

PAPER • OPEN ACCESS

# Self-force framework for merger-ringdown waveforms

To cite this article: Lorenzo Küchler *et al* 2026 *Class. Quantum Grav.* **43** 015018

View the [article online](#) for updates and enhancements.

## You may also like

- [Can you fall into a McVittie black hole? Will you survive?](#)  
Brien C Nolan
- [Black holes, gravitational waves and fundamental physics: a roadmap](#)  
Abbas Askar, Chris Belczynski, Gianfranco Bertone et al.
- [Parameter estimation of gravitational waves from hyperbolic black hole encounters](#)  
Chad Henshaw, Jacob Lange, Peter Lott et al.

# Classical and Quantum Gravity



## PAPER

# Self-force framework for merger-ringdown waveforms

### OPEN ACCESS

RECEIVED  
17 June 2025

REVISED  
30 October 2025

ACCEPTED FOR PUBLICATION  
10 December 2025

PUBLISHED  
31 December 2025

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Lorenzo Küchler<sup>1,\*</sup> , Geoffrey Compère<sup>2</sup>  and Adam Pound<sup>1</sup> 

<sup>1</sup> School of Mathematical Sciences and STAG Research Centre, University of Southampton, Southampton SO17 1BJ, United Kingdom

<sup>2</sup> Université Libre de Bruxelles, BLU-ULB Brussels Laboratory of the Universe and International Solvay Institutes, C.P. 231, B-1050 Bruxelles, Belgium

\* Author to whom any correspondence should be addressed.

E-mail: [l.m.kuchler@soton.ac.uk](mailto:l.m.kuchler@soton.ac.uk), [geoffrey.compere@ulb.be](mailto:geoffrey.compere@ulb.be) and [A.Pound@soton.ac.uk](mailto:A.Pound@soton.ac.uk)

**Keywords:** gravitational self-force, gravitational waves, quasinormal modes, merger-ringdown waveforms, intermediate- and extreme-mass-ratio inspirals, waveform modelling, black holes

## Abstract

The prospect of observing asymmetric compact binaries with next-generation gravitational-wave detectors has motivated the development of fast and accurate waveform models in gravitational self-force theory. These models are based on a two-stage process: in a (slow) offline stage, waveform ingredients are pre-computed as functions on the orbital phase space; in a (fast) online stage, the waveform is generated by evolving through the phase space. While this framework has traditionally been restricted to the inspiral stage of a binary, we recently extended it across the transition to plunge, where the small companion crosses the innermost stable circular orbit around the primary black hole. In this paper, for the special case of quasicircular, nonspinning binaries, we show how the ‘offline/online’ phase-space paradigm also extends through the final plunge, which generates the binary’s merger-ringdown signal. We implement the method at leading, geodesic order in the plunge. The resulting plunge waveform agrees well with a stationary-phase approximation at early times and with a (self-consistently calculated) quasinormal mode sum at late times, but we highlight that neither of the two approximations reaches the peak of the full plunge waveform. Finally, we compare the plunge waveform to numerical relativity simulations. Our framework offers the prospect of fast, accurate inspiral-merger-ringdown waveform models for asymmetric binaries.

## Contents

1. Introduction	2
2. Merger-ringdown in a phase-space description	5
2.1. Self-force primer	5
2.2. Inspiral	6
2.3. Transition to plunge	7
2.4. Plunge	9
3. Plunge: orbital motion	10
3.1. Expansion at fixed orbital frequency	11
3.1.1. Leading-order match with the late-time transition-to-plunge solution	11
3.1.2. Geodesic order	12
3.1.3. First post-geodesic order	13
3.2. Asymptotic match with the transition to plunge	14
3.3. Expansion at fixed orbital radius	17
3.3.1. Geodesic order	17
3.3.2. First post-geodesic order	18
4. Plunge: field equations at first and second order	19
5. First-order RWZ waveforms	21
5.1. RWZ equations and sources	21
5.2. Punctured RWZ equations	23
5.3. Frequency-domain equations	26
5.4. Inhomogeneous solutions	26
5.5. SPA	29
5.6. QNM sum	29
6. Numerical implementation and comparisons	31
6.1. Implementation and validation	31
6.2. Comparison with the SPA and the QNM sum	35
6.3. Comparison with NR	38
7. Discussion and conclusions	39
Data availability statement	43
Acknowledgment	43
Appendix A. Asymptotic match between transition to plunge and plunge	43
A.1. Coefficients of the asymptotic late-time transition-to-plunge solutions	43
A.2. Coefficients of the asymptotic near-ISCO plunge solutions	45
A.3. Self-force matching conditions	45
Appendix B. Transformation between the fixed- $\Omega$ and fixed- $r_p$ expansions	46
Appendix C. Vector and tensor spherical harmonics	48
Appendix D. First-order Regge–Wheeler–Zerilli (RWZ) equations in generic slicing	48
D.1. Punctured RWZ equations	48
D.2. Frequency-domain equations	49
D.3. Inhomogeneous solutions	50
D.4. Regularity, slicing, and punctures at the boundaries	51
Appendix E. Contribution from the early-time effective source	52
Appendix F. Regularization procedure for the excitation coefficients	53
References	53

## 1. Introduction

Since the first detection of a binary black hole merger in 2015 [1], three observing runs of the LIGO–Virgo–KAGRA Collaboration have seen close to a hundred events [2–4]. The majority of the observed signals originated from the merger of compact binaries with mass ratios typically close to 1, but several binaries have been detected with much more disparate masses. This includes one signal thought to have originated from the merger of a binary with mass ratio  $\approx 1:26$  (GW191219\_16312), outside the range in which current waveform models have been validated [4]. Future observing runs and next-generation ground-based detectors such as the Einstein Telescope [5] and Cosmic Explorer [6] promise an increase in the number of detections [7, 8] and likely also in the number of events with smaller mass ratios. Space-based gravitational-wave (GW) detectors such as the Laser Interferometer Space Antenna (LISA) [9] will be able to detect GWs in the mHz frequency band, allowing us to access a greater variety

of sources such as intermediate- and extreme-mass-ratio inspirals (I/EMRIs) [10, 11]. In addition, LISA will also detect the coalescence of massive black hole binaries, which could exhibit a long tail in their distribution of mass ratios extending to around  $1:10^3$  [12–14]. Intermediate-mass-ratio binaries are also possible sources of multiband signals detectable with both space- and ground-based detectors [15]. The prospect of these future observations motivates the production of fast, accurate inspiral-merger-ringdown (IMR) models for asymmetric binaries, in which the secondary of mass  $m_p$  (labelled with ‘ $p$ ’ because we treat it as a particle) is significantly lighter than the primary of mass  $M$ .

Gravitational self-force (GSF) theory is recognized as the primary method of modelling I/EMRIs [16] and has been used to inform effective models that cover the full binary parameter space [17–20]. Although originally designed for EMRIs, second-order GSF waveforms have proven to be very accurate even at mass ratios  $\approx 1:10$ , with sub-radian dephasing compared to fully nonlinear numerical relativity (NR) simulations [21, 22]. GSF waveforms can also now be generated on a timescale of milliseconds, fast enough for data analysis [23]. However, the methods leading to these GSF waveforms are specialized to slow evolutions; they cannot capture the rapid final stages of a binary merger. Although these final stages contribute negligible signal-to-noise ratio for EMRIs [24], they are critical for more moderate mass ratios and for ground-based detectors (which are most sensitive to the end stages of an asymmetric binary due to their low-frequency floor [25]).

The basic anatomy of a binary evolution consists of three stages: the inspiral, which is well modelled by post-Newtonian (PN) theory [26] (for comparable-mass binaries) or GSF theory [27, 28] (for asymmetric binaries); the merger, which has historically been the realm of fully nonlinear NR simulations [29]; and the post-merger ringdown, which is accurately described by vacuum black hole perturbation theory [30, 31].

Effective-one-body theory (EOB) [32, 33] offered the simplest semi-analytical understanding of this evolution. Rather than a three-stage process, EOB proposed that there are only two stages: an extended inspiral that persists all the way to the light ring of the effective-one-body black-hole metric, followed by a ringdown. This description stems from the reduction of the two-body problem to a (reduced-mass) secondary object orbiting in an effective-one-body black hole spacetime. In that reduced problem, we can apply the following physical intuition. Before the secondary object crosses the effective spacetime’s light ring, the waveform is dominated by radiation propagating directly from the orbiting secondary, and the amplitude of the radiation grows steadily as the secondary moves deeper into the strong-field, relativistic regime. After it crosses the light ring, any radiation the secondary emits falls into the black hole. The waveform is then dominated by radiation that was trapped on the light ring, which then leaks out to future null infinity as quasinormal modes (QNMs).

To meet accuracy requirements for GW science, this simple description has had to be corrected by non-quasicircular corrections (NQC) [34–38] and other phenomenological adaptations calibrated to NR waveforms. However, the core idea proved remarkably accurate.

Complementary to EOB, the Phenom family of IMR models [39–42] takes advantage of another feature of the merger. The amplitude and phase of the waveform through merger are surprisingly simple and can be very well approximated by elementary functions. Directly approximating the amplitude and phase observed in NR allows Phenom models to avoid stitching together an inspiral to a ringdown. Instead, an inspiral waveform based on PN and tuned to EOB is effectively stitched (in an NR-calibrated way) to an NR-informed merger-ringdown waveform.

