. [29] in the test-particle limit, and compared it to Teukolsky waveforms computed along the same EOB trajectory. We have found that, up to a spin \( \sim 0.5 \), the EOB waveforms (based on the factorized resummation of PN formulas in Refs. [25,26]) perform well, with phase differences within 0.1 rads and amplitude errors which are negligible up to merger (see Fig. 11). For larger spins, instead, while the EOB model can produce a reasonable \((2,2)\) mode waveform (see Fig. 12), still it disagrees with the Teukolsky data, due to the poor performance of the current factorized waveforms in such highly relativistic regimes—for example for \( q = 0.8 \), we find an amplitude error around 5% during the late inspiral and a dephasing of about 0.8 rads 50 GW cycles before merger, and growing as one moves to lower frequencies. We have also discussed the limitations of the current factorized EOB energy flux, namely the lack of horizon-absorption terms and the truncation at \( \ell = 8 \) modes.

The natural extension of this project will consider inclined orbits in Kerr spacetime. Even at the level of geodetic motion, there exist orbits with constant separation and inclination (with respect to the direction of the Kerr spin), which display precession of the orbital plane. Thus, these orbits will radiate waveforms which carry amplitude and phase modulations due to the precession. On the analytical side, we have shown in this paper several limitations of the current EOB factorized flux [25,26] for large spins. Thus, it will be crucial to improve this flux in the future either by designing a new resummation scheme, or by incorporating higher-order PN terms that have been recently computed [59]. Moreover, the current
EOB flux was developed for nonprecessing BH binaries only; we plan to test different prescriptions that could extend its validity to the precessing case. Such work can help the more challenging EOB modeling of precessing, comparable-mass BH binaries, which has first been tackled in Ref. [57].

ACKNOWLEDGMENTS

We thank Enrico Barausse, Yi Pan, Achilleas Porfyriadis, and Nico Yunes for useful, informative discussions. A. B. and A. T. acknowledge partial support from NSF Grants No. PHY-0903631 and No. PHY-1208881. A. B. also acknowledges partial support from the NASA Grant No. NNX12AN10G and A. T. from the Maryland Center for Fundamental Physics. This work was supported at MIT by NSF Grant No. PHY-1068720. G. K. acknowledges research support from NSF Grants No. PHY-0916906, No. CNS-0959382, No. PHY-1135664, and No. PHY-1303724, and from the U.S. Air Force Grant No. FA9550-10-1-0354 and No. 10-RI-CRADA-09.

APPENDIX: INPUT VALUES FOR NON-QUASICIRCULAR CORRECTIONS TO MERGER WAVEFORMS

In this appendix, we provide useful information about the Teukolsky merger waveforms that can be exploited in the construction of comparable-mass, spinning, nonprecessing EOB models that span the entire physical parameter space, as discussed in Sec. VI. We omit spin 0.99 because it is difficult to determine its peak positions \( r_{\text{peak}}^{m} \), due to the extreme flatness of the mode amplitudes, as shown in Sec. IV. We also omit the negative spins for the (2,1) and (3,2) modes since, as discussed in Sec. VB, QNM mixing has an early onset [around the turning point of the azimuthal motion for (2,1); slightly later than that for (3,2)], and affects the peak of the waveform; it is therefore ambiguous where to measure the input values for these cases.

In Fig. 13 we show how the time delay between the orbital frequency peak \( t_{\Omega}^{\text{peak}} \) and the Teukolsky amplitude peak \( t_{l}^{m}\text{peak} \) changes with the Kerr spin. As pointed out in Sec. IV, the amplitudes tend to peak earlier and earlier as \( q \) increases, well before the ISCO when \( q > 0.8 \). This creates difficulties when applying the non-quasicircular procedure to correct the EOB merger waveforms at \( t_{l}^{m}\text{peak} \), as elucidated in Sec. VI. In fact, in the comparable-mass EOB model of Ref. [29], we chose a delay \( \Delta t_{22}\text{peak} \) which decreases after spin 0.8, thus departing from the blue curve in Fig. 13.

In Figs. 14 and 15 we plot the input values computed at the time \( t_{l}^{m}\text{peak} \) when the Teukolsky amplitudes peak. The largest numerical uncertainties are visible on the curvature, but, as it turns out, the EOB waveforms are only mildly sensitive to such input value; in order to get a good modeling, the crucial input values are rather the values of the amplitude and the frequency.
[64] I. Hinder, A. Buonanno, M. Boyle, Z. B. Etienne, J. Healy et al., Classical Quantum Gravity 31, 025012 (2014).