Anatomy of the binary black hole recoil: A multipolar analysis

Jeremy Schnittman,1 Alessandra Buonanno,1 James R. van Meter,2,3 John G. Baker,2 William D. Boggs,4 Joan Centrella,2 Bernard J. Kelly,2 and Sean T. McWilliams4

1Maryland Center for Fundamental Physics, Department of Physics, University of Maryland, College Park, Maryland 20742
2Gravitational Astrophysics Laboratory, NASA Goddard Space Flight Center, 8800 Greenbelt Rd., Greenbelt, MD 20771
3Center for Space Science & Technology, University of Maryland Baltimore County, Physics Department, 1000 Hilltop Circle, Baltimore, MD 21250
4Department of Physics, University of Maryland, College Park, Maryland 20742

(Dated: June 7, 2007)

We present a multipolar analysis of the recoil velocity computed in recent numerical simulations of binary black hole coalescence, for both unequal masses and non-zero, non-precessing spins. We show that multipole moments up to and including l = 4 are sufficient to accurately reproduce the final recoil velocity (≈ 98%) and that only a few dominant modes contribute significantly to it (≈ 95%). We describe how the relative amplitude, and more importantly, the relative phase, of these few modes control the way in which the recoil builds up throughout the inspiral, merger, and ring-down phases. We also find that the numerical results can be reproduced, to a high level of accuracy, by an effective Newtonian formula for the multipole moments obtained by replacing in the Newtonian formula the radial separation with an effective radius computed from the numerical data. Beyond the merger, the numerical results are reproduced by a superposition of three Kerr quasi-normal modes. Analytic formulae, obtained by expressing the multipole moments in terms of the fundamental QNMs of a Kerr BH, are able to explain the onset and amount of “anti-kick” for each of the simulations. Lastly, we apply this multipolar analysis to understand the remarkable difference between the amplitudes of planar and non-planar kicks for equal-mass spinning black holes.

PACS numbers: 04.25.Dm, 04.30.Db, 04.70.Bw, 04.25.Nx, 04.30.-w

I. INTRODUCTION

After the recent breakthrough in numerical relativity (NR) [1–3], a number of different groups [4–6] are now able to evolve binary black holes (BHs) through merger. Recently, a great deal of effort has been directed towards the computation of the recoil velocity of the final BH [7–15]. The fundamental cause of this recoil is a net linear momentum flux in the gravitational radiation, due to some asymmetry in the system [16–18], typically unequal masses or spins in the case of BH binaries. The recoil has astrophysical importance because it can affect the growth of supermassive black holes (SMBHs) through hierarchical mergers [19, 20]. In those scenarios dark-matter halos grow through hierarchical mergers. The SMBHs at the centers of such haloes are thought to merge unless they had been kicked out of the gravitational potential well because the recoil velocity gained in a previous merger is larger than the halo’s escape velocity.

Other astrophysical implications include the displacement of the SMBH, along with its gaseous accretion disk, forming an “off-center” quasar. These quasars might also have emission lines highly red- or blue-shifted relative to the host galaxy due to the doppler shift of the recoil velocity. Additionally, these displaced SMBHs could in turn displace a significant amount of stellar mass from the galactic nucleus as they sink back to the center via dynamical friction, forming a depleted core of missing mass on the order of twice the SMBH mass [20–22].

Until now, numerical simulations have computed recoil velocities for non-spinning unequal-mass BH binary systems [7–9] in the range $m_2/m_1 = (1 \cdots 4)$, where $m_1$ and $m_2$ are the individual BH masses; for spinning, non-precessing binary BHs [10–12], and also for precessing BHs with both equal [13, 14] as well as unequal masses [15]. Quite interestingly, there exist initial spin configurations for which the recoil velocity can be quite large, e.g., ≈ 2500 km/sec [13–15]. However, it is not yet clear whether those very large recoil velocities are astrophysically likely [24–26]. So far, due to limited computational resources, the numerical simulations have explored a rather small portion of the total parameter space.

Analytic calculations, based on the post-Newtonian (PN) expansion of Einstein field equations [27] and its siblings [28–33], have made predictions for the recoil velocity [34–38] before the NR breakthrough. Since the majority of the linear momentum flux is emitted during the merger and ringdown (RD) phases, it is difficult to make definitive predictions for the recoil using only analytic methods. These methods need to be somehow calibrated to the NR results, so that they can be accurately extended during the transition from inspiral to RD. So far, in the non-spinning case, the PN model [37] has provided results consistent with NR all along the adiabatic inspiral; the effective-one-body (EOB) model [28, 29, 31] can reproduce the total recoil, i.e., also the contribution from the RD phase, but with large uncertainties [38]. More recently, Ref. [23] provided the first estimates of the distribution of recoil velocities from spinning BH mergers using the EOB model, calibrated to the NR results.
In Ref. [39], perturbative calculations which make use of the so-called close-limit approximation [40] have been used to predict the recoil for unequal-mass binary BHs moving on circular and eccentric orbits.

In this paper we present a diagnostic of the physics of the recoil, trying to understand how it accumulates during the inspiral, merger, and RD phases. We use several numerical simulations of non-spinning, unequal mass binary systems, as well as spinning, non-precessing binary systems obtained by the Goddard numerical relativity group.

We frame our understanding using the multipolar formalism originally laid out by Thorne [41-45]. We work out which (pairs of) multipole moments contribute most significantly to the recoil. We employ analytic, but leading order, formulae for the multipole moments of the linear momentum flux during the inspiral phase, and express the multipole moments in terms of a linear superposition of quasi-normal modes (QNMs) during the RD phase [46]. These analysis tools help us understand why for some binary mass and spin configurations the so-called “anti-kick” is present, and what determines the difference between the peak and the final recoil velocity. What we learn in this study will be used in a forthcoming paper to improve the analytic models, so that they can be used to interpolate between NR results, efficiently and accurately covering the entire parameter space.

This paper is organized as follow. In Sec. II, after introducing our definitions and notations, we review the main features of the numerical simulations. In Sec. III we discuss the multipolar expansion of the linear momentum, angular momentum and energy fluxes given in terms of the symmetric trace-free radiative mass and current moments, and show how to compute those fluxes from the multipole decomposition of the Weyl scalar $\Psi_4$. In Sec. IV, we analyse the multipole content of the numerical waveforms during the inspiral and ring-down phases. In Sec. V we show that, by properly normalizing the binary radial separation, the multipole moments computed at leading order in an expansion in $1/c$ can approximate quite well the numerical results. Moreover, a superposition of three QNMs matches the RD phase. In Sec. VI we apply the tools developed in the previous sections to understand, using analytic expressions, how the kick builds up during the inspiral, merger, and ring-down phases. Finally, Sec. VII contains a brief discussion of our mains results and future research directions.

II. NUMERICAL SIMULATIONS

In this section we introduce our definitions and notations, and review the main features of the numerical simulations. Throughout the paper, we adopt geometrical units with $G = c = 1$ (unless otherwise specified) and metric signature $(-1,1,1,1)$.

A. Definitions and conventions

Our complex null tetrad is defined using the time-like unit vector normal to a given hypersurface $\hat{\tau}$, the radial unit vector $\hat{r}$ and an ingoing (\bar{\ell}) and outgoing (\bar{n}) null vectors as

\[ \bar{\ell} \equiv \frac{1}{\sqrt{2}} (\hat{\tau} + \hat{r}), \]
\[ \bar{n} \equiv \frac{1}{\sqrt{2}} (\hat{\tau} - \hat{r}). \]

We define the complex null vectors $\bar{m}$ and $\bar{m}^*$ by

\[ \bar{m} \equiv \frac{1}{\sqrt{2}} (\bar{\ell} + \bar{n}), \]
\[ \bar{m}^* \equiv \frac{1}{\sqrt{2}} (\bar{\ell} - \bar{n}), \]

with the standard spherical metric at infinity $ds^2 = -d\tau^2 + dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\phi^2)$. In terms of this tetrad, the complex Weyl scalar $\Psi_4$ is given by

\[ \Psi_4 \equiv C_{abcd} n^a (m^b)^* n^c (m^d)^*, \]

where $C_{abcd}$ is the Weyl tensor and $*$ denotes complex conjugation.

To relate $\Psi_4$ to the gravitational waves (GWs), we note that in the transverse-traceless (TT) gauge,

\[ \frac{1}{4} (\hat{h}_{TT}^\phi - \hat{h}_{TT}^\theta) = - R_{\hat{\theta}\hat{\phi}} = - R_{\hat{\phi}\hat{\theta}} = - R_{\hat{\theta}\hat{\phi}} = - R_{\hat{\phi}\hat{\theta}}, \]

\[ = R_{\hat{\phi}\hat{\theta}} = R_{\hat{\theta}\hat{\phi}} = R_{\hat{\phi}\hat{\theta}}, \]

\[ = R_{\hat{\phi}\hat{\theta}} = R_{\hat{\theta}\hat{\phi}} = R_{\hat{\phi}\hat{\theta}}. \]

Following usual convention, we take the $h_+$ and $h_\times$ polarizations of the GW to be given by

\[ \hat{h}_+ = \frac{1}{2} (\hat{h}_{TT}^\theta - \hat{h}_{TT}^\phi), \]
\[ \hat{h}_\times = \hat{h}_{TT}^\phi. \]

Since the Riemann and Weyl tensors coincide in vacuum regions of the spacetime: $R_{abcd} = C_{abcd}$, we find

\[ \Psi_4 = - (\hat{h}_+ - i \hat{h}_\times). \]
\[ \pm 2 Y_{\ell m}(\theta, \varphi) = \left[ \frac{(\ell - 2)!}{(\ell + 2)!} \right]^{1/2} \left[ a_{(\ell m)}^\pm(\theta) Y_{\ell m}(\theta, \varphi) + \beta_{(\ell m)}^\pm(\theta) Y_{\ell-1m}(\theta, \varphi) \right], \]  

for \( \ell \geq 2 \) and \( |m| \leq \ell \), and with the functional coefficients

\[ a_{(\ell m)}^\pm(\theta) = \frac{2m^2 - \ell(\ell + 1)}{\sin^2 \theta} \mp 2m(\ell - 1) \frac{\cot \theta + \ell - 1}{\sin \theta} \cot^2 \theta, \]

\[ \beta_{(\ell m)}^\pm(\theta) = 2 \left[ \frac{2\ell + 1}{2\ell - 1} (\ell^2 - m^2) \right]^{1/2} \left( \frac{\pm m}{\sin^2 \theta} + \frac{\cot \theta}{\sin \theta} \right). \]

Finally, in the far field (\( r \gg M \)) we decompose the dimensionless Weyl scalar \( M r \Psi_4 \) as

\[ M r \Psi_4(t, \tilde{r}) = \sum_{\ell m} -2 C_{\ell m}(t) Y_{\ell m}(\theta, \varphi). \]

where \( M = m_1 + m_2 \) is the total mass of the binary system with \( m_1 \) and \( m_2 \) the masses of the individual BHs (see below for explanations), and \( \tilde{r} \) is the radial distance to the binary center of mass.