Both these models leverage the fact that during the inspiral, the waveform can be written as a function of the two-body, mechanical phase-space variables—the bodies’ relative positions and momenta, for example. GSF waveform models double down on this, using it to formulate a multiscale expansion of the Einstein field equations that puts the field equations in a practical form and simultaneously enables rapid waveform generation [28, 43, 44]. In the multiscale expansion, waveform generation is divided into offline and online calculations. The offline calculations consist of solving the field equations to precompute waveform ingredients as functions on the binary’s mechanical phase space. The online stage then consists of a rapid, inexpensive evolution through the phase space, together with a summation of waveform modes [23].

This GSF framework for modelling the inspiral has traditionally relied on a separation of time scales: the system’s parameters (e.g. its fundamental frequencies) evolve on a radiation-reaction time scale that is large compared to the periods of orbital motion. During the inspiral, the radiation-reaction time scale is of order  $M^2/m_p$ , and the multiscale expansion takes the form of a post-adiabatic expansion [28, 43–45]. This scaling breaks down as the secondary approaches the innermost stable circular orbit (ISCO), or more generally the separatrix between stable and plunging orbits [46], where the post-adiabatic expansion becomes singular. However, there is still a separation of time scales, as the ISCO-crossing time is long, of order  $M(M/m_p)^{1/5}$ , compared to the orbital period [33, 47]. In [48–51], we exploited this

fact to develop a multiscale expansion adapted to the transition across the ISCO. In [51], in particular, we showed how the phase-space paradigm persists in this regime and how it continues to facilitate an offline-online split of the field equations, ultimately enabling rapid waveform generation. (EOB models, in contrast, do not fully exploit the separation of scales, which allows them to smoothly evolve across the ISCO but prevents a complete offline-online split.)

In this paper, we continue this development by building a merger-ringdown GSF framework within the phase-space paradigm, restricting to the case of quasicircular, nonspinning binaries (as we also did for the transition to plunge). There is no separation of time scales during the final plunge, when the secondary falls from the ISCO down to the black hole horizon. Nevertheless, we show that one can continue to treat the waveform as a function on the binary’s mechanical phase space, linking the waveform to the orbital dynamics all the way into the infinite future. Unlike in EOB, where the connection to the orbit is lost at the light ring, we obtain the entire merger-ringdown waveform without having to switch to a separate ringdown approximation. Like in the inspiral and transition to plunge, this enables an offline-online split in which we can pre-compute waveform ingredients as functions on the orbital phase space.

Our phase-space formulation differs in two ways from previous approaches to modelling the merger-ringdown at leading order within GSF theory [52–56]: first, our approach maintains the rapid waveform generation framework of the inspiral; second, it can be systematically applied beyond leading order. We ultimately aim to combine our treatment of the plunge with the treatments of the inspiral and transition to plunge in order to construct a model that can seamlessly and rapidly generate complete IMR waveforms for asymmetric binaries. In the present paper, we limit our ambitions to three goals: (i) developing an appropriate ‘post-geodesic’ expansion of the motion and field equations in the plunge regime; (ii) implementing the method to generate plunge waveforms at leading, geodesic order; (iii) exploring this leading-order waveform’s features and accuracy.

For our formulation of the post-geodesic expansion in the plunge, we emphasize how the early-time behaviour of this expansion, when the secondary is near the ISCO, appropriately matches the late-time behaviour of our transition-to-plunge expansion. We then make use of this matching in our implementation of the leading, geodesic-order plunge. We compute plunge waveforms at that order using a Fourier transform adapted to the phase-space representation of the problem. Our calculations then closely follow earlier ones in [57, 58]. However, we go beyond those calculations by placing them within a framework that can be applied at higher orders and by using the matching to the transition to plunge to more rigorously justify various steps.

With the geodesic plunge waveform in hand, we then explore how well it can be separated into two distinct segments: an extended inspiral and a ringdown, as in EOB. Our final waveform is given by an inverse Fourier transform, which presents two clear approximations. Before the waveform’s peak amplitude, we can approximate the inverse Fourier transform by a stationary-phase approximation (SPA), in which the waveform frequency becomes equal to an integer multiple of the orbital frequency; this is in the spirit of EOB’s waveform generation in its extended inspiral. The inverse Fourier transform can also be written in terms of a sum of QNMs, power-law tails, and prompt response, and we are able to internally compute the excitation coefficients of the QNMs. In line with EOB’s basic description of merger, we find that the SPA works remarkably well until near the waveform’s peak, and we find good agreement with the QNM sum after the peak. However, neither approximation is accurate at the waveform’s peak, and stitching the two together does not reproduce the full plunge waveform. It is possible that inclusion of higher-order terms in the SPA will bring it closer to the peak. On the other hand, the QNM sum clearly breaks down near the peak and can only be sensibly used at times  $\gtrsim 10M$  after the peak, consistent with studies of numerical fits to numerical merger-ringdown waveforms [59–62].

The paper is structured as follows. Section 2 contains the phase-space description of our merger-ringdown model. In section 3 we present the post-geodesic expansion of the orbital motion and asymptotically match it with the late-time transition-to-plunge solution of [51]. We present the plunge field equations through second order in the mass ratio in section 4. We then proceed to solve the Regge–Wheeler–Zerilli (RWZ) equations to obtain first-order plunge waveforms in section 5. This section also explains the SPA to the plunge waveform and the construction of the QNM sum. In section 6 we present our numerical implementation, perform internal consistency checks and compare our first-order waveforms to NR simulations. Finally, we discuss our results and future directions in section 7.

We work in geometric units where the gravitational constant  $G$  and the speed of light  $c$  are set to unity,  $G = c = 1$ . The small mass ratio is defined as  $\varepsilon := m_p/M$ . It will also be useful to introduce the large mass ratio  $q := 1/\varepsilon$ .

## 2. Merger-ringdown in a phase-space description

Before detailing our method, in this section we explain how the merger-ringdown regime can be described within the same phase-space paradigm as the inspiral and transition to plunge. For concreteness, we specialize immediately to the case of a nonspinning particle orbiting a slowly spinning primary black hole.

### 2.1. Self-force primer

We start from the equations of second-order self-force theory in the self-consistent approach [63–66], which are valid on all timescales. These equations will then be expanded on the relevant timescales of the inspiral, transition to plunge, and plunge.

The spacetime metric is split into a background plus a perturbation,  $g_{\alpha\beta} + h_{\alpha\beta}$ , where  $g_{\alpha\beta}$  is the Schwarzschild metric of the primary black hole as if it were isolated, with constant mass  $M$ . The perturbation  $h_{\alpha\beta} \sim \varepsilon$ , due to the presence of the orbiting secondary, encodes all corrections sourced by the secondary as well as all nonlinear effects of the two bodies' gravitational interaction, such as the primary's slow accumulation of mass and spin due to absorption of radiation.  $h_{\alpha\beta}$  will be expanded in a series for small  $\varepsilon$ , but the form of that series will depend on the stage of the binary's evolution.

Without specifying the form of that expansion, we can write the Einstein equations perturbatively in  $h_{\alpha\beta}$  as

$$\delta G_{\alpha\beta}[h] + \delta^2 G_{\alpha\beta}[h, h] + \mathcal{O}(\varepsilon^3) = 8\pi T_{\alpha\beta}, \quad (2.1)$$

where  $\delta G_{\alpha\beta}$  is the linearized Einstein tensor on the background  $g_{\alpha\beta}$ ,  $\delta^2 G_{\alpha\beta}$  is quadratic in  $h_{\alpha\beta}$ ,<sup>3</sup> and so on. On the right-hand side, the secondary is represented by the point-particle Detweiler stress-energy tensor [66, 67],

$$T_{\alpha\beta} = m_p \int \tilde{u}_\alpha \tilde{u}_\beta \frac{\delta^4(x^\mu - z^\mu(\tilde{\tau}))}{\sqrt{-\tilde{g}}} d\tilde{\tau}, \quad (2.2)$$

where  $z^\mu(\tilde{\tau})$  is the particle's orbital trajectory.  $\tilde{g}_{\alpha\beta} = g_{\alpha\beta} + h_{\alpha\beta}^{\mathcal{R}}$  is a certain *effective* metric, which is regular at the particle's position, and in which the particle moves as a test mass. Here we have split the metric perturbation into  $h_{\alpha\beta} = h_{\alpha\beta}^{\mathcal{P}} + h_{\alpha\beta}^{\mathcal{R}}$ , where  $h_{\alpha\beta}^{\mathcal{P}}$  is an analytically known 'puncture', which is singular at the particle's position, and  $h_{\alpha\beta}^{\mathcal{R}}$  is the regular residual field, as defined in [68]. The parameter  $\tilde{\tau}$  is proper time in  $\tilde{g}_{\alpha\beta}$ , and the particle's 4-velocity is  $\tilde{u}^\alpha := dz^\alpha/d\tilde{\tau}$  (with  $\tilde{u}_\alpha := \tilde{g}_{\alpha\beta}\tilde{u}^\beta$ ).

While the particle obeys the geodesic equation in the effective metric  $\tilde{g}_{\alpha\beta}$ , its motion is accelerated in the background metric  $g_{\alpha\beta}$ , such that

$$\frac{D^2 z^\mu}{d\tau^2} = f^\mu. \quad (2.3)$$

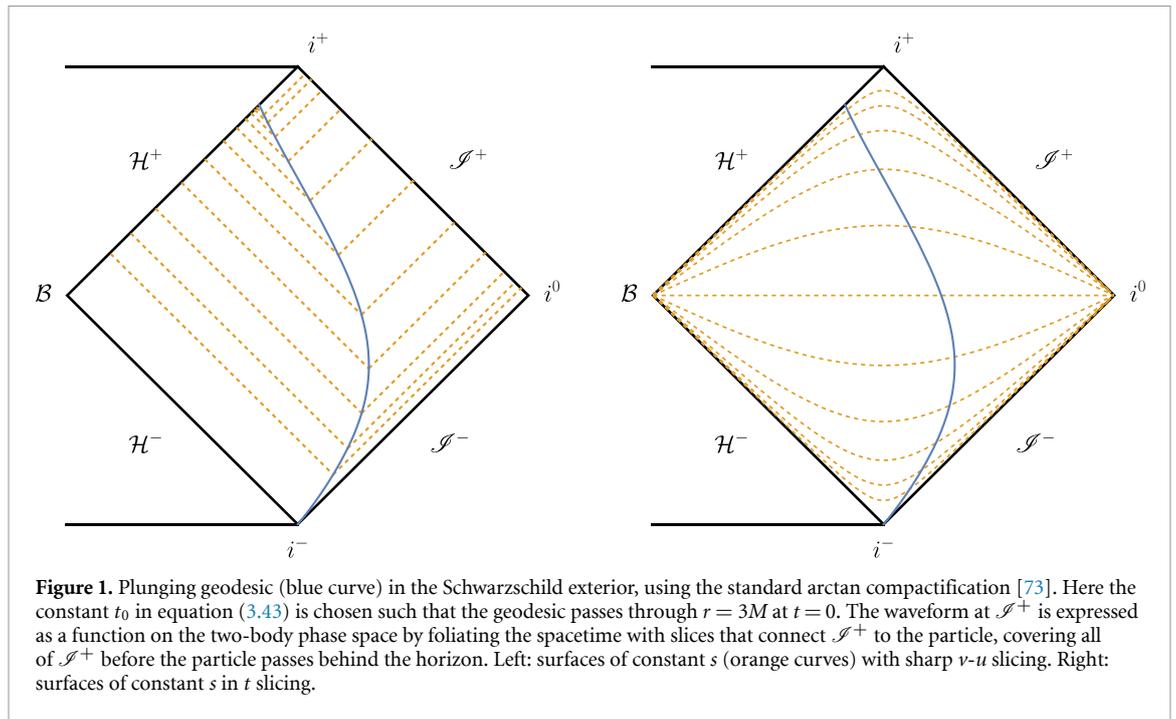
Here  $\tau$  is proper time in  $g_{\alpha\beta}$  and  $D/d\tau := u^\alpha \nabla_\alpha$ , with the 4-velocity  $u^\alpha := dz^\alpha/d\tau$  and covariant derivative  $\nabla_\alpha$  compatible with  $g_{\alpha\beta}$ . Explicitly, the self-force (per unit mass) acting on the secondary is given by [63, 65]

$$f^\mu = -\frac{1}{2} P^{\mu\nu} (\delta_\nu^\rho - h_\nu^{\mathcal{R}\rho}) (2\nabla_\alpha h_{\beta\rho}^{\mathcal{R}} - \nabla_\rho h_{\alpha\beta}^{\mathcal{R}}) u^\alpha u^\beta + \mathcal{O}(\varepsilon^3), \quad P^{\mu\nu} := g^{\mu\nu} + u^\mu u^\nu. \quad (2.4)$$

Just as the particle's orbit evolves self-consistently in response to the metric perturbation, the primary's mass and spin evolve due to the GW fluxes of energy and angular momentum through its horizon. In order to build a consistent perturbative expansion, we need to take into account this dynamical change. We write the black hole's total mass as  $M + \varepsilon \delta M$  and total spin as  $\varepsilon \delta J$ , where  $M$  is the constant mass of the Schwarzschild background and  $\delta M(\nu, \varepsilon)$  and  $\delta J(\nu, \varepsilon)$  are the evolving corrections (normalized by  $\varepsilon$ ), which are functions of advanced time  $\nu$  along the horizon. These perturbations appear in  $h_{\alpha\beta}$ , rather than in the background, following the self-consistent prescription of [43, 69]. Adopting the conventions of [51], we define  $\delta M := \delta M^+$  and  $\delta J := \delta M^-$  and collectively denote them as  $\delta M^\pm$ .

To describe the three stages of binary evolution, we now foliate the spacetime with surfaces of constant time  $s$ . As explained in detail in [43, 70],  $s$  is most conveniently chosen to be a hyperboloidal time (or quasi-hyperboloidal, allowing null segments), such that each  $s = \text{constant}$  slice extends from the

<sup>3</sup> Due to its strong singularity at the particle,  $\delta^2 G_{\alpha\beta}$  is not uniquely defined on a domain that includes the particle's worldline. We assign it the distributional definition of [66], which is consistent with our use of the Detweiler stress-energy tensor.



black hole’s future horizon  $\mathcal{H}^+$  to future null infinity  $\mathcal{S}^+$ ; this avoids some of the spurious divergences that arise at subleading orders in  $\varepsilon$  if using Schwarzschild time  $t$  [71, 72]. Here we leave the choice of time generic, only restricting it to the form  $s = t - \kappa(x)$ , where  $t$  is Schwarzschild time and  $x$  is the tortoise coordinate. In the present section, for simplicity we also assume  $s = t$  at the particle’s worldline, but in sections 4 and 5.1 we will allow for more general choices. Figure 1 illustrates the two most extreme examples: ‘sharp’ slicing in which  $s = v$  to the left of the particle and  $s = u$  to the right; and ‘flat’ slicing  $s = t$  everywhere. On each slice, we use Schwarzschild coordinates  $x^i = (r, \theta, \phi)$ . We can then conveniently use  $t$  as a parameter along the particle’s worldline, such that  $z^\mu(t, \varepsilon) = (t, x_p^i(t, \varepsilon))$ , with the spatial trajectory

$$x_p^i(t, \varepsilon) = (r_p(t, \varepsilon), \pi/2, \phi_p(t, \varepsilon)). \tag{2.5}$$

We refer to the space spanned by  $(x_p^i, \dot{x}_p^i, \delta M^\pm)$  as the binary’s phase space, where an overdot denotes  $d/dt$ . A particular binary system evolves through this space, and, via the Einstein field equations, the phase-space evolution is linked to the metric’s evolution on spacetime. The particular form of the evolution, and the congruous form of the metric perturbation, depends on the regime: inspiral, transition to plunge, or plunge.

### 2.2. Inspiral

During the binary’s quasicircular inspiral stage, the particle’s orbit is characterized by a single slowly evolving orbital frequency,

$$\frac{d\phi_p}{dt} = \Omega, \tag{2.6}$$

and a slowly evolving orbital radius. The mechanical variables describing the binary are then the constants  $M$  and  $m_p$ , the orbital phase  $\phi_p$ , which varies on a ‘fast’ orbital time scale  $\sim 2\pi/\Omega \sim M$ , and the evolving parameters  $J^a = (\Omega, \delta M^\pm)$ , which evolve on the ‘slow’ radiation-reaction time scale  $\sim M/\varepsilon$ . The orbital radius, rather than being independent, is expressed in terms of the other variables through an expansion of the form  $r_p = r_{(0)}(\Omega) + \varepsilon r_{(1)}(J^a) + \mathcal{O}(\varepsilon^2)$ ,<sup>4</sup> where  $r_{(0)}(\Omega) = M(M\Omega)^{-2/3}$  is the geodesic relationship, and the correction  $r_{(1)}$  is due to the radial self-force.

We define the mechanical variables  $(\phi_p, J^a)$  as functions on spacetime by making them constant on slices of constant global time  $s$ , recalling that  $s = t$  on the particle’s worldline. Their evolution from slice

<sup>4</sup> Note that here and throughout this paper, we suppress functional dependence on the background mass  $M$ .

to slice is then governed by simple differential equations of the form [43, 70]

$$\frac{d\phi_p}{ds} = \Omega, \quad (2.7a)$$

$$\frac{d\Omega}{ds} = \varepsilon \left[ F_{(0)}^\Omega(\Omega) + \varepsilon F_{(1)}^\Omega(J_a) + \mathcal{O}(\varepsilon^2) \right], \quad (2.7b)$$

$$\frac{d}{ds} \delta M^\pm = \varepsilon F_{(1)}^\pm(\Omega) + \mathcal{O}(\varepsilon^2). \quad (2.7c)$$

Here numeric labels ( $n$ ) denote the post-adiabatic order ( $n$ PA) at which a term contributes to the orbital phase. The forcing terms  $F_{(n)}^\Omega$  are obtained from the self-force using the equation of motion (2.3), while  $F_{(n)}^\pm$  are determined from the fluxes of energy and angular momentum through the horizon.

The fundamental feature of our phase-space approach is that the metric perturbation only depends on  $s$  through a dependence on the evolving mechanical variables, such that

$$h_{\alpha\beta}(s, x^i, \varepsilon) = h_{\alpha\beta}(\phi_p(s, \varepsilon), J^a(s, \varepsilon), x^i, \varepsilon). \quad (2.8)$$

In most treatments [28, 43, 44, 51, 70], this form is treated as an ansatz, which is justified *a posteriori* by the fact that it leads to a solution to the Einstein equations with the desired physical properties, satisfying the desired boundary condition, and with uniform, controlled small- $\varepsilon$  accuracy in each of the regimes of interest. More recently, [69] showed that the ansatz in the inspiral regime can be derived from the more general self-consistent formulation [74] of GSF theory. It is then inherited, through matching conditions, in the transition regime and finally in the plunge. However, for practical purposes we can continue to treat it simply as an ansatz.