### B. Details of numerical simulations

We set up the simulations by placing the BHs on an initial Cauchy surface using the Brandt-Brügmann prescription [49]; the Hamiltonian constraint is solved using the second-order accurate multigrid solver AMR MG [50]. We use the Bowen-York [51] framework to incorporate the BH spins and momenta, with the choice of initial tangential momentum informed by the quasi-circular PN approximation of Damour et al. [32], Eq.(5.3).

The parameters for the runs considered in this paper are shown in Table I. We use the following notation. We indicate with EQ and NE the equal-mass and unequal-mass runs. The subscripts 0, +, - refer to zero spin, spin aligned, and spin anti-aligned with the orbital angular momentum, respectively. For the unequal-mass cases we use a superscript to indicate the mass ratio \( m_1 : m_2 \). We denote by \( m_1 \) the BH horizon mass computed as

\[ m_1 = \sqrt{m_{\text{irr},1}^2 + S_1^2 / 4m_{\text{irr},1}^2}, \]

where \( S_1 = a_1 m_1 \tilde{S}_1 = S_1 \tilde{S}_1 \) is the spin angular momentum of hole 1, \( m_{\text{irr},1} \) is its irreducible mass \( m_{\text{irr},1} = \sqrt{A_1 / 16\pi} \) [52], and \( A_1 \) is its apparent horizon area. Similar definitions hold for hole 2. The binary's total mass is \( M = m_1 + m_2 \), \( \delta m = m_1 - m_2 \), the mass ratio is \( q = m_1 / m_2 \leq 1 \), and the symmetric mass ratio is \( \eta = m_1 m_2 / M \). Following Kidder [35], we further define the spin vectors \( S = S_1 + S_2, \Delta \equiv M(S_2 / m_2 - S_1 / m_1) \), and \( \xi = S - (\delta m / M) \Delta \).

The mass and spin parameters of the final BH are \( M_f \) and \( a_f \). The values of \( M_f \) and \( a_f \) listed in Table I are computed from the loss of energy and angular momentum from the initial time to the end of the RD phase. They are compatible with the values obtained by extracting the fundamental QNM (see below Sec. IVB). All spins are orthogonal to the orbital plane, so \( \Delta^\varphi = \Delta^\varphi = 0 \) (the exception is a single run with planar spins discussed in Sec. VI D).

The simulations were carried out using the moving puncture method [53, 54] in the finite-differencing code Hahndol [55], which solves the Einstein equations in a standard 3+1 BSSN conformal formulation. Dissipation [56] terms (tapered to zero near the punctures) and constraint-damping [57] terms were added for robust stability. We used the gauge condition recommended in Ref. [58] for moving punctures, fourth-order accurate mesh-adapted differencing [59] for the spatial derivatives, and a fourth-order accurate Runge-Kutta algorithm for the time-integration. The adaptive mesh refinement and most of the parallelization was handled by the software package Paramesh [60], with fifth-order accurate interpolation between mesh refinement regions.

The grid spacing in the finest refinement region around each BH is \( h_f = 3M / 160 \). We extract data for the radiation at a radius \( r_{\text{ext}} = 45M \). The wave extraction was performed by 4th order interpolation to a sphere followed by angular integration with a Newton-Cotes formula. We have found satisfactory convergence of the results. For example, for the 2:1 mass ratio run, for which a higher resolution of \( h_f = 1M / 64 \) was run in addition to \( h_f = 3M / 160 \), the rates of convergence of the Hamiltonian and momentum constraints are comparable to that found in our equal mass runs reported in [61], and the radiated momenta from the two resolutions agree to within 2%. This was also true for a 3:2 mass ratio test case with spin (NE++ below).
### TABLE I: Parameters of the numerical simulations (see Sec. II B for explanations).

<table>
<thead>
<tr>
<th>Run</th>
<th>$m_1$</th>
<th>$m_2$</th>
<th>$\delta m$</th>
<th>$q$</th>
<th>$a_1/m_1$</th>
<th>$a_2/m_2$</th>
<th>$\Delta^2$</th>
<th>$\xi^2$</th>
<th>$\Sigma_{33}$</th>
<th>$M_{f}$</th>
<th>$a_{f}/M_{f}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ$_{+}$</td>
<td>0.503</td>
<td>0.503</td>
<td>0.0</td>
<td>1.0</td>
<td>-0.198</td>
<td>-0.198</td>
<td>-0.2</td>
<td>0.0</td>
<td>0.075</td>
<td>0.9668</td>
<td>0.697</td>
</tr>
<tr>
<td>NE$_{+}$</td>
<td>0.401</td>
<td>0.593</td>
<td>-0.192</td>
<td>0.677</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.9599</td>
<td>0.675</td>
<td></td>
</tr>
<tr>
<td>NE$_{0}$</td>
<td>0.333</td>
<td>0.667</td>
<td>-0.333</td>
<td>0.500</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.9662</td>
<td>0.633</td>
<td></td>
</tr>
<tr>
<td>NE$_{0}$</td>
<td>0.2</td>
<td>0.8</td>
<td>-0.6</td>
<td>0.250</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.961</td>
<td>0.423</td>
<td></td>
</tr>
<tr>
<td>NE$_{+}$</td>
<td>0.399</td>
<td>0.610</td>
<td>-0.210</td>
<td>0.655</td>
<td>0.201</td>
<td>-0.194</td>
<td>-0.2</td>
<td>0.002</td>
<td>0.072</td>
<td>0.9707</td>
<td>0.640</td>
</tr>
<tr>
<td>NE$_{+}$</td>
<td>0.399</td>
<td>0.610</td>
<td>-0.212</td>
<td>0.653</td>
<td>-0.201</td>
<td>0.193</td>
<td>-0.2</td>
<td>-0.002</td>
<td>-0.072</td>
<td>0.9667</td>
<td>0.704</td>
</tr>
</tbody>
</table>

### III. MULTIPOLAR FORMALISM

In this Section we review the most relevant results from Thorne [41], showing how a multipole decomposition of the gravitational radiation field can be used to calculate the energy, angular momentum, and linear momentum fluxes from a BH binary system. One of the fundamental assumptions we employ here is that it is possible to relate in a simple way the multipole moments of the source and the radiation field, so in much of the discussion below we will use these two descriptions interchangeably.

#### A. Linear momentum flux

In the literature [7–11, 15] it is common to compute the linear momentum flux, and then the recoil, using the following formula

$$\frac{dP_i}{dt} = \frac{r^2}{16\pi} \int d\Omega \frac{x_i}{r} \int_{-\infty}^{t} dt \Psi_4^2,$$  \hspace{1cm} (14)

where $r$ is the extraction radius and the antiderivative of $\Psi_4$ is used because the linear momentum flux scales as the square of the first derivative of the waveform, whereas $\Psi_4$ is proportional to the second derivative of the waveform [see Eq. (9) above]. To study how the different multipole moments contribute to the recoil, we could plug Eq. (12) into Eq. (14), as done, e.g., in Ref. [10]. Here, we prefer to use the expression of the linear momentum flux given in terms of the symmetric and trace-free (STF) radiative mass and current multipole moments, as done in Refs. [41–45].

Starting from Eq. (4.20') in Ref. [41], we can write the linear momentum flux as

$$F_j \equiv \frac{dP_j}{dt} = \frac{G}{c^7} \sum_{l=2}^{\infty} \left[ \frac{2(l + 2)(l + 3)}{(l + 1)(2l + 3)!} \mathbf{I}_{jA_l}^{(l+1)} \mathbf{I}_{A_l} \left( \frac{1}{c} \right)^{2(l-2)} + \frac{8(l + 3)}{(l + 1)(2l + 3)!} \mathbf{S}_{jA_l}^{(l+1)} \mathbf{S}_{A_l} \left( \frac{1}{c} \right)^{2(l-1)} \right],$$  \hspace{1cm} (15)

where $\mathbf{I}_{jA_l}$ ($\mathbf{S}_{A_l}$) are the STF $l$-dimensional radiative mass (current) multipole moments and left-hand superscripts represent time derivatives. From these tensors, we can construct the radiative multipole moments $T_l$ and $S_l$ according to the normalization given by Eq. (4.7) of Ref. [41]:

$$T_l \equiv \frac{16\pi}{(2l + 1)!!} \left( \frac{l + 1}{2} \right)^{1/2} l! \mathbf{Y}_{lA_l} \mathbf{Y}_{A_l}^{*},$$

$$S_l \equiv \frac{-32\pi l}{(l + 1)(2l + 1)!!} \left( \frac{l + 1}{2} \right)^{1/2} l! \mathbf{S}_{A_l} \mathbf{Y}_{lA_l}^{*},$$  \hspace{1cm} (16)

where $\mathbf{Y}_{lA_l}$ are $l$-dimensional STF tensors that are closely related to the usual scalar spherical harmonics by

$$Y_{l}^{(m)}(\theta, \phi) = Y_{11 \cdots 1}^{(m)} n^1 \cdots n^l,$$  \hspace{1cm} (17)

with $n^i = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)$. Note that the radiative moments $T_l$ and $S_l$ are complex scalars and have no explicit spatial dependence. To simplify the notation below, we incorporate the $(l + 1)$ time derivatives into the radiative multipole moments, and define

$$T^{(l)} \equiv (l+1)T_l, \quad S^{(l)} \equiv (l+1)S_l.$$  \hspace{1cm} (18)

By combining Eqs. (15), (16), and (18), we find that at
leading order (in a $1/c$ expansion) the linear momentum flux is given by

\[ F_x^{(0)} + i F_y^{(0)} = \frac{1}{336\pi} \left[ -14i\sqrt{6} I^{22*} + \sqrt{14} I^{31} I^{22*} - \sqrt{210} I^{22} I^{33*} + 7i\sqrt{6} I^{20} S^{21*} - 7i\sqrt{6} S^{20} I^{21} + 14i I^{21} S^{22*} + \sqrt{42} I^{30} I^{21*} - 2\sqrt{21} I^{20} I^{31} - 2\sqrt{35} I^{21} I^{32*} \right], \tag{19} \]

and

\[ F_x^{(0)} = \frac{1}{336\pi} \left[ 4\sqrt{14} R(I^{31} I^{21*}) - 14\Im(I^{21} S^{21*}) + 2\sqrt{35} R(I^{22} I^{32*}) - 28\Im(I^{22} S^{22*}) + 3\sqrt{7} I^{20} I^{30} \right]. \tag{20} \]

Note that Eq. (19) coincides with Eq. (9) in Ref. [38] when we express the radiative multipole moments in terms of the source moments [42–45] and reduce ourselves to a circular, non-spinning orbit in the $(x, y)$ plane. In this case only the first three terms in Eq. (19) survive.