We can then treat the perturbation as a function on phase space,  $h_{\alpha\beta}(\phi_p, J^a, x^i, \varepsilon)$ . In the inspiral regime, we expand this function at fixed values of the phase-space coordinates, such that

$$h_{\alpha\beta}(\phi_p, J^a, x^i, \varepsilon) = \sum_{n \geq 1} \varepsilon^n h_{\alpha\beta}^{(n)}(\phi_p, J^a, x^i). \quad (2.9)$$

Moreover, since  $\phi_p$  is a  $2\pi$ -periodic variable, assuming continuity on phase space allows us to expand the metric perturbations in discrete Fourier series:

$$h_{\alpha\beta}(\phi_p, J^a, x^i, \varepsilon) = \sum_{n \geq 1} \varepsilon^n \sum_{m=-\infty}^{\infty} h_{\alpha\beta}^{(n),m}(J^a, x^i) e^{-im\phi_p}. \quad (2.10)$$

For the metric perturbations, a label ( $n$ ) indicates their absolute order in  $\varepsilon$ .

Concrete calculations are performed by rewriting the Einstein equations as equations on phase space, as we review in section 4 below. The leading-order field equations, for  $h_{\alpha\beta}^{(1),m}$ , are identical to traditional frequency-domain field equations for a particle on a precisely circular geodesic of frequency  $\Omega$ , even though  $\phi_p$  and  $J^a$  are never ascribed the time dependence they would have for a geodesic. The forcing functions  $F_{(n)}^a = (F_{(n)}^\Omega, F_{(n)}^\pm)$  in equations (2.7b) and (2.7c) are then calculated from the amplitudes  $h_{\alpha\beta}^{(n),m}$ .

In this approach, the asymptotic waveform is obtained as a function on phase space simply by taking the  $r \rightarrow \infty$  limit of equation (2.10). Rapid waveform generation is made possible by the fact that the waveform's inputs (the amplitudes  $h_{\alpha\beta}^{(n),m}$  and forcing functions  $F_{(n)}^\Omega$  and  $F_{(n)}^\pm$ ) are all pre-computed in advance as functions on phase space, prior to specifying the value of  $\varepsilon$  or any particular trajectory through phase space. Given the pre-computed inputs, waveforms are then generated by solving the ordinary differential equations (2.7) to determine the waveform's time dependence.

### 2.3. Transition to plunge

When the particle approaches the ISCO, which lies at a radius  $r_* := 6M$  or frequency  $\Omega_* := 1/(6^{3/2}M)$ ,<sup>5</sup> the frequency begins to evolve more rapidly, and the assumption that  $d\Omega/dt \sim \varepsilon$  breaks down. The particle then transitions across the ISCO, on a time scale  $\sim M/\varepsilon^{1/5}$  and over a frequency band of size  $\sim \varepsilon^{2/5}$ .

<sup>5</sup> Note that we chose to use  $x$  rather than  $r^*$  to denote the tortoise coordinate in order to avoid confusion with the numerical constant  $r_*$ .

**Table 1.** Labelling and numbering conventions for the forcing functions and metric perturbations in the inspiral, transition-to-plunge and plunge expansions.

	Forcing functions	Metric perturbations
Inspiral	$F_{(n)}^\Omega, n = 0, 1, \dots$ $F_{(n)}^\pm, n = 1, 2, \dots$	$h_{\alpha\beta}^{(n)}, n = 1, 2, \dots$
Transition to plunge	$F_{[n]}^{\Delta\Omega}, n = 0, 1, \dots$ $F_{[n]}^\pm, n = 3, 4, \dots$	$h_{\alpha\beta}^{[n]}, n = 5, 6, \dots$
Plunge	$F_{\{n\}}^\Omega, n = 0, 1, \dots$ $F_{\{n\}}^\pm, n = 1, 2, \dots$	$h_{\alpha\beta}^{\{n\}}, n = 1, 2, \dots$

We describe this transition to plunge in essentially the same way as we did the inspiral. In place of  $\Omega$ , we adopt a scaled frequency

$$\Delta\Omega := \frac{\Omega - \Omega_*}{\varepsilon^{2/5}}, \quad (2.11)$$

which is of order unity in the transition regime. We then work with mechanical variables  $\phi_p$  and  $\Delta J^a = (\Delta\Omega, \delta M^\pm)$ , and equations (2.7) are replaced by evolution equations of the form [51]

$$\frac{d\phi_p}{ds} = \Omega = \Omega_* + \varepsilon^{2/5} \Delta\Omega, \quad (2.12a)$$

$$\frac{d\Delta\Omega}{ds} = \varepsilon^{1/5} \left[ F_{[0]}^{\Delta\Omega}(\Delta\Omega) + \varepsilon^{2/5} F_{[2]}^{\Delta\Omega}(\Delta\Omega) + \varepsilon^{3/5} F_{[3]}^{\Delta\Omega}(\Delta J^a) + \mathcal{O}(\varepsilon^{4/5}) \right], \quad (2.12b)$$

$$\frac{d}{ds} \delta M^\pm = \varepsilon F_{[3]}^\pm(\Delta\Omega) + \mathcal{O}(\varepsilon^{6/5}). \quad (2.12c)$$

We use numeric labels  $[n]$ , rather than  $(n)$ , to denote the post-leading-transition order ( $n$ PLT) at which a term contributes to the orbital phase<sup>6</sup>. Here, functions on phase space admit series expansions in powers of the small parameter  $\varepsilon^{1/5}$  rather than in powers of  $\varepsilon$ . This behaviour, as well as the  $\varepsilon^{2/5}$  scaling of  $\Delta\Omega$  in equation (2.11), is readily derived from the requirement that the transition-to-plunge expansion asymptotically matches the expansions in the inspiral and plunge. For the ease of the reader, we summarize the labelling and numbering conventions on table 1.

Although the frequency  $\Omega$  now evolves at a rate  $d\Omega/ds = \varepsilon^{2/5}(d\Delta\Omega/ds) \sim \varepsilon^{3/5}$ , the rate of change of  $\delta M^\pm$  remains  $\mathcal{O}(\varepsilon)$ . This is true in all regimes, including even the plunge, because the fluxes of energy and angular momentum across the horizon are always  $\mathcal{O}(\varepsilon^2)$ , as they are proportional to the square of the time derivative of the metric perturbation (and we recall that  $\delta M^\pm$  are normalized by  $\varepsilon$ ).

Just as for the inspiral, we expand the metric perturbation for small  $\varepsilon$  at fixed values of our phase-space variables:

$$h_{\alpha\beta} = \sum_{m=-\infty}^{\infty} \left[ \varepsilon h_{\alpha\beta}^{[5],m}(\Delta J^a, x^i) + \varepsilon^{7/5} h_{\alpha\beta}^{[7],m}(\Delta J^a, x^i) + \varepsilon^{8/5} h_{\alpha\beta}^{[8],m}(\Delta J^a, x^i) + \mathcal{O}(\varepsilon^{9/5}) \right] e^{-im\phi_p}. \quad (2.13)$$

Similarly to the inspiral, for the metric perturbations a label  $[n]$  indicates the absolute order of  $\varepsilon^{1/5}$ . Here the dependence on  $\varepsilon^{1/5}$  is inherited from the orbital evolution through the field equations. Like in the inspiral, we express the field equations as equations on phase space, allowing us to directly compute the metric amplitudes  $h_{\alpha\beta}^{[n],m}(\Delta J^a, x^i)$  and forcing functions  $F_{[n]}^a(\Delta J^b)$  as functions of  $\Delta J^a$ . Again, we review this method in section 4; and again, the forcing functions  $F_{[n]}^a(\Delta J^b)$  can be computed from the amplitudes  $h_{\alpha\beta}^{[n],m}$ .

Since we are able to precompute all the waveform ingredients, we maintain the rapid waveform-generation framework of the inspiral. Given the waveform amplitudes and forcing functions, we rapidly generate waveforms by solving the orbital evolution equations (2.12).

<sup>6</sup> The quantity  $F_{[n]}^{\delta M^\pm}$  defined in [75] is now denoted as  $F_{[n+3]}^\pm$ .

## 2.4. Plunge

Once the particle has fallen sufficiently far below the ISCO, after a time  $\gtrsim M/\varepsilon^{1/5}$ , the transition-to-plunge approximation breaks down and the orbital frequency begins to evolve on a time scale  $\sim M$ . This occurs as the steepness of the radial potential begins to drive the inward motion, dominating over radiation-reaction effects. The secondary then plunges following a nearly geodesic orbit in the primary's geometry with almost constant orbital energy and angular momentum [47].

One might intuitively doubt that the phase-space picture can be extended to this final plunge regime. That intuition stems from thinking that the ringdown portion of the waveform occurs in some sense 'after' the particle has fallen behind the black hole horizon. However, the meaning of 'after' entirely depends on one's choice of spacetime foliation. From the perspective of an asymptotic observer, the particle *never* falls behind the horizon; the final ringdown does not occur 'after' the particle's passage through the horizon, since that passage is not in the causal past of  $\mathcal{I}^+$ . Using either our sharp or flat slicings, we can foliate the entire black hole exterior with constant- $s$  slices, and the particle's trajectory intersects every such slice, never crossing the horizon at any time  $s < \infty$ . See again figure 1 for an illustration of this. As a consequence, the entire merger and ringdown can be parameterized by the binary's trajectory through phase space.