The next highest order terms ($1/c^2$) are proportional to the mass octupole $I^{3m}$, or current quadrupole $S^{2m}$.

\[ F_x^{(1)} + i F_y^{(1)} = \frac{1}{672\pi} \left[ -7i\sqrt{6} S^{32} I^{33*} - 14\sqrt{6} I^{33} I^{44*} + 4\sqrt{21} S^{20} S^{31*} - 4\sqrt{35} S^{21} S^{32*} - 2\sqrt{210} S^{22} S^{33*} + 2\sqrt{42} S^{30} S^{21*} + 14i\sqrt{3} S^{30} S^{31*} - 14i\sqrt{3} S^{30} S^{31*} - 7i\sqrt{10} I^{31} S^{32*} - 7i\sqrt{10} I^{31} S^{32*} - 2\sqrt{105} I^{30} I^{41*} + 6\sqrt{10} I^{30} I^{31*} - 3\sqrt{70} I^{30} I^{42*} + 3\sqrt{14} I^{41} I^{32*} + 21\sqrt{2} I^{32} I^{43*} + 2\sqrt{14} I^{31} S^{22*} + \sqrt{42} I^{42} I^{33*} + 7i\sqrt{6} I^{32} S^{33*} \right], \tag{21} \]

and

\[ F_x^{(1)} = \frac{1}{336\pi} \left[ 3\sqrt{7} S^{30} S^{30} + 4\sqrt{14} R(S^{31} S^{31*}) + 2\sqrt{35} R(S^{22} S^{32*}) - 7\Im(I^{31} S^{31*}) - 14\Im(I^{32} S^{32*}) - 21\Im(I^{33} S^{33*}) + 2\sqrt{21} I^{30} I^{40} + 3\sqrt{35} R(I^{31} I^{41*}) + 6\sqrt{7} R(I^{32} I^{42*}) + 7\sqrt{3} R(I^{33} I^{43*}) \right]. \tag{22} \]

Note that all of the terms in Eqs. (19) and (21) contain products of multipoles with $m' = m \pm 1$, while the terms in Eqs. (20) and (22) have $m' = m$, as with familiar quantum-mechanical operators that involve similar $x_i$-weighted integrations over the sphere. Also note that for mass-mass and current-current terms, $l' = l \pm 1$.

The above formulae (19)–(22) are valid for completely general orbits, including spin terms and even for binary systems precessing out of the plane. However, we can simplify them significantly by rotating into the frame where the instantaneous orbital angular momentum is along the $z$-axis. Furthermore, by relating the radiative multipole moments to the source multipole moments [42–45] and assuming that terms proportional to $\bar{R}$ ($\bar{R}$ being the binary radial separation) are negligible, we find $I^{20} = I^{30} = \bar{S}^{30} = I^{32} = I^{40} = I^{41} = I^{43} = 0$. In the approximation of $\bar{R} = 0$, the inclusion of $\bar{R} \neq 0$ terms adds no new multipole modes. In the case of non-spinning BHs, the formulae (19)–(22) can be additionally simplified by setting $S^{20} = I^{21} = S^{22} = S^{31} = S^{33} = 0$. Quite interestingly, we obtain that the latter conditions are also valid in the special case of non-precessing BHs where the spins are aligned or anti-aligned with the orbital angular momentum. Since these are the cases we consider in this paper, we shall refer often to the following approximate formula for the linear momentum flux:

\[ F_x + i F_y \approx \frac{1}{672\pi} \left[ -28i S^{21} I^{22*} - 2\sqrt{210} I^{22} I^{33*} - 14\sqrt{6} I^{33} I^{44*} + 2\sqrt{14} I^{41} I^{22*} - 7i\sqrt{6} S^{32} I^{33*} \right], \quad F_z = 0. \tag{23} \]

As we will see below in Sec. IV A, the linear momentum flux contributions from $I^{31} I^{22*}$ as well as other higher-$l$
modes are typically smaller by at least an order of magnitude. When integrating Eq. (23) to get the recoil velocity, we also find that (due in large part to the relative phases between the modes) the contribution from $S^{32}_{33}*$ is rather minimal. Thus for most of the analysis that follows, we will focus solely on the first three terms of Eq. (23).

In the following, sometime we shall use

\[ F = \{ F_x, F_y, F_z \}, \quad \tilde{F} = \frac{F}{|F|}. \quad (24) \]

All the (non-precessing) numerical simulations we will analyze have $F_z = 0$, so we can introduce a complex scalar flux

\[ F = F_x + i F_y. \quad (25) \]

Since what we extract from the numerical simulations are the modes $-2C_{lm}$ computed over the sphere surrounding the binary, we need to relate the $-2C_{lm}$ to the radiative mass and current multipole moments defined above. From Eq. (4.3) of [41],

\[ h_{mm} = \sum_{lm} (\ell^{lm}T_{ab}^{E2,lm}m^a m^b + \ell^{lm}T_{ab}^{B2,lm}m^a m^b) \]

where $h_{mm} = h_{ab}m^a m^b$ and $h_{ab}$ is the metric perturbation $g_{ab} - \eta_{ab}$ in the transverse traceless gauge, which satisfies Eqs. (6a-6b), and $T_{ab}^{E2,lm}$ and $T_{ab}^{B2,lm}$ are the "pure-spin" harmonics of Thorne. From Appendix A of [62],

\[ T_{ab}^{E2,lm} = \frac{1}{\sqrt{2}} (-2Y^{lm} m_a m_b + 2Y^{lm} \bar{m}_a \bar{m}_b) \quad (27) \]

\[ T_{ab}^{B2,lm} = -\frac{i}{\sqrt{2}} (-2Y^{lm} m_a m_b - 2Y^{lm} \bar{m}_a \bar{m}_b). \quad (28) \]

Substituting Eqs. (27-28) into Eq. (26) and recalling that $m^a m_a = 0$ gives

\[ h_{mm} = \frac{1}{\sqrt{2}} \sum_{lm} (\ell^{lm}T_{ab}^{E2,lm}m^a m^b + \ell^{lm}T_{ab}^{B2,lm}m^a m^b). \quad (29) \]

Now taking the complex conjugate and using the fact that $+2Y^{lm} = (-1)^m -2Y^{l-m}$ [note there is a typo in Eq. (3.1) of Ref. [47]] we obtain

\[ \bar{h}_{mm} = \frac{1}{\sqrt{2}} \sum_{lm} (-1)^m (\ell^{lm}T_{ab}^{E2,lm} - i\ell^{lm}T_{ab}^{B2,lm}) -2Y^{l-m} \quad (30) \]

\[ = \frac{1}{\sqrt{2}} \sum_{lm} (-1)^m (\ell^{lm}T^{l-m} - i\ell^{lm}S^{l-m}) -2Y^{l}\quad (31) \]

Using the tetrad choice of Eqs. (1-9), $d_t^2 h_{mm} = \ddot{h}_x - i\ddot{h}_y = -i \psi$, which decomposed into spin -2 weighted harmonics gives

\[ d_t^2 h_{mm} = -\sum_{lm} -2C_{lm} -2Y^{lm}, \quad (32) \]

we then see term-by-term that

\[ (-1)^m(-l+2)T^{l+m} - i(-l+2)S^{l-m} = -\sqrt{2} -2C_{lm} \quad (33) \]

Recall that $(-1)^mT^{l+m} = T^{lm}$ and $(-1)^mS^{l-m} = S^{lm}$, which allows us to write

\[ (-l+2)T^{lm} = \frac{-1}{\sqrt{2}} [-2C_{lm} + (-1)^m -2C_{l-m}], \quad (34a) \]

\[ (-l+2)S^{lm} = \frac{-i}{\sqrt{2}} [-2C_{lm} - (-1)^m -2C_{l-m}], \quad (34b) \]

Equations (19)-(23) are expressed in terms of $T^{lm} = (l+1)T^{lm}$ and $S^{lm} = (l+1)S^{lm}$ which can be computed by integrating Eqs. (34a), (34b). To avoid the complication of an undetermined constant of integration, we typically integrate $-2C_{lm}(t)$ backwards in time, since in the numerical data (and what we expect happens in reality) all the moments go to zero exponentially after the merger. At early times on the other hand, most of the modes are significantly non-zero and also include a large amount numerical noise due to the initial conditions.

**B. Energy and angular momentum flux**

Unlike the equations for the linear momentum flux, which all involve "beating" between pairs of different modes, the energy and angular- momentum flux expressions involve terms of the form $|T^{lm}|^2$. As we will see below, for the comparable-mass binary systems that we analyse ($m_1 : m_2 = 1 : 1, 2 : 3, 1 : 2$), the amplitude of the mass quadrupole moment $T^{22}$ is roughly an order of magnitude larger than the next largest mode. Thus it almost completely dominates the energy and angular momentum fluxes, and we can write [see Eq. (4.16) in Ref. [41]]

\[ \frac{dE}{dt} = \frac{1}{32\pi} \sum_{l=2}^{\infty} \sum_{m=-l}^{l} (|T^{lm}|^2 + |S^{lm}|^2) \approx \frac{1}{16\pi} |T^{22}|^2. \quad (35) \]

For the numerical simulations considered in this paper, the only non-zero modes have $l + m$ even for $T^{lm}$ and $l + m$ odd for $S^{lm}$, so we can neglect the $(m, m \pm 1)$ cross-terms in Eq. (4.23) of Ref. [41]. These cross-terms are responsible for angular momentum loss in the $x$-$y$ plane, so it is reasonable that they must be zero for non-precessing planar orbits. In this case, where the angular momentum is in the $z$ direction, we have

\[ \frac{dJ_z}{dt} = \frac{i}{32\pi} \sum_{l=2}^{\infty} \sum_{m=-l}^{l} m(l) T^{lm} \ast (l+1)T^{lm} + m(l) S^{lm} \ast (l+1)S^{lm} \]

\[ = \frac{i}{8\pi} (2)T^{22} \ast (3)T^{22}, \quad (36) \]

where in the second equation we have restored the explicit time derivatives [see Eq. (18)].
Integrating Eqs. (35) and (36) term-by-term, we can calculate how much energy and angular momentum are radiated in each of the dominant modes, similar to the approach of [68]. These results are shown in Table II, along with the contributions from just the inspiral phase \( t < t_{\text{peak}} \) and just the RD phase \( t > t_{\text{peak}} \). We will see below in Section V that these various energy contributions agree closely with the Newtonian predictions for the relative mass-scalings. For example, the energy \( E_{22} \) in the inspiral phase should scale as \( \eta^4 \), while the RD contribution should scale like \( \eta^{12} \). It is important to note that the different moments have different scalings: \( E_{33} \sim \eta^2 \), while the \( E_{44} \) contribution has a much weaker dependence on mass ratio: \( E_{44} \sim \eta^2 (1 - 3\eta)^2 \).