In the bulk of the paper, we show how to apply this idea to develop an appropriate expansion for the plunge. The fundamental difference between the plunge and the other regimes is that the orbital frequency now evolves on the same time scale  $\sim M$  as the orbital phase. Hence, instead of equations of the form (2.7) or (2.12), in the plunge we have equations of the form

$$\frac{d\phi_p}{ds} = \Omega, \quad (2.14a)$$

$$\frac{d\Omega}{ds} = F_{\{0\}}^\Omega(\Omega) + \varepsilon F_{\{1\}}^\Omega(J^a) + \mathcal{O}(\varepsilon^2), \quad (2.14b)$$

$$\frac{d}{ds}\delta M^\pm = \varepsilon F_{\{1\}}^\pm(\Omega) + \mathcal{O}(\varepsilon^2), \quad (2.14c)$$

where we have reverted to phase-space coordinates  $J^a = (\Omega, \delta M^\pm)$ , and we use labels  $\{n\}$  to denote post-geodesic orders ( $n$ PG). The metric perturbation, still treated as a function on phase space, has a corresponding expansion

$$h_{\alpha\beta} = \sum_{m=-\infty}^{\infty} \left[ \varepsilon h_{\alpha\beta}^{\{1\},m}(J^a, x^i) + \varepsilon^2 h_{\alpha\beta}^{\{2\},m}(J^a, x^i) + \mathcal{O}(\varepsilon^3) \right] e^{-im\phi_p}. \quad (2.15)$$

The label  $\{n\}$  indicates the metric perturbation's absolute order in  $\varepsilon$ . Just as for the inspiral and transition to plunge, we will rewrite the field equations as equations for the functions  $h_{\alpha\beta}^{\{n\},m}(J^a, x^i)$ .

This formulation of merger and ringdown markedly contrasts with EOB's. As alluded to in the Introduction, in the EOB description, the waveform's link to the binary phase space is broken when the effective particle passes the light ring of the effective-one-body metric; after that point, a phenomenological ringdown waveform is attached. By maintaining the waveform's link to the phase-space trajectory, we maintain the structure that enables fast waveform generation: all waveform ingredients are pre-computed by solving field equations on phase space, and waveforms are then rapidly generated by solving the evolution equations (2.14).

We also emphasize that our treatment of the plunge fundamentally differs from an expansion around a geodesic solution, even though we refer to our expansion as 'post-geodesic' for lack of a better name. An expansion around a geodesic solution would involve an expansion of the phase-space trajectory itself, as in  $\phi_p(t, \varepsilon) = \phi_0(t) + \varepsilon \phi_1(t) + \mathcal{O}(\varepsilon^2)$ ,  $\Omega(t, \varepsilon) = \Omega_0(t) + \varepsilon \Omega_1(t) + \mathcal{O}(\varepsilon^2)$ , and  $\delta M^\pm(t, \varepsilon) = \delta M_0^\pm(t) + \mathcal{O}(\varepsilon)$ , where a subscript 0 denotes a geodesic solution. Correspondingly, the metric perturbation would be expanded as  $h_{\alpha\beta} = \varepsilon h_{\alpha\beta}^{\{1\}}(s, x^i) + \varepsilon^2 h_{\alpha\beta}^{\{2\}}(s, x^i) + \mathcal{O}(\varepsilon^3)$ , where  $h_{\alpha\beta}^{\{1\}}$  would be sourced by the particle on the geodesic trajectory. This would be a 'Gralla-Wald' formulation [76, 77], which is an expansion in powers of  $\varepsilon$  at fixed spacetime coordinates  $(s, x^i)$ ; in our approach, we instead expand all quantities, including  $\dot{\Omega}$  and  $\delta \dot{M}^\pm$ , at fixed values of  $(\phi_p, J^a, x^i)$ . Our leading perturbation  $h_{\mu\nu}^{\{1\}}$  is a function of  $(\phi_p, \Omega, \delta M^\pm)$ , which becomes a function of time when evaluated along phase-space trajectories obtained by solving equations (2.14). Those phase-space trajectories are only geodesic trajectories when  $F_{\{1\}}^a$  and higher terms are omitted.

Historically, a Gralla-Wald expansion has not been used in the inspiral because it is only valid on time scales much shorter than the dephasing time  $\sim M/\sqrt{\varepsilon}$  over which the accelerated trajectory dephases significantly from the geodesic trajectory [71]. This restriction would not in itself be problematic

in the transition or plunge, which occur on the time scales  $\sim M/\varepsilon^{1/5}$  and  $\sim M$ , respectively. However, a Gralla-Wald expansion would not allow us to assemble the inspiral, transition to plunge and plunge in an efficient asymptotically matched expansion scheme.

By treating all three regimes in the same manner, we put the problem in a uniform offline/online format. Ultimately, one can hybridize the waveforms between the inspiral, transition-to-plunge, and plunge regimes at the level of the precomputed forcing functions  $F_{(n)}^a$ ,  $F_{[n]}^a$ , and  $F_{\{n\}}^a$  and the amplitudes  $h_{\alpha\beta}^{(n),m}$ ,  $h_{\alpha\beta}^{[n],m}$ ,  $h_{\alpha\beta}^{\{n\},m}$  (with  $r \rightarrow \infty$ ), such that the full IMR waveform can be generated from pre-computed ingredients by solving a single set of three ordinary differential equations for  $\phi_p(s, \varepsilon)$ ,  $\Omega(s, \varepsilon)$ , and  $\delta M^\pm(s, \varepsilon)$ .

In this paper, we defer the presentation of such a hybridization, focusing on the plunge regime and its asymptotic matching to the transition-to-plunge regime.

### 3. Plunge: orbital motion

We now develop the equations (2.14a) and (2.14b) describing the plunging orbit. We work in Schwarzschild coordinates  $(t, r, \theta, \phi)$ , such that the background metric reads

$$ds^2 = -f(r) dt^2 + \frac{dr^2}{f(r)} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2), \quad f(r) := 1 - \frac{2M}{r}, \quad (3.1)$$

and the particle's spatial trajectory  $x_p^i(t, \varepsilon)$  is as in equation (2.5).

We present two different formulations of the plunge expansion, based on two different choices of phase-space coordinates  $(\phi_p, J^a)$ . In section 3.1, we choose  $J^a = (\Omega, \delta M^\pm)$  as in section 2.4. This choice exhibits potentially problematic features because  $\Omega$  does not provide a single coordinate chart over the whole plunge; it increases to a maximum at the light ring and then decreases to zero at the horizon, such that  $(\phi_p, \Omega, \delta M^\pm)$  actually represents two distinct coordinate patches (completely analogous to the 'isofrequency' orbits described in [78]). In section 3.3 instead we make a choice that avoids this issue:  $J^a = (r_p, \delta M^\pm)$ , where  $r_p$  is the orbital radius. Since the two formulations only differ in their choice of phase-space coordinates, they are formally equivalent. In appendix B we derive the transformation between them.

In either approach, the variables  $J^a$  satisfy evolution equations of the form

$$\frac{dJ^a}{dt} = F^a(J^b, \varepsilon) = F^a_{\{0\}}(J^b) + \varepsilon F^a_{\{1\}}(J^b) + \mathcal{O}(\varepsilon^2). \quad (3.2)$$

If  $J^a = (\Omega, \delta M^\pm)$ , then the functions  $F^a_{\{n\}}$  are the only needed orbital input in the online waveform generation described in the previous section. If  $J^a = (r_p, \delta M^\pm)$ , then  $\Omega$  in equation (2.14a) becomes  $\Omega(r_p, \varepsilon) = \Omega_{\{0\}}(r_p) + \varepsilon \Omega_{\{1\}}(J^a) + \mathcal{O}(\varepsilon^2)$ , and the functions  $\Omega_{\{n\}}$  become necessary inputs in the waveform generation.

To derive the forcing functions  $F^a_{\{n\}}$  and the frequency corrections  $\Omega_{\{n\}}$ , we rewrite the orbital equation of motion (2.3) in terms of  $t$ . Defining the redshift  $U := dt/d\tau$ , we can write the four-velocity  $u^\mu := dz^\mu/d\tau$  as

$$u^\mu(J^a, \varepsilon) = U \left( 1, \frac{dr_p}{dt}, 0, \Omega \right). \quad (3.3)$$

If we choose  $J^a = (\Omega, \delta M^\pm)$ , then  $dr_p/dt$  here is expressed in terms of  $F^a_{\{n\}}$  by applying the chain rule to  $r_p(t, \varepsilon) = r_p(J^a(t, \varepsilon), \varepsilon)$ . The normalization of the four-velocity for timelike curves,  $g_{\mu\nu} u^\mu u^\nu = -1$ , leads to an equation for the redshift,

$$U^{-2} = -g_{\mu\nu} \frac{dz^\mu}{dt} \frac{dz^\nu}{dt}. \quad (3.4)$$

Rewritten in terms of the parameter  $t$ , the equation of motion (2.3) becomes

$$\frac{d^2 z^\mu}{dt^2} + U^{-1} \frac{dU}{dt} \frac{dz^\mu}{dt} + \Gamma_{\nu\sigma}^\mu \frac{dz^\nu}{dt} \frac{dz^\sigma}{dt} = U^{-2} f^\mu, \quad (3.5)$$

where  $\Gamma_{\nu\sigma}^\mu$  are the background Schwarzschild Christoffel symbols. The self-force  $f^\mu$  has only two independent components because  $f^\theta = 0$  on equatorial orbits and because the normalization  $g_{\mu\nu} u^\mu u^\nu = -1$  implies  $u^\mu f_\mu = 0$ .