In the limit of very large initial separation (small initial frequency), each of the \( E_{lm} \) and \( J_{lm} \) should converge to a finite value, with the notable exception of \( J_{22} \). It is well-known that the angular momentum of a binary system scales as \( R^{3/2} \), and is thus unbound in the limit of \( R \to \infty \), but it is interesting to see that the higher-order contributions to the angular momentum all converge at large \( R \). This can be understood directly from Eq. (36) in the Keplerian limit of \( R = \omega^{-2/3} \). At leading order, radiation reaction follows the relation \( dt \sim \omega^{-11/3} d\omega \) so the angular momentum in the inspiral is

\[
J_{22} = \frac{i}{8\pi} \int_{t=-\infty}^{t_0} dt (2) \overline{r}^{22*} \overline{r}^{22} \sim \int_{\omega_0}^{\omega_0} \omega^{2/3} \omega^{5/3} \omega^{-11/3} \to \infty. \tag{37}
\]

For all the other energy and angular momentum modes, the fluxes from Eqs. (35),(36) scale as \( \omega^{10/3} \) or higher powers, and thus converge when integrated over \( \omega^{-11/3} d\omega \).

IV. MULTIPOLAR ANALYSIS OF THE NUMERICAL SIMULATIONS

In this section we want to investigate how the different multipole moments contribute to the inspiral phase and the RD phase of unequal-mass binary systems.

A. Inspiral phase

As can be derived in PN theory [27] and it has been confirmed numerically in Refs. [53, 54], the \( l = 2, m = 2 \) mode in Eq. (12) is circularly polarized during the inspiral phase of a non-spinning equal-mass binary system. Because of that, Ref. [66] defined the (dominant) orbital angular frequency as

\[
\omega_{l^m}^{\text{D}} = -\frac{1}{m} \Im \left( \frac{-2C_{l^m}}{\omega_{l^m}} \right). \tag{38}
\]

Here, we extend (38) by defining several (dominant) angular-orbital frequencies, each of them being related to a specific multipole moment, \( I_{l^m} \) or \( S_{l^m} \), as

\[
\omega_{l^m}^{I^m} = -\frac{1}{m} \Im \left( \frac{I_{l^m}}{I_{l^m}} \right), \quad \omega_{l^m}^{S^m} = -\frac{1}{m} \Im \left( \frac{S_{l^m}}{S_{l^m}} \right). \tag{39}
\]
We plot these frequencies in Fig. 1 for the dominant multipole moments, i.e., $I^{22}$, $S^{21}$, $I^{33}$, $I^{44}$, and $S^{32}$. The amplitudes of the $I^{31}$ and $I^{42}$ modes are too weak and dominated by noise to extract a dominant frequency. In this figure, as well as most shown in the rest of the paper, we plot the time variable with respect to $t_{\text{peak}}$, the point at which $J^{22}$ reaches its peak, closely corresponding to the peak in GW energy emission [see Eq. (35)]. We notice that the frequencies associated to the modes with $l = m$ agree quite well during the inspiral, but for the frequency of the $S^{21}$ mode decouples from the others approximately $50M$ before the peak in the $I^{22}$ mode. As we shall see in Sec. VI, this is due to the fact that, during the ringdown phase, the dominant angular frequency associated to the $S^{21}$ mode is almost twice as large as those of the other leading modes [63–65]. This decoupling plays a major role in determining the shape of the kick and anti-kick (see Sec. VI below).

In Fig. 2 we show the amplitudes of the multipole moments in Eq. (23). The left panel in Fig. 2 refers to the NE$^{23}_{00}$ run, while the right panel to the NE$^{13}_{00}$ run. The mass-quadrupole moment $I^{22}$ clearly dominates in both cases, while the $I^{31}$ and $I^{42}$ modes are so weak as to be almost completely overwhelmed by numerical noise. In addition to having dissimilar amplitudes, the different moments also peak at slightly different times, which may be related to the fact that RD modes are excited at different times. In particular, the modes mentioned above with $l \neq m$ tend to peak later in time, perhaps due to a longer transition to the higher QNM frequency.
As we shall see in Sec. V, as the mass ratio becomes more extreme, the higher-order modes increase in relative amplitude, with $I^{33}$ and $S^{21}$ both proportional to $\delta m$. $T^{44}$ and $S^{22}$, however, scale as $\eta(1 - 3\eta)$, so they increase only slightly in the range of masses considered here.

Next, in Fig. 3, we show the amplitude of the linear momentum flux from the mode-pairs included in Eq. (23). Again, the mass-quadrupole terms dominate, with significantly smaller contributions from the $S^{22}$ and $S^{12}$ modes. However, note the appreciable flux amplitude from the $I^{22}I^{22*}$ term, which is formally a higher-order correction in a $(1/c)$ expansion [37, 38]. From Fig. 3, we expect that the first three pair of modes in Eq. (23) should contribute most significantly to the recoil. Including the complex phase relations between the different modes, we find this result will be supported further by the analysis in Sec. VIA.

### B. Ringdown phase

We now extract the QNMs, notably the fundamental and the first two overtones, present in the most significant multipole moments. We follow the procedure outlined in Ref. [66]. To avoid possible constant offsets introduced by integrating Eqs. (34a),(34b), we prefer to extract the QNMs directly from the $I_{lm}$'s instead of using $I_{lm}$'s or $S_{lm}$'s. Following the approach of [65], we define the complex frequencies $\sigma_{lmn}$:

$$\sigma_{lmn} \equiv \omega_{lmn} - i/\tau_{lmn},$$

and each RD mode is proportional to $\exp(-i\sigma_{lmn}t)$. In this format, $\omega_{lmn}$ are the QNM oscillation frequencies (not to be confused with the dominant frequencies of Eq. (39)) and $\tau_{lmn}$ are the mode decay times, all functions of the final black hole mass and spin. In this notation, $l$ and $m$ are the same spherical wavenumbers used above, and $n = 0$ denotes the fundamental mode, with $n = 1, 2, \ldots$ corresponding to the higher harmonics. The fundamental QNM frequencies $\sigma_{lm0}$ are shown in Table III for the NR runs listed above. All frequencies and decay times are measured in units of the final mass $M_f$.

We present the RD analysis only for the NEE$^{2,3}_{00}$ run, but the others are qualitatively very similar. By applying a 4-parameter fit to the most dominant mode, i.e., $-2C^{22}$, we obtain $a_f/M_f = 0.669$ and $M_f/M_f = 0.965$ together with the amplitude and phase of the fundamental QNM. Using the results of Ref. [65], we find $M_\omega = 0.538$ and $M_\tau = 0.0847$. We then do a 6-parameter fit and successfully extract also the first overtone, obtaining slightly different values for $a_f/M_f = 0.661$ and $M_f/M_f = 0.958$. We find it impossible to extract, with a 8-parameter fit, also the second overtone. By contrast if we keep $a_f/M_f$ and $M_f/M_f$ fixed and equal to the values obtained when extracting the fundamental QNM, we obtain that we can fit up to the second overtone. Moreover, quite interestingly, the fit provides a waveforms which compares very well with the NR waveforms up to $\sim 5M$ before the peak of $-2C_{22}$, as can be seen in the upper left panel of Fig. 4.

The other panels of Fig. 4 show results for the other relevant modes $-2C_{33}, -2C_{44}$ and $-2C_{32}$. As obtained in Ref. [66], we find a mode-mixing in $-2C_{32}$, i.e., the QNM is a combination of $l = 2, m = 2$ and $l = 3, m = 2$. By fitting the fundamental QNM we obtain $a_f/M_f = 0.670$ and $M_f/M_f = 0.972$; $a_f/M_f = 0.412$ and $M_f/M_f = 0.838$; $a_f/M_f = 0.628$ and $M_f/M_f = 0.972$, for $-2C_{33}, -2C_{44}$ and $-2C_{32}$, respectively. Those values for the final BH spin and mass are rather consistent, except for $-2C_{44}$. This discrepancy might be due to numerical resolution effects, and will be the object of future investigations.
TABLE III: Frequencies and decay times for the fundamental QNMs for each of the numerical simulations.