Obtaining the forcing functions  $F_{\{n\}}^\Omega$  will additionally require information from the transition-to-plunge regime. This is because the PG expansion is not self-contained; without information from the transition to plunge, we would have no way of determining *which* geodesic is the correct one to expand around, for example. Picking out the correct PG dynamics requires asymptotic matching conditions between the PG and PLT expansions. These conditions come from the requirement that the PG and PLT expansions commute. If we expand a function for small  $\varepsilon$  at fixed  $\Omega$  (a PG expansion) and then re-expand for small  $\varepsilon$  at fixed  $\Delta\Omega$  (a PLT expansion), we must obtain the same result as we do by first expanding for small  $\varepsilon$  at fixed  $\Delta\Omega$  and then re-expanding for small  $\varepsilon$  at fixed  $\Omega$ . In both cases, we obtain a double series in  $\varepsilon$  and  $(\Omega - \Omega_*)$ , and the two double series must agree term by term. Equivalently, we can say that a late-time expansion (for  $M\Delta\Omega \gg 1$ ) of the PLT dynamics must match, term by term, an early-time expansion ( $M(\Omega - \Omega_*) \ll 1$ ) of the PG dynamics.

### 3.1. Expansion at fixed orbital frequency

We start by expanding the orbital radius  $r_p$  and the redshift  $U$  in powers of the mass ratio  $\varepsilon$  at fixed phase-space coordinates  $\Omega$  and  $\delta M^\pm$ :

$$r_p(\Omega, \delta M^\pm, \varepsilon) = r_{\{0\}}(\Omega) + \sum_{n=1}^{\infty} \varepsilon^n r_{\{n\}}(\Omega, \delta M^\pm), \quad (3.6)$$

$$U(\Omega, \delta M^\pm, \varepsilon) = U_{\{0\}}(\Omega) + \sum_{n=1}^{\infty} \varepsilon^n U_{\{n\}}(\Omega, \delta M^\pm). \quad (3.7)$$

Note that these quantities are independent of  $\phi_p$ : as a consequence of the background's axisymmetry, the metric perturbation only ever depends on  $\phi_p$  in the combination  $(\phi - \phi_p)$  [28], such that it is independent of  $\phi_p$  when evaluated on the particle at  $\phi = \phi_p$ ; this implies the self-force and other dynamical quantities derived from it (such as  $r_{\{n\}}$ ) are likewise independent of  $\phi_p$ .

Since *all* functions are expanded at fixed phase-space coordinates, we also expand the coordinates' rates of change (with respect to  $t$ ) as

$$\dot{\Omega}(\Omega, \delta M^\pm, \varepsilon) = F_{\{0\}}^\Omega(\Omega) + \sum_{n=1}^{\infty} \varepsilon^n F_{\{n\}}^\Omega(\Omega, \delta M^\pm), \quad (3.8)$$

$$\delta \dot{M}^\pm(\Omega, \delta M^\pm, \varepsilon) = \sum_{n=1}^{\infty} \varepsilon^n F_{\{n\}}^\pm(\Omega, \delta M^\pm). \quad (3.9)$$

Finally, the self-force can be similarly expanded as

$$f^\mu(\Omega, \delta M^\pm, \varepsilon) = \sum_{n=1}^{\infty} \varepsilon^n f_{\{n\}}^\mu(\Omega, \delta M^\pm). \quad (3.10a)$$

We now substitute the post-geodesic expansion of the worldline (2.5) and the four-velocity (3.3) into the equation of motion (3.5) and the normalization condition (3.4). This leads straightforwardly to a sequence of equations for  $r_{\{n\}}$ ,  $U_{\{n\}}$ , and  $F_{\{n\}}^a$ , describing the plunging motion at each  $n$ PG order. Those equations are then completed using information from the transition-to-plunge regime.

#### 3.1.1. Leading-order match with the late-time transition-to-plunge solution

Our derivation of the 0PG dynamics in the next section will crucially rely on information from the asymptotic match with the transition to plunge close to the ISCO. Here we give a brief overview of how this matching works, focusing on the 0PG information it implies; we defer a more detailed analysis to section 3.2.

The transition-to-plunge expansions to 2PLT order for the orbital radius and the rate of change of the orbital frequency read [51]

$$r_p = 6M + \varepsilon^{2/5} \left[ r_{[0]}(\Delta\Omega) + \varepsilon^{2/5} r_{[2]}(\Delta\Omega) + \mathcal{O}(\varepsilon^{3/5}) \right], \quad (3.11)$$

$$\frac{d\Omega}{dt} = \varepsilon^{3/5} \left[ F_{[0]}^{\Delta\Omega}(\Delta\Omega) + \varepsilon^{2/5} F_{[2]}^{\Delta\Omega}(\Delta\Omega) + \mathcal{O}(\varepsilon^{3/5}) \right]. \quad (3.12)$$

The terms  $r_{[n]}$  are algebraically determined, while  $F_{[n]}^{\Delta\Omega}$  solve sourced linearized Painlevé transcendental equations of the first kind [51]. In the late-time limit  $\Delta\Omega \rightarrow +\infty$  (where the transition to plunge

asymptotically matches with the plunge) we have

$$r_{[0]} = r_{[0]}^{(2,1)} \Delta\Omega, \quad r_{[2]} = r_{[2]}^{(4,2)} \Delta\Omega^2 + \mathcal{O}(\Delta\Omega^{-1}), \quad (3.13)$$

and

$$F_{[0]}^{\Delta\Omega} = F_{[0]}^{(3,3/2)} \Delta\Omega^{3/2} + \mathcal{O}(\Delta\Omega^{-1}), \quad F_{[2]}^{\Delta\Omega} = F_{[2]}^{(5,5/2)} \Delta\Omega^{5/2} + \mathcal{O}(\Delta\Omega^0), \quad (3.14)$$

where  $r_{[0]}^{(2,1)}$ ,  $r_{[2]}^{(4,2)}$ ,  $F_{[0]}^{(3,3/2)}$  and  $F_{[2]}^{(5,5/2)}$  are numerical constants given in appendix A.1. Plugging these solutions into equations (3.11) and (3.12) and re-expressing  $\Delta\Omega$  as  $(\Omega - \Omega_*)/\varepsilon^{2/5}$  yields

$$r_p = \left[ 6M + r_{[0]}^{(2,1)} (\Omega - \Omega_*) + r_{[2]}^{(4,2)} (\Omega - \Omega_*)^2 + \mathcal{O}\left[(\Omega - \Omega_*)^3\right] \right] + \mathcal{O}(\varepsilon), \quad (3.15)$$

$$\frac{d\Omega}{dt} = \left[ F_{[0]}^{(3,3/2)} (\Omega - \Omega_*)^{3/2} + F_{[2]}^{(5,5/2)} (\Omega - \Omega_*)^{5/2} + \mathcal{O}\left[(\Omega - \Omega_*)^{7/2}\right] \right] + \mathcal{O}(\varepsilon). \quad (3.16)$$

The  $\mathcal{O}(\varepsilon)$  terms come from the subleading terms in the late-time expansion of the PLT orders, e.g. the  $\mathcal{O}(\Delta\Omega^{-1})$  term in  $F_{[0]}^{\Delta\Omega}$ . Subleading terms in the square brackets originate from the leading-order terms in the late-time expansion of higher PLT orders.

As explained above, the  $\varepsilon^0$  terms in equations (3.15) and (3.12) must match, term by term, with the near-ISCO expansion of OPG quantities in the plunge. Hence, we can infer that, in the near-ISCO limit  $\Omega \rightarrow \Omega_*$ ,

$$r_{\{0\}} = 6M + r_{[0]}^{(2,1)} (\Omega - \Omega_*) + r_{[2]}^{(4,2)} (\Omega - \Omega_*)^2 + \mathcal{O}\left[(\Omega - \Omega_*)^3\right], \quad (3.17)$$

$$F_{\{0\}}^\Omega = F_{[0]}^{(3,3/2)} (\Omega - \Omega_*)^{3/2} + F_{[2]}^{(5,5/2)} (\Omega - \Omega_*)^{5/2} + \mathcal{O}\left[(\Omega - \Omega_*)^{7/2}\right]. \quad (3.18)$$

From this, we see that the OPG plunge trajectory must asymptote to a circular orbit with  $r_{\{0\}} \rightarrow 6M$  and  $\dot{\Omega} \rightarrow 0$ —i.e. a circular geodesic infinitesimally below the geodesic ISCO.