<table>
<thead>
<tr>
<th>Run</th>
<th>$a_3/M_3$</th>
<th>$\sigma_{210}$</th>
<th>$\tau_{210}$</th>
<th>$\sigma_{220}$</th>
<th>$\tau_{220}$</th>
<th>$\sigma_{320}$</th>
<th>$\tau_{320}$</th>
<th>$\sigma_{330}$</th>
<th>$\tau_{330}$</th>
<th>$\sigma_{440}$</th>
<th>$\tau_{440}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ+</td>
<td>0.697</td>
<td>0.454</td>
<td>12.2</td>
<td>0.531</td>
<td>12.4</td>
<td>0.758</td>
<td>11.9</td>
<td>0.841</td>
<td>12.0</td>
<td>1.14</td>
<td>11.8</td>
</tr>
<tr>
<td>NE_{20}</td>
<td>0.675</td>
<td>0.450</td>
<td>12.1</td>
<td>0.521</td>
<td>12.2</td>
<td>0.740</td>
<td>11.7</td>
<td>0.827</td>
<td>11.9</td>
<td>1.12</td>
<td>11.7</td>
</tr>
<tr>
<td>NE_{10}</td>
<td>0.633</td>
<td>0.442</td>
<td>11.9</td>
<td>0.505</td>
<td>12.1</td>
<td>0.734</td>
<td>11.6</td>
<td>0.803</td>
<td>11.7</td>
<td>1.09</td>
<td>11.5</td>
</tr>
<tr>
<td>NE_{14}</td>
<td>0.423</td>
<td>0.411</td>
<td>11.5</td>
<td>0.445</td>
<td>11.5</td>
<td>0.674</td>
<td>11.1</td>
<td>0.711</td>
<td>11.1</td>
<td>0.963</td>
<td>10.9</td>
</tr>
<tr>
<td>NE_{33}</td>
<td>0.640</td>
<td>0.443</td>
<td>11.9</td>
<td>0.507</td>
<td>12.1</td>
<td>0.736</td>
<td>11.6</td>
<td>0.806</td>
<td>11.7</td>
<td>1.09</td>
<td>11.5</td>
</tr>
<tr>
<td>NE_{23}</td>
<td>0.704</td>
<td>0.456</td>
<td>12.2</td>
<td>0.533</td>
<td>12.4</td>
<td>0.760</td>
<td>11.9</td>
<td>0.845</td>
<td>12.1</td>
<td>1.14</td>
<td>11.9</td>
</tr>
</tbody>
</table>

FIG. 4: Comparison of numerical and QNM waveforms for the NE_{20} run. The dominant modes analyzed are $-2C_{22}$ (upper left), $-2C_{33}$ (upper right), $-2C_{32}$ (lower left), and $-2C_{44}$ (lower right). We denote with $t_{\text{peak}}$ the time of the peak of $I_{\text{peak}}$.

V. EFFECTIVE NEWTONIAN MODEL

In an attempt to better understand the amplitudes and frequencies of the various modes during the inspiral and merger phases, we present here what we call the "effective Newtonian" (eN) model. It begins with calculating the leading-order Newtonian formulae for each multipole moment of the source, as a function of the BH masses, binary separation, and orbital frequency. To extend these formulae through the end of the inspiral and into the merger phase, we introduce an effective radius coordinate that is a way of absorbing PN effects into the leading-order multipole expressions. Each multipole moment is then individually matched to a linear superposition of ring down modes, as is done in the effective-one-body model. Taken together, this eN model provides an excellent framework within which we can understand the details of the linear momentum flux and integrated recoil velocity.
A. Newtonian Multipole Moments

Working at leading Newtonian order for each mode, we equate the radiative multipole moments to the source multipole moments. Restricting ourselves to circular, planar orbits, and setting \( M = 1 \), we find that for non-spinning systems, the dominant modes are [42-45]:

\[
S_{n=1}^{21} = -\frac{8}{3} \sqrt{\frac{2\pi}{5}} \delta m \eta R^3 \omega^4 e^{-i\phi}, \quad (41a)
\]

\[
I_{n=1}^{22} = i16 \frac{2\pi}{5} \eta R^2 \omega^3 e^{-2i\phi}, \quad (41b)
\]

\[
I_{n=1}^{31} = -\frac{2}{3} \frac{\pi}{35} \delta m \eta R^3 \omega^4 e^{-i\phi}, \quad (41c)
\]

\[
S_{n=2}^{22} = -\frac{16}{3} \sqrt{\frac{2\pi}{7}} (1 - 3\eta) R^4 \omega^5 e^{-2i\phi}, \quad (41d)
\]

\[
I_{n=2}^{33} = \frac{54}{21} \delta m \eta R^3 \omega^4 e^{-3i\phi}, \quad (41e)
\]

\[
I_{n=2}^{32} = \frac{16}{63} \frac{2\pi}{7} (1 - 3\eta) R^4 \omega^5 e^{-2i\phi}, \quad (41f)
\]

\[
I_{n=2}^{44} = -\frac{256}{9} \frac{2\pi}{7} (1 - 3\eta) R^4 \omega^5 e^{-4i\phi} \quad (41g)
\]

where \( R \) is the radial separation and \( \omega = \dot{\phi} \) is the binary orbital frequency. Considering only the mass quadrupole terms in the linear momentum flux, we obtain the well-known result valid at Newtonian order [38]:

\[
F^{(0)} = -\frac{464}{105} \delta m \eta^2 R^5 \omega^7 e^{i\phi}. \quad (42)
\]

Including the next-highest order moments in Eq. (21), we get

\[
F^{(1)} = -i\frac{11120}{1323} \delta m \eta^2 (1 - 3\eta) R^7 \omega^9 e^{i\phi}. \quad (43)
\]

While there may also be next-to-leading order contributions from a PN expansion of the multipole moments included in Eq. (19) that would show up in Eq. (43), we can effectively absorb those corrections into the \( R \) variable, as will be described below. Combining Eqs. (42) and (43) we find the linear momentum flux scales like

\[
|F^{(0)} + F^{(1)}| \propto \delta m \eta^2 \left[ 1 + \frac{3475}{1827} (1 - 3\eta) R^2 \omega^2 \right]
\]

\[
\approx \frac{3}{2} \delta m \eta^2 (1 - 0.9\eta), \quad (44)
\]

which is remarkably similar to the result found in [9]. Here we have used \( R^2 \omega^2 \approx 0.23 - 0.25 \) at the peak of the energy flux, which seems to be quite standard for a range of mass ratios. However, the extremely close agreement with Ref. [9] is probably to some degree a coincidence, since this simple Newtonian formula does not include any details of the phase relations between different modes. As we show below in Sec. VIIB, in the transition from inspiral to ringdown, these phases play a major role in determining the amplitude of the un-kick, and thus the final recoil velocity. We find that more extreme-mass ratio BH binaries have a relatively smaller un-kick, which should also play an important role in the scaling relation of [9].

If we compute the above multipole moments \((41a)-(41g)\) using \( \omega \) as given by Eq. (39) and \( R \) as obtained from the puncture trajectories, we do not find a very good agreement with the numerical results. This is not surprising since there is no reason to believe that the Newtonian approximation should work well all along the inspiral phase. We should expect that higher-order PN corrections become important as we approach the merger. Since our scope is limited to a diagnostic of the NR results, and not to a precise comparison with PN calculations, instead of including PN corrections in Eqs. (41a)-(41g) and (42), we investigate whether by properly scaling the Newtonian expressions we can get a better agreement until the merger. We can also think of this normalization as a way of resumming the PN expansion.

Quite interestingly, if we compute the amplitudes \(|I_{lm}|\) or \(|S_{l}^{lm}|\) from the numerical data, and the angular frequency \( \omega \) from Eq. (39), we find that the radii \( R_{lm} \) which appear in the RHS of Eqs. (41a)-(41g) are rather independent of the multipole moments \( l \) and \( m \), as Fig. 5 shows. We denote the radii \( R_{lm} \) computed numerically as effective radii \( R_{\text{eff}}^{lm} \). The close agreement between the frequencies (see Fig. 1) and effective radii for each mode suggests we can use the Newtonian expressions and a single \( R_{\text{eff}}(t) \), e.g., the \( R_{22}^{(2)}(t) \) for all modes with a high degree of accuracy for the entire inspiral phase and even during the transition to merger.

For comparison we show in Fig. 5 also the radius from the puncture trajectory (dot-dashed curves) and the radius computed using the Arnowitt-Deser-Misner transverse-traceless gauge (dashed curves), given as function of frequency through 3PN order by [67]
FIG. 5: Effective radius for different modes, derived from Eqs. (39), (41a)-(41g). The close agreement for the $R_{\text{eff}}^{(i)}$ suggests we can use a single effective radius $R_{\text{eff}}^{(i)}(t)$ for the Newtonian expressions. We believe that the large oscillations in $R_{\text{eff}}^{(i)}$ are due to initial eccentricity at early times. Also plotted is the ADM radius (dashed curves) derived from the orbital frequency via Eq. (VA) and the coordinate separation of the BH punctures (dot-dashed curves). The results correspond to the NE$^{2,5}$ (left panel) and NE$^{1,2}$ (right panel) runs. We denote with $t_{\text{peak}}$ the time at which $I^{22}$ reaches its maximum.

In the next section, we shall investigate how this simple effective Newtonian model can be combined with a superposition of QNMs, as described in Sec. IV B, giving a good representation of the NR results.

B. Matching to ringdown

We now match the inspiral and RD waveforms in a mode-by-mode fashion following the philosophy of the EOB approach [30]. A similar attempt was followed in Ref. [38], where for simplicity the authors performed the matching to the Schwarzschild QNM frequencies, while we use the Kerr QNM frequencies and match to the fundamental QNM frequency and the first two overtones, as done in Ref. [66]. We obtain the QNM frequencies and decay times from Ref. [65]. For the fundamental and two overtone QNMs, we can match a given multipole mode by equating it and two time derivatives to a linear combination of QNMs.

We write

$$I^{lm}(t) = A(t) e^{-i\phi(t)} = \sum_{r_i=0}^{\infty} A_{lmn} e^{-i\sigma_{lnm}(t-t_{\text{match}})}, \quad (46)$$

where the complex QNM frequencies are known functions of the final BH mass and spin, and we must solve for the complex amplitudes $A_{lmn}$. Matching three QNMs we get

$$I^{lm}(t_{\text{match}}) = \sum_{n=0}^{2} A_{lmn},$$

$$\frac{d}{dt} I^{lm}(t_{\text{match}}) = -i \sum_{n=0}^{2} \sigma_{lmn} A_{lmn},$$

$$\frac{d^2}{dt^2} I^{lm}(t_{\text{match}}) = - \sum_{n=0}^{2} \sigma_{lmn}^2 A_{lmn}, \quad (47)$$

or as a simple matrix equation

$$\begin{pmatrix} 1 & 1 & 1 \\ -i\sigma_{l0m} & -i\sigma_{l1m} & -i\sigma_{l2m} \\ -\sigma_{l0m}^2 & -\sigma_{l1m}^2 & -\sigma_{l2m}^2 \end{pmatrix} \begin{pmatrix} A_{l0m} \\ A_{l1m} \\ A_{l2m} \end{pmatrix} = \begin{pmatrix} I^{lm} \\ j^{lm} \\ \tilde{I}^{lm} \end{pmatrix} \quad (48)$$

In Fig. 6, we compare the NR modes to the modes obtained by the effective Newtonian model defined in Sec. VA until $t_{\text{match}}$ and by the superposition of three QNMs for $t > t_{\text{match}}$. During the inspiral, the different moments are calculated according to Eqs. (41a)-(41g), using a single $R_{\text{eff}}$ and $\omega_D$ determined from the $I^{22}$ mode, with the exception of the $S^{21}$ mode, where we instead use the higher frequency $\omega_D^{21}$. The QNM frequencies and decay time are obtained from the mass and spin of the final BH (see Tables I, III), so they are not computed $a$ priori. We treat $t_{\text{match}}$ as a free parameter: if we stop the inspiral too early, the effective Newtonian mode amplitudes are still growing, so the sudden transition to decaying RD modes prematurely reduces them. On the other hand, if the inspiral is continued too long, we tend to lose the important phase shifts between the modes that only begin during the transition to RD. This is particularly evident in the $I^{44}$ mode, which undergoes
FIG. 6: Comparison of the effective Newtonian and NR radiative modes during inspiral, merger and RD phases. The data refer to the NE3GB run. We denote with $t_{\text{peak}}$ the time at which $I_{22}$ reaches its maximum.

An unexplained 180° phase-shift around the transition to RD, and also decays at somewhat different rates than is predicted from QNM theory (see above, Sec. IV B). Motivated by the results of Sec. IV B, notably by the fact that a superposition of three QNMs can fit very well the NR waveforms starting from the peak of the energy flux, we choose as best-matching point the peak of the energy flux.

Having shown a reasonably close match for each of the radiative multipoles between the effective Newtonian model and the numerical data, it stands to reason that the total recoil calculated with this model should agree, as well. This is shown in Fig. 7, where we have also varied the matching point around $t_{\text{peak}}$. The significant range in final recoil velocities is due to the phase shifting of the different modes around $t = t_{\text{peak}}$, most notably that of the $I_{44}$ mode described above. In Section VI B below, we will explore this phasing in greater detail and show how it affects the overall kick. At this point, we unfortunately do not have a clear understanding of the underlying cause of the de-phasing, but it may well be related to the slightly different times of transition from inspiral to ringdown for the different modes. Preliminary results also suggest that this de-phasing effect is reduced in more extreme-mass-ratio systems.

VI. ANATOMY OF THE KICK

In the above Sections, we have laid the groundwork for a multipolar analysis of the gravitational recoil, describing the momentum flux as a combination of radiative multipole modes. Along with the pseudo-analytic models for the inspiral and ringdown phases, we can now give a detailed description of the “anatomy” of the kick, namely the way the different modes combine to produce a peak recoil velocity, followed by a characteristic un-kick and then asymptotic approach to the final value of the BH recoil.

A. Contribution from different moments

In Sec. III A, we showed how the individual multipole moments contribute to the linear momentum flux through the integral of the $\Psi_4$ scalar [Eqs. (12), (14)].
FIG. 7: Comparison of the effective Newtonian model and NR predictions for the recoil velocity for a range of inspiral-RD matching points. We denote with $t_{\text{peak}}$ the time at which $r_2'$ reaches its maximum. The data refer to the $\text{NE}_{\text{gb}}$ run.

Here, we want to determine exactly which modes we need to include in the multipole expansion Eq. (15) to get a good representation of the full recoil, and which are the pairs of modes in Eq. (23) that contribute most.

By including only a select choice of terms in the $\Psi_4$ expansion Eq. (12), we can calculate the linear momentum flux by direct integration of Eq. (14) and compare it with the predictions of Eqs. (19)-(23), in each case including only the appropriate moments. This is a good way of double-checking those lengthy equations term-by-term, and in practice we find excellent agreement, limited only by the numerical accuracy of the simulations. Similarly, we can use this method of truncated expansion to determine which modes are necessary for calculating the recoil up to a given accuracy. The results of using higher and higher order multipolar moments are shown in Figs. 8, 9 for the $\text{NE}_{\text{gb}}$ and $\text{NE}_{\text{b}}$ runs.

In the left panel of Figs. 8, 9 we show with a solid curve the integrated momentum kick from Eq. (14), with a dashed curve the contribution from terms up to $l = 4$, i.e., the ones obtained by Eq. (19) and (21), and with a dotted curve the contribution from the leading three terms in Eq. (23). We conclude that the linear momentum flux is dominated by the $I^{22}I^{22*}$, $I^{33}I^{44*}$, and $S^{21}I^{22*}$ terms, which combine to produce the primary kick and anti-kick agreeing with the exact result within $\lesssim 10\%$ throughout the entire merger. Note that the flux from the $S^{33}I^{33*}$ term, while not insignificant in Fig. 3, contributes almost nothing to the net recoil velocity. This is largely due to phase relations between the various modes during the transition from inspiral to ringdown, described below in Sec. VI B.

In the right panel of Figs. 8, 9 we show the difference between the calculation obtained including terms up to $l = 3, 4, 5, 6$, and the exact result. It seems clear that we need modes up to and including $l = 4$ to get an accurate estimate of the recoil velocity. For more extreme mass ratios, higher-order moments become relatively more important, but remain strongly sub-dominant to the $l \leq 4$ modes [68].

To understand more clearly the relative contributions of the different modes to the total recoil, we will include analysis of a few more simulations including non-precessing spins. As mentioned above in Sec. III A, non-precessing spins do not introduce any additional moments compared to the non-spinning simulations, but simply modify the relative amplitudes of the different modes in Eq. (23) by adding the spin terms. Thus, once we determine how the spins modify the individual modes, we can use the same analysis for the spinning and non-spinning cases.

Again setting $M = 1$ and relating the radiative multipole moments to the source moments, we get the leading order spin-orbit modifications to Eqs. (41a)-(41g) [see Eqs. (3.14),(3.20) in Ref. [35] and Eq. (5.5) in Ref. [69]]:

$$S_{SO}^{31} = -4i \sqrt{\frac{2\pi}{3}} \eta R \omega^3 e^{-i\phi} \Delta z,$$  \hspace{1cm} (49a)
$$I_{SO}^{32} = \frac{64}{3} \frac{2\pi}{5} \eta R^2 \omega^4 e^{-2i\phi} \xi z,$$  \hspace{1cm} (49b)
$$S_{SO}^{33} = -\frac{32}{3} \frac{2\pi}{7} \eta R^2 \omega^4 e^{-2i\phi} \xi z,$$  \hspace{1cm} (49c)
$$I_{SO}^{31} = -\frac{2}{3} \frac{\pi}{35} \eta R^3 \omega^5 e^{-i\phi} \Sigma_{31},$$  \hspace{1cm} (49d)
$$I_{SO}^{33} = 54 \frac{\pi}{21} \eta R^3 \omega^5 e^{-3i\phi} \Sigma_{33},$$  \hspace{1cm} (49e)

where we have introduced the spin vectors

$$\Sigma_{31} \equiv \frac{11}{2} \delta m S + \frac{1}{2} (11 - 39\eta) \Delta,$$ \hspace{1cm} (50a)
$$\Sigma_{33} \equiv \frac{3}{2} \delta m S + \frac{3}{2} (1 - 5\eta) \Delta.$$ \hspace{1cm} (50b)
In the left panel, we show the net recoil kick, integrated from the linear momentum flux via Eq. (14) (solid curve), from all modes with $l \leq 4$ (dashed curve) and also limiting the modal composition of $\Psi_+$ to just the three dominant mode pairs in Eq. (23) (dotted curve). In the right panel we show the difference between the exact result and the $P_1$ expansion Eq. (12), limited to $l \leq 3,4,5,6$. The data refer to the NE$^{40}_{12}$ run. We denote with $t_{\text{peak}}$ the time at which $I^{22}$ reaches its maximum.

In all of the simulations considered here, where the dimensionless spins are equal ($a_1/m_1 = a_2/m_2$) and point in opposite directions, $\xi^2 = 0$, so we are left only with the modification of $S^{21}$ due to $\Delta^2$ and $I^{33}$ due to $\Sigma^3_3$.

Then Eqs. (41a)-(41g), and (49) give the linear momentum flux $F = F_x + iF_y$ during the inspiral for each of the first three dominant terms in Eq. (23):

$$F_{21,22}^{\text{insp}} = \frac{16}{45} \eta^2 R^3 \omega^6 (2\delta m R^2 \omega + 3\Delta^2) e^{i\phi},$$

$$F_{22,33}^{\text{insp}} = -\frac{36}{7} \eta^2 R^5 \omega^7 (\delta m + \omega \Sigma_{33}) e^{i\phi},$$

$$F_{33,44}^{\text{insp}} = -\frac{64}{7} \eta^2 (1 - 3\eta) R^7 \omega^9 (\delta m + \omega \Sigma_{33}^2) e^{i\phi}.$$
for the four runs $\text{NE}^2_3$, $\text{NE}^3_3$, $\text{NE}^3_0$, and $\text{EO}^+$. As seen in Table I, the $\text{NE}^2_3$ run has $\Delta^2 = 0.2M^2$, while the $\text{NE}^3_3$ run has $\Delta^2 = -0.2M^2$, respectively, adding destructively and constructively with the $\delta m$ term in Eq. (51a). This difference is clearly seen in the blue curves in Fig. 10. Also notable in these plots is the difference in mode amplitudes due to a similar effect from the constructive/destructive additions of $\delta m$ and $\Sigma_{53}$ in Eq. (51b). As we see in Fig. 10, $\text{NE}^3_{-3}$ appears to be the average of $\text{NE}^3_3$ and $\text{NE}^2_3$, while the flux from $\text{EO}^+$ is strongly suppressed due to the $\delta m = 0$ terms in Eq. (51c), leaving only the flux from $\Delta^2 = -0.2M^2$ and $\Sigma_{53} = 0.075$.

In each panel of Fig. 10, we also plot with dashed lines the eN prediction for the various flux amplitudes. In almost all cases, the eN flux is quite close to the NR results up to about $10M$ before $t_{\text{peak}}$, when the effective radii of Fig. 5 begin to diverge. The amplitude differences near the peaks are comparable to those see in Fig. 6 for the $\text{NE}^3_0$ run. The notable exception is the $F^{21,22}$ flux from the $\text{NE}^3_3$ and $\text{EO}^+$ runs, where there may be additional spin-spin terms that become important at these relatively low amplitudes.

B. Transition to ringdown and the de-phasing of the multipole modes

Since the flux vectors defined in the previous section will not generally be co-linear (and in the case of precessing spins, not even co-planar), to understand the time evolution of the recoil velocity, we must first understand the phase relations between the different modes. From Eqs. (41a)–(41g) and (51a)–(51c), we see that during the inspiral phase, the individual moments and the resulting flux vectors evolve according to a single orbital phase $\phi$, with $F^{21,22}$ pointing in the opposite direction of $F^{21,22}_{\text{insp}}$ and $F^{33,44}_{\text{insp}}$. However, as we can see from Fig. 1, as the binary evolves from inspiral to RD, the frequency (and thus phase) of the $S^{21}$ mode decouples from the other dominant modes. Upon closer inspection, we find that even the $I^{22}$, $I^{33}$ and $I^{44}$ modes deviate from each other enough to undergo a significant phase shift at the inspiral-RD transition.