### 3.1.2. Geodesic order

At leading order, we can use the  $t$  component of the equation of motion (3.5) to solve for  $\partial_\Omega^2 r_{\{0\}}$ :

$$\begin{aligned} \partial_\Omega^2 r_{\{0\}} &= \frac{3M (\partial_\Omega r_{\{0\}})^2}{r_{\{0\}} (r_{\{0\}} - 2M)} - \frac{\partial_\Omega F_{\{0\}}^\Omega \partial_\Omega r_{\{0\}}}{F_{\{0\}}^\Omega} \\ &+ \frac{2M^2 - r_{\{0\}} \left( r_{\{0\}}^2 \Omega^2 (r_{\{0\}} - 4M) + M \right) - r_{\{0\}}^4 \Omega (r_{\{0\}} - 2M) / \partial_\Omega r_{\{0\}}}{r_{\{0\}}^3 \left( F_{\{0\}}^\Omega \right)^2}. \end{aligned} \quad (3.19)$$

Plugging this result into the radial component of the equation of motion allows us to obtain the following simple expression for  $\partial_\Omega r_{\{0\}}$ :

$$\partial_\Omega r_{\{0\}} = -\frac{r_{\{0\}} (r_{\{0\}} - 2M)}{2\Omega (r_{\{0\}} - 3M)}. \quad (3.20)$$

We can use this to simplify  $U_{\{0\}}$ , which we obtain from equation (3.4) at order  $\mathcal{O}(\varepsilon^0)$ ,

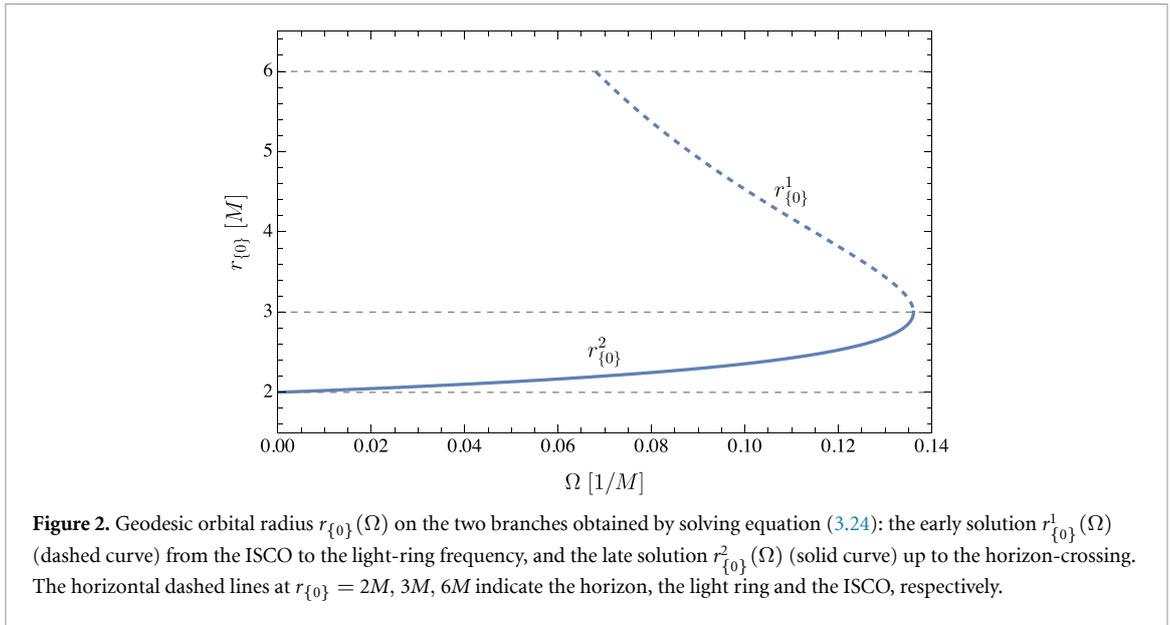
$$U_{\{0\}} = \frac{2\Omega r_{\{0\}}^{1/2} (r_{\{0\}} - 3M)}{\sqrt{4\Omega^2 (r_{\{0\}} - 3M)^2 \left( r_{\{0\}} - 2M - r_{\{0\}}^3 \Omega^2 \right) - (r_{\{0\}} - 2M) r_{\{0\}}^4 \left( F_{\{0\}}^\Omega \right)^2}}. \quad (3.21)$$

The geodesic orbital energy and angular momentum per unit mass of the plunging particle,  $E_{\{0\}}$  and  $L_{\{0\}}$ , are given by

$$E_{\{0\}} = -u_t^{\{0\}} = -g_{tt} (r_{\{0\}}) U_{\{0\}} = f(r_{\{0\}}) U_{\{0\}}, \quad (3.22a)$$

$$L_{\{0\}} = u_\phi^{\{0\}} = g_{\phi\phi} (r_{\{0\}}, \pi/2) \Omega U_{\{0\}} = r_{\{0\}}^2 \Omega U_{\{0\}}. \quad (3.22b)$$

Since  $u_t$  and  $u_\phi$  are constant along any geodesic of Schwarzschild,  $E_{\{0\}}$  and  $L_{\{0\}}$  must be constant ‘on shell’, when  $\Omega$  satisfies  $\dot{\Omega} = F_{\{0\}}^\Omega$ . Moreover, making use of equation (3.19), it is straightforward to verify that they are constants even if  $\Omega$  satisfies  $\dot{\Omega} = F^\Omega(\Omega, \varepsilon)$  because  $dE_{\{0\}}/d\Omega = dL_{\{0\}}/d\Omega = 0$ . The



**Figure 2.** Geodesic orbital radius  $r_{\{0\}}(\Omega)$  on the two branches obtained by solving equation (3.24): the early solution  $r_{\{0\}}^1(\Omega)$  (dashed curve) from the ISCO to the light-ring frequency, and the late solution  $r_{\{0\}}^2(\Omega)$  (solid curve) up to the horizon-crossing. The horizontal dashed lines at  $r_{\{0\}} = 2M, 3M, 6M$  indicate the horizon, the light ring and the ISCO, respectively.

values of these constants are fixed by the requirement for the plunge to asymptotically match the transition to plunge close to the ISCO. We can take the limit as  $\Omega \rightarrow \Omega_*$  of equations (3.22) keeping in mind that the geodesic radius reduces to the ISCO radius while  $F_{\{0\}}^\Omega$  vanishes as prescribed by equations (3.17) and (3.18). We get

$$E_{\{0\}} = E_* := \frac{2\sqrt{2}}{3}, \quad L_{\{0\}} = L_* := 2\sqrt{3}M, \quad (3.23)$$

which correspond to the (specific) energy and angular momentum of a test particle on the ISCO. The plunging geodesics with these values of  $E$  and  $L$  are the ‘universal’ geodesics studied in the past [57, 58, 79–81]; at OPG order, all quasicircular inspirals transition onto one of these geodesics, with the only difference between them being a constant shift in  $\phi_p$ .

By imposing  $E_{\{0\}}/L_{\{0\}} = E_*/L_*$ , where  $E_{\{0\}}$  and  $L_{\{0\}}$  are given in equation (3.22) and the constants  $E_*$  and  $L_*$  are defined in equation (3.23), we obtain a cubic master equation for the orbital radius in terms of the orbital frequency,

$$\sqrt{2}\Omega r_{\{0\}}^3 - 3\sqrt{3}Mr_{\{0\}} + 6\sqrt{3}M^2 = 0. \quad (3.24)$$

The solution has two branches, as shown in figure 2: on the first branch,  $r_{\{0\}}^1(\Omega)$ , the orbital frequency increases monotonically from  $\Omega = \Omega_* := 1/(6\sqrt{6}M)$  at the ISCO to its light ring value  $\Omega = \Omega_{\text{LR}} := 1/(3\sqrt{6}M)$ . On the second branch,  $r_{\{0\}}^2(\Omega)$ , the orbital frequency monotonically decreases to  $\Omega = 0$  at the horizon-crossing.

The condition  $(E_{\{0\}})^2 = (E_*)^2$ , together with equations (3.21) and (3.22), also yields the expression for the geodesic forcing term,

$$F_{\{0\}}^\Omega = \frac{\Omega (r_{\{0\}} - 3M) \left( 36M^2 - 20Mr_{\{0\}} + r_{\{0\}}^2 + 8\Omega^2 r_{\{0\}}^4 \right)^{1/2}}{\sqrt{2}r_{\{0\}}^{5/2} (2M - r_{\{0\}})^{1/2}}. \quad (3.25)$$

Above the light ring, the frequency increases with  $t$  ( $F_{\{0\}}^\Omega(r_{\{0\}}) > 0$  for  $3M < r_{\{0\}} < 6M$ ), while it decreases between the light ring and the horizon ( $F_{\{0\}}^\Omega(r_{\{0\}}) < 0$  for  $2M < r_{\{0\}} < 3M$ ), consistently with figure 2.