To quantify these effects, we define the phase differ-
Throughout the inspiral, merger, and ringdown phases. In the inspiral phase, we can see that for the unequal-mass runs, where \( \delta m \) dominates Eqs. (51a)–(51c), we have
\[
\cos \psi^{2-3}_{\text{insp}} = \cos \psi^{2-4}_{\text{insp}} = -1, \quad \cos \psi^{3-4}_{\text{insp}} = 1.
\] (53)

For the EQ+ run with \( \delta m = 0 \), Eq. (51a) predicts that all phases have \( \cos \psi_{\text{insp}} = 1 \) during the inspiral. During the RD phase, using Eq. (46), we can write the flux vectors and phase evolution in terms of the fundamental QNM frequencies \( \omega_{l0m} \):

\[
F_{21,22}^{\text{RD}} = F_{\text{match}}^{21,22} \exp[-i(\sigma_{210} - \sigma_{220})(t - t_{\text{match}})],
\]
(54a)
\[
F_{22,33}^{\text{RD}} = F_{\text{match}}^{22,33} \exp[-i(\sigma_{220} - \sigma_{330})(t - t_{\text{match}})],
\]
(54b)
\[
F_{33,44}^{\text{RD}} = F_{\text{match}}^{33,44} \exp[-i(\sigma_{330} - \sigma_{440})(t - t_{\text{match}})].
\]
(54c)

Here \( \Phi_{\text{match}} \) is a phase offset determined at the transition from inspiral to ringdown. Quite interestingly, we find that for the range of BH spin parameters \( 0.5 \leq a_1/M_f \leq 0.9 \), the linear combinations of frequencies in Eqs. (55a)–(55c) vary by less than 10%. Thus, if we compute the above expressions for the \( \omega_{0m0} \) corresponding to \( a_1/M_f = 0.7 \), we have [65]

\[
\cos \psi^{2-3}_{\text{RD}} \simeq \cos \left[ \frac{0.23}{M_f} (t - t_{\text{match}}) + \Phi^{2-3}_{\text{match}} \right],
\]
(56a)
\[
\cos \psi^{2-4}_{\text{RD}} \simeq \cos \left[ \frac{0.22}{M_f} (t - t_{\text{match}}) + \Phi^{2-4}_{\text{match}} \right],
\]
(56b)
\[
\cos \psi^{3-4}_{\text{RD}} \simeq \cos \left[ \frac{0.012}{M_f} (t - t_{\text{match}}) + \Phi^{3-4}_{\text{match}} \right].
\]
(56c)

Even more intriguing, we find that for the unequal-mass simulations described above, the phase relations during the inspiral and RD are almost identical, regardless of spin orientations. This can be seen clearly in Fig. 11, which plots \( \cos \psi \) during inspiral, merger and RD for the different runs. The colinearity of the flux vectors is clear during the inspiral phase, and the sinusoidal oscillations of the phases during RD agree well with the analytic predictions (plotted in dashed curves in Fig. 11). Since the analytic models are most reliable during the inspiral and RD phases (but have more difficulty tracing the merger portion), we omit in Fig. 11 the transition region of \(-10M \leq (t - t_{\text{peak}}) \leq 10M\). The phase relations during inspiral are determined by Eq. (53) and during ringdown by Eqs. (56a)–(56c). Here we use a \( t_{\text{match}} \) about 20M after \( t_{\text{peak}} \) to ensure that the phase relations are truly dominated by the fundamental QNMs, and thus Eqs. (56a)–(56c) are valid. Note that the phase differences for EQ+ are particularly noisy since the amplitude of the \( f^{33}_m \) moment is zero to leading order, and it is thus more difficult to extract a clear phase for that mode.

The feature that is most difficult to explain from an analytic model alone (and is thus omitted from the eN curves in Fig. 11) is the 180-degree jump in phase between \( F_{22,33}^{\text{insp}} \) and \( F_{33,44}^{\text{insp}} \) around \(-20M \) before the peak. This appears to be a general feature in all the unequal-mass runs, but preliminary results suggest that it is less significant (i.e., a smaller phase shift) for more extreme-mass ratio systems. We are not able to explain it with the additional RD overtone modes described in Sec. VB, but using slightly different RD matching points for the different multipoles may help explain the issue.

**C. The anti-kick**

These flux amplitudes and phase relations can now be used to understand the amplitude of the kick and anti-kick. Throughout the inspiral phase, the amplitude and rotational frequency of the flux vectors in Eqn. (51a) are monotonically increasing, giving the familiar outward-spiraling trajectory for the velocity vector. Then, in the RD phase [Eqs. (54),(55)], the frequency is nearly
FIG. 11: Phase differences between different mode-pair flux vectors, as defined by Eqs. (52a)-(52c). The data refer to the $\text{NE}^{\pm,-}_2$ (upper left panel), $\text{NE}^{\pm,-}_3$ (upper right panel), $\text{NE}^{\pm,0}_2$ (lower left panel), and $\text{EQ}^{+,0}$ (lower right panel) runs. The dashed curves are the eN model predictions of Eqs. (53),(56). We denote with $t_{\text{peak}}$ the time at which $F^{22}$ reaches its maximum.

constant while the mode amplitudes decay exponentially, giving an inward-spiral that decays like a damped oscillator around the final asymptotic recoil velocity. These trajectories in velocity space can be seen in Figs. 12, along with the instantaneous flux vectors from the competing mode-pairs. Clearly, even small changes in the mass ratios and spins orientations of the BHs can give a rather diverse selection of velocity trajectories.

To calculate the recoil velocity, we must integrate the linear momentum flux vectors in time. We can get a reasonable analytic approximation by using Eqs. (51a) and (54) for the inspiral and RD phases, respectively. In the adiabatic inspiral, the complex recoil velocity $v = v_x + iv_y$ can be written as

$$v_{\text{insp}} = \int_{-\infty}^{t_{\text{match}}} F(t') dt' \simeq \frac{1}{i\omega_{\text{match}}} F(t_{\text{match}}),$$

while for the RD portion we have

$$v_{\text{RD}} = \int_{t_{\text{match}}}^{\infty} F(t') dt' = \sum_{l,m,l',m'} \frac{-i}{\sigma_{l'm'0} - \sigma_{lm0}} F^{l,m,l'm'}_{\text{match}},$$

summing over each pair of modes $(lm, l'm')$. Then the total velocity $v$ in each of the dominant mode pairs is given
FIG. 12: The recoil velocity vector evolving in the $v_x$–$v_y$ plane (black solid curve), along with the flux vectors due to the three mode pairs at each time interval (same color scheme as Fig. 10). The data refer to the NE$^{2,3}$ (upper left panel), NE$^{2,3}$ (upper right panel), NE$^{3,0}$ (lower left panel), and EQ$^+$ (lower right panel) runs. We denote with the label peak the time at which $f_2$ reaches its maximum.

The phase $\phi_{\text{match}}$ is defined as the angle made between the flux vector $F^{2,1,22}$ and the integrated velocity vector $v$ at the beginning of the ringdown (with other phases $\phi_{\text{match}}$, $\phi_{\text{match}}$ defined analogously). Because of the anomalous phase shifts and departure from adiabaticity at the transition from inspiral to ringdown, these angles are difficult to predict with an independent analytic model, but can be calculated easily from plots like Fig. 12.

In the case of the NE$^{2,3}$ run, where the recoil is almost entirely dominated by the mode pair $f_2^2 f_3^3$, the ratio of the final velocity to the peak velocity (i.e., the amplitude of the anti-kick) is given by the magnitude of the complex number $|1 - i \omega_{\text{match}} e^{i \phi_{2,22}}/(\sigma_{220} - \sigma_{220})| \approx 0.5$. For the EQ$^+$ run on the other hand, we see that there is no anti-kick, which can be explained by the slowly rotating flux vector that does not spiral back inwards, but rather drifts off slowly towards infinity during the RD. The difference between these two modes can be explained entirely by examining their fundamental QNM frequen-
cies: $\omega_{220} - \omega_{210} = 0.08/M_f$, a much slower frequency than $\omega_{330} - \omega_{220} = 0.31/M_f$. From Eqs. (59a), the final kick should be roughly $|1-i\omega_{\text{match}}e^{i\pi/4}/(\sigma_{210}-\sigma_{220})| \approx 2$ times larger than at the end of inspiral, as seen in Fig. 12. It is somewhat more complicated to calculate the amplitude of the antikick for the other runs, which have multiple dominant flux-pairs, and thus requires the summation of Eqs. (59a)-(59c). For the NE$^{2,3}_{-}$ run we estimate $v_{\text{final}}/v_{\text{insp}} \approx 0.7$ and for the NE$^{2,3}_0$ run $v_{\text{final}}/v_{\text{insp}} \approx 0.85$, in qualitative agreement with the numerical simulations.

In general, we find the relative magnitude of the antikick is mostly dependent on the relative magnitudes of the $S^{21}$ and $I^{33}$ moments. When $S^{21}$ dominates, the ringdown rotation is slow and there is a small unkick, whereas a dominant $I^{33}$ mode gives a rapidly rotating ringdown flux and thus a large unkick. Furthermore, from Eq. (51a), we see that for non-spinning BHs, both the $S^{21}$ and $I^{33}$ modes share the same mass and frequency scaling, so the relative size of the unkick should be roughly independent of mass ratio.

We want to get a more quantitative picture of how these flux vectors add constructively and destructively to give the total recoil velocity to support the analytic estimates presented above. Using $v = \int F dt$ we can write

$$\frac{d}{dt} |v| = \frac{d}{dt} (\hat{v} \cdot v) = \hat{v} \cdot F,$$

where $\hat{v} \cdot v = 1$. Breaking as above $F$ up into the contributions of the dominant modes and then integrating in time gives

$$v^{21,22} = \int \hat{v} \cdot F^{21,22} dt,$$

$$v^{22,33} = \int \hat{v} \cdot F^{22,33} dt,$$

$$v^{33,44} = \int \hat{v} \cdot F^{33,44} dt$$

which add linearly to give the total recoil velocity:

$$|v| = v^{21,22} + v^{22,33} + v^{33,44}.$$  (62)

These different velocities are plotted in Figs. 13, with the same color scheme as in Fig. 12, along with the total integrated recoil velocity in solid black curves. Also shown in Fig. 13 is the velocity $v^{22,33}$ (dashed blue curves), defined analogously to Eq. (61) for the $S^{22}I^{33*}$ flux terms. The remarkably small contribution from this mode pair further justifies our focus on the more dominant pairs of Eq. (23) and Fig. 3.

In the NE$^{2,3}_{-}$ run, where the modal analysis shows the $F^{21,22}$ and $F^{33,44}$ flux terms canceling out, we see that the total integrated recoil velocity (black curves in Fig. 13) is almost entirely dominated by the $F^{22,33}$ flux (red curves). On the other hand, for the NE$^{1,2}_{+}$ run, the $F^{21,22}$ flux is much stronger, adding constructively with the $F^{22,33}$ flux. This has the effect of both increasing the peak velocity and also decreasing the relative strength of the unkick, due to the slow rotation frequency of the $F^{21,22}$ flux during ringdown, as described above. As expected, the NE$^{2,3}_{-}$ run displays behavior intermediate between these two extremes. The EQ run, however, is entirely dominated by the $F^{21,22}$ flux, and thus experiences almost no unkick, but rather drifts off slowly in a nearly constant direction, as seen in the bottom-right panel of Fig. 12.

D. Application to non-planar kicks

One of the most remarkable results from the recent renaissance in numerical relativity was the prediction of extremely large kicks from equal-mass BHs with spins pointing opposite to each other and normal to the orbital angular momentum, producing a recoil out of the orbital plane [13-15]. While this configuration can produce recoils of nearly 4000 km/sec, the analogous non-precessing configuration (EQ run in this paper) gives kicks of only ~ 500 km/sec in the case of maximal spin [10-12]. It turns out that the multipole analysis tools developed above are ideal for understanding and explaining this remarkable difference.

First, we should note that leading-order estimates of the linear momentum flux during inspiral suggest that the discrepancy should be less than a factor of two. For example, Eq. (3.31b) of Kidder [35] gives the spin-orbit contribution to the momentum flux for circular, Keplerian orbits as

$$F_{SO} = \frac{16}{15} \eta^2 \frac{\omega^2}{R^3} [\hat{n} \times \Delta + (\hat{n} \times \hat{\nu})(\hat{\nu} \cdot \Delta)],$$

with $\hat{n}$ and $\hat{\nu}$ being the normalized separation and velocity vectors, respectively. For spins parallel to the angular momentum, the term in brackets has magnitude $\Delta^2$, while for planar spins, it is $2|\Delta| \sin \phi_\Delta$, where $\phi_\Delta$ is the angle between $\Delta$ and $\hat{n}$.

Not surprisingly, we see the exact same scaling from the multipole analysis of Eqs. (19), (20), (49), and one new moment:

$$S_{SO}^{22} = -4i \frac{\sqrt{2\pi}}{5} \eta R \omega^3 e^{-i\phi}(\Delta^2 - i\Delta^\nu).$$  (64)

Combining these equations, we get

$$F_x + iF_y \approx \frac{1}{336\pi} (-14iS^{21}T^{22*}) = \frac{16}{15} \eta^2 R^3 \omega^3 \Delta^2 e^{i\phi}.$$  (65)
FIG. 13: Relative contributions to the integrated recoil velocity from the different multipole mode-pairs. $I^{22}J^{33*}$ (red curve) is the dominant mode for unequal-mass binary systems, while $S^{21}I^{22*}$ (blue curve) dominates for spinning, equal-mass binary systems. Also plotted is the contribution from the $S^{32}J^{33*}$ flux terms (blue dashed curve), demonstrating its very small contribution to the total recoil velocity. The data refer to the NE$^+$ (upper left panel), NE$^-$ (upper right panel), NE$^0$ (lower left panel), and EQ$^+$ (lower right panel) runs. We denote with $t_{\text{peak}}$ the time at which $I^{22}$ reaches its maximum.

So in both paradigms, we see that, when maximizing over $\sin \phi_A$, the planar-spin orientation should result in a recoil twice as large as the parallel-spin case, leaving a factor of roughly 4 difference unexplained.

From Eqs. (65),(66) we see that the only relevant modes involved should be $I^{22}$, $S^{21}$, and $S^{22}$. In the left panel of Fig. 14 we plot the amplitude of $I^{22}$ from the EQ$_{\text{planar}}$ simulation, along with that of a planar-spin simulation EQ$_{\text{planar}}$. All other binary parameters and the initial conditions are the same. Remarkably, the mass quadrupole moments are nearly identical in both runs, suggesting that the overall dynamics of the inspiral and merger are the same. This is in fact quite reasonable since the total spin of the system is zero in both cases. However, we see in the right-hand panel of Fig. 14 that the peak amplitude of the $S^{22}$ mode is a factor of $\sim 2.5$ greater than that of the $S^{21}$ mode from the EQ$_{\text{planar}}$ and EQ$_{\text{planar}}$ runs, respectively.

Yet Eqs. (49), (64) suggest that these two modes should have the same magnitudes, at least during the inspiral phase, and presumably during the RD as well, since the RD amplitudes are completely determined by the mode amplitudes at the matching point. It appears from Fig. 14 that $S^{22}$ and $S^{21}$ do in fact have the same amplitude at early times, but the relatively noisy data and short duration of the simulations make it impossible to say for certain. If this is the case, our best explanation for the sudden remarkable increase in the amplitude of $S^{22}$ is due to mode-mixing with the dominant $I^{22}$ mode, much like that of $S^{32}$ and $I^{22}$ described above in Sec. IVB. This effect is apparently only important between modes with the same $m$-number, and may pos-
FIG. 14: **left panel**: Comparison of the multipole amplitudes $I^{22}$ for the two different equal-mass simulations: EQ\_planar (solid line) and EQ\_+ (dashed line). **right panel**: The $S^{22}$ amplitude from the planar-spins run (EQ\_planar, solid line) and the parallel-spins run (EQ\_+, dashed line).

FIG. 15: **left panel**: Comparison of the linear momentum flux for the two different equal-mass simulations: EQ\_planar (solid line) and EQ\_+ (dashed line). **right panel**: The total recoil velocity from the planar-spins run (EQ\_planar, solid line) and the parallel-spins run (EQ\_+, dashed line).

possibly be explained by the fact that the QNMs are really expressed as spheroidal, not spherical harmonics [65, 66].

Lastly, from the ringdown contribution to the velocity [Eqs. (58),(59)], we can understand another difference between the planar- and parallel-spin orientations. Instead of having two different RD frequencies $\sigma_{210}$ and $\sigma_{220}$ combine to give a slowly rotating flux vector, for the planar-spin case, we have two identical RD frequencies, giving precisely zero rotation to the RD flux. Furthermore, as the spin vector $\Delta$ is precessing faster and faster in a positive direction around the orbital angular momentum vector, even during the inspiral the two modes $I^{22}$ and $S^{22}$ become nearly locked in phase, producing a relatively long-duration burst of linear momentum flux in a single direction during the merger phase. Combined, these effects essentially straighten out the spiral curve in the lower-right panel of Fig. 12, providing another factor of $\sim 1.5$ of increased recoil velocity for planar spins.

In Fig. 15 we show the combination of the above effects. In the left panel, we plot the linear momentum flux from Eqs. (65),(66), showing the factor of two increase predicted by the Kidder formula, along with the factor of 2.5 increase in the amplitude of $S^{22}$ from mode-mixing. In the right panel, we plot the integrated recoil velocity for both runs, which includes the effect of flux rotation during the merger and inspiral phases, accounting for another factor of $\sim 1.6$, giving a total discrepancy of $v(EQ\_planar)/v(EQ\_+) \approx 2.5 \times 2 \times 1.6 = 8$.

**VII. DISCUSSION**

In this paper we analysed several numerical simulations of binary BH coalescence focusing on the physics of
the recoil. We developed tools, based on the multipolar expansion [41–45], that can be used as a diagnostic of the numerical results, and understand how the recoil velocity evolves during the inspiral, merger, and ringdown phases of the coalescence.

We wrote explicit expressions for the linear momentum flux expressed in terms of radiative multipole moments through \( l = 4 \), valid for generic spinning, precessing BH binary systems. We show that these formulae are sufficient to obtain the final recoil with high accuracy. By comparing the amplitudes of the different multipole moments, we found that only three pairs of modes contribute significantly to most of the recoil, notably \( S^{21}I^{22*}, I^{22}I^{33*} \) and \( I^{33}I^{44*} \). Those modes account for the total recoil with an accuracy on the order of \( \sim 5 - 10\% \) throughout the simulations. (see Figs. 8, 9).

The way in which the contribution from these three pairs of modes builds up is not trivial, since not only the relative amplitudes, but especially, the relative phases are also quite important. We find that the relative phases between the three pair modes are nearly constant during the inspiral phase, but start evolving at the onset of the transition from inspiral to RD (see Fig. 11). This late-time evolution can be described reasonably well with analytic formula obtained expressing the mode-pairs in terms of fundamental QNMs of Kerr BHs. We showed that it is the relative magnitude of the current-quadrupole mode \( S^{21} \) and mass-octupole mode \( I^{33} \) that determines the difference between the recoil at the peak of the linear momentum flux, and the final recoil velocity, i.e., the amount of anti-kick.

With the final goal of improving analytic PN models, we also explored whether simple modifications of the Newtonian formula for the linear momentum flux, allow to match the numerical results all along the binary evolution. We found that, if we treat the binary radial separation in the Newtonian multipole modes \( (41a)-(41e) \) with an effective radius, which is computed from the numerical simulations assuming that each multipole mode is described by a dominant frequency (see Fig. 1), the leading Newtonian modes reproduce quite well the numerical ones (see Figs. 6, 7) up to the end of the inspiral phase. We also found, confirming the results in Ref. [66], that a superposition of three QNMs, can fit very well the numerical waveforms beyond the peak of the radiation, all along the RD phase.

The tools developed in this paper will be employed to improve the PN analytic models [27] and its siblings, notably the EOB model [28, 29, 31], so that they can accurately interpolate between numerical results. In this way, fast Monte Carlo simulations will be able to predict recoil distributions from BH mergers with uncertainties smaller than in Ref. [23]. Those recoil distribution can be included in simulations of hierarchical merger models of supermassive BHs providing more robust predictions.

Acknowledgments

A.B. and J.S. acknowledge support from NSF grant PHYS-0603762, and A.B. was also supported by the Alfred Sloan Foundation. We thank Emanuele Berti for providing us with tabulated data for the Kerr QNM frequencies.