### 3.1.3. First post-geodesic order

At first post-geodesic order, the  $t$  and  $r$  components of the equation of motion (3.5) lead to hierarchical differential equations for  $r_{\{1\}}$  and  $F_{\{1\}}^\Omega$ ,

$$\partial_{\Omega} r_{\{1\}} + \frac{r_{\{0\}}^2 - 6Mr_{\{0\}} + 6M^2}{2\Omega(r_{\{0\}} - 3M)^2} r_{\{1\}} = \frac{1}{32r_{\{0\}}^4 \Omega^3 (r_{\{0\}} - 3M)^2} \left[ 9r_{\{0\}}^3 (r_{\{0\}} - 2M)^3 f_{\{1\}}^r \right. \\ \left. - \frac{18\sqrt{2}r_{\{0\}}^{5/2} (2M - r_{\{0\}})^{7/2} (r_{\{0\}}^3 \Omega^2 - r_{\{0\}} + 2M)}{(36M^2 - 20Mr_{\{0\}} + r_{\{0\}}^2 + 8\Omega^2 r_{\{0\}}^4)^{1/2}} f_{\{1\}}^t \right], \quad (3.26a)$$

$$\partial_{\Omega} F_{\{1\}}^{\Omega} - \frac{4\Omega(3M^2 + \Omega^2 r_{\{0\}}^4 - M(3\Omega^2 r_{\{0\}}^3 + r_{\{0\}})) + r_{\{0\}}^4 F_{\{0\}}^{\Omega} \partial_{\Omega} F_{\{0\}}^{\Omega}}{r_{\{0\}}^4 (F_{\{0\}}^{\Omega})^2} F_{\{1\}}^{\Omega} \\ = \frac{1}{4\Omega^2 r_{\{0\}}^5 (r_{\{0\}} - 3M)^2 (r_{\{0\}} - 2M)^2 (F_{\{0\}}^{\Omega})^2} \left[ 8\Omega^3 r_{\{0\}}^4 (r_{\{0\}} - 2M)(r_{\{0\}} - 3M)^3 (F_{\{0\}}^{\Omega})^3 \partial_{\Omega}^2 r_{\{1\}} \right. \\ \left. - 8\Omega^2 r_{\{0\}}^3 (r_{\{0\}} - 3M)^2 F_{\{0\}}^{\Omega} \left[ 4\Omega^4 (r_{\{0\}} - 3M)^3 - 3Mr_{\{0\}} (r_{\{0\}} - 2M)(F_{\{0\}}^{\Omega})^2 \right. \right. \\ \left. \left. - \Omega r_{\{0\}} (r_{\{0\}} - 3M)(r_{\{0\}} - 2M) F_{\{0\}}^{\Omega} \partial_{\Omega} F_{\{0\}}^{\Omega} \right] \partial_{\Omega} r_{\{1\}} \right. \\ \left. - 4\Omega (r_{\{0\}} - 3M) F_{\{0\}}^{\Omega} \left[ 2\Omega^2 (r_{\{0\}} - 3M)^2 (12M^3 + 3\Omega^2 r_{\{0\}}^5 + 2M^2 r_{\{0\}} (6\Omega^2 r_{\{0\}}^2 - 5) \right. \right. \\ \left. \left. - M(14\Omega^2 r_{\{0\}}^4 - 2r_{\{0\}}^2) \right) - 3Mr_{\{0\}}^4 (r_{\{0\}}^2 - 3Mr_{\{0\}} + 2M^2) (F_{\{0\}}^{\Omega})^2 \right] r_{\{1\}} \right. \\ \left. + (r_{\{0\}} - 2M) \left[ 4\Omega^2 (r_{\{0\}} - 3M)^2 (\Omega^2 r_{\{0\}}^3 - r_{\{0\}} + 2M) + r_{\{0\}}^4 (r_{\{0\}} - 2M) (F_{\{0\}}^{\Omega})^2 \right]^2 f_{\{1\}}^t \right]. \quad (3.26b)$$

$$+ (r_{\{0\}} - 2M) \left[ 4\Omega^2 (r_{\{0\}} - 3M)^2 (\Omega^2 r_{\{0\}}^3 - r_{\{0\}} + 2M) + r_{\{0\}}^4 (r_{\{0\}} - 2M) (F_{\{0\}}^{\Omega})^2 \right]^2 f_{\{1\}}^t \right]. \quad (3.26c)$$

Contrary to the 0PG equations, which describe a geodesic in the background spacetime, the motion at 1PG order is driven by both the  $t$  and  $r$  components of the first-order self-force,  $f_{\{1\}}^t$  and  $f_{\{1\}}^r$ .

We notice that the post-geodesic expansion derived in this section breaks down at the light ring, where the evolution equations (3.26) encounter a pole. This is a consequence of the fact that the frequency does not provide a global coordinate for the whole plunge, as mentioned previously. It motivates an alternative choice of coordinates, which we develop in section 3.3.

Like the 0PG dynamics, the 1PG dynamics is not fully specified by the orbital equations of motion (3.26). Since those equations are first-order differential equations, they only determine  $F_{\{1\}}^{\Omega}$  up to a constant. Fixing that constant requires additional input from the transition-to-plunge regime. Moreover, we note that equation (3.26) becomes singular at the ISCO as  $F_{\{0\}}^{\Omega}$  vanishes there. In general, we expect the post-geodesic expansion to become increasingly singular at the ISCO at higher PG orders. Matching to the transition-to-plunge solution then becomes increasingly important.

### 3.2. Asymptotic match with the transition to plunge

In this section, following [75], we derive the asymptotic match with the late-time transition-to-plunge motion of [51]. Specifically, we generalize the brief description in the previous section to 1PG order and high PLT orders. This will allow us to determine how the transition to plunge selects the correct solution to the 1PG equations (3.26), analogously to what we have done at geodesic order in section 3.1.2.

To explain the structure of the asymptotic matching, we consider a function  $g(\varepsilon, J^a)$  that admits both a transition-to-plunge expansion (a regular expansion in powers of  $\lambda := \varepsilon^{1/5}$  at fixed  $\Delta J^a$ ) and a plunge expansion (a regular expansion in powers of  $\varepsilon$  at fixed  $J^a$ ). We write the PLT expansion as  $g(\lambda, \Delta J^a) = g_* + \lambda^p \sum_{n \geq 0} \lambda^n g_{[n]}(\Delta J^a)$ ; for examples, see equations (3.11) and (3.12) for  $r_p$  and  $d\Omega/dt$ . At late times, where  $\Delta\Omega \rightarrow +\infty$ , we can expand in half-integer powers of  $\Delta\Omega$ . Since  $\Delta\Omega = (\Omega - \Omega_*)/\varepsilon^{2/5}$ , re-expanding  $g$  for large  $\Delta\Omega$  is equivalent to re-expanding for small  $\varepsilon$  at fixed  $\Omega > \Omega_*$ :

$$g(\lambda, \Delta J^a) = g_* + \sum_{n \geq 0} \sum_m \lambda^{p+n} \Delta\Omega^{m/2} g_{[n]}^{(p+n, m/2)} (\delta M^{\pm}) \quad (3.27a)$$

$$= g_* + \sum_{n \geq 0} \sum_m \varepsilon^{(p+n-m)/5} (\Omega - \Omega_*)^{m/2} g_{[n]}^{(p+n, m/2)} (\delta M^{\pm}). \quad (3.27b)$$

In principle, there is no restriction on the integers  $m$ . However, the range of  $m$  is restricted by the requirement of asymptotically matching the early-time behaviour of the plunge expansion, which contains only integer powers of the mass ratio,  $\varepsilon^i$  with  $i = 0, 1, 2, \dots$ ; see section 3.1. This restricts the powers of  $\Delta\Omega$  that can appear in the late-time expansion of a given  $n$ PLT order:

$$m = p + n - 5i, \quad n, i \geq 0. \quad (3.28)$$

As an example, if we consider the 0PLT forcing term  $F_{[0]}^{\Delta\Omega}$  we have the following: as anticipated in section 3.1.1, since  $p = 3$  and  $n = 0$ , the late-time expansion is given by a series expansion with powers  $\Delta\Omega^{(3-5i)/2}$ . We obtain the coefficients  $r_{[n]}^{(p+n, m/2)}$  for  $r_p$  and  $F_{[n]}^{(p+n, m/2)}$  for  $\dot{\Omega}$  by substituting expansions of the form (3.27a) into the transition-regime equations for  $r_{[n]}$  and  $F_{[n]}^{\Delta\Omega}$  (see [51]). These coefficients are given explicitly in appendix A.1.

We can follow the analogous procedure for the plunge. The plunge expansion of a function  $g$  is given by  $g(\varepsilon, J^a) = \sum_{i \geq 0} \varepsilon^i g_{\{i\}}(J^a)$ . Re-expanding close to the ISCO and substituting  $(\Omega - \Omega_*) = \lambda^2 \Delta\Omega$  we obtain

$$g(\varepsilon, J^a) = g_* + \sum_{i \geq 0} \sum_j \lambda^{5i+j} \Delta\Omega^{j/2} g_{\{i\}}^{(5i+j, j/2)}(\delta M^\pm). \quad (3.29)$$

We can restrict the range of  $j$  by requiring the powers of  $\lambda$  in the near-ISCO plunge solution to match those in the late-time transition-to-plunge solution (3.27a). This yields the condition

$$j = p + n - 5i, \quad n, i \geq 0. \quad (3.30)$$

At geodesic order ( $i = 0$ ), we plug the asymptotic expansions (3.29) into equations (3.24) and (3.25) and solve for the coefficients  $g_{\{0\}}^{(j, j/2)}$  with  $j \geq p + n$  and  $n \geq 0$ . These coefficients are given explicitly in appendix A.2.

At first post-geodesic order, we need to know the near-ISCO structure of the self-force. The first-order field equations in the plunge regime yield a metric perturbation that contains pieces which are smooth at the ISCO and pieces  $\sim F_{\{0\}}^{\Omega} \sim (\Omega - \Omega_*)^{3/2}$  due to the presence of time derivatives (see section 5.1, where the first-order field equations are presented in a slightly different expansion scheme, but the general statement made here remains unchanged). The self-force inherits this structure through equation (2.4), where at most one time derivative acts on the metric perturbations and gives rise to half-integer powers  $\geq 3/2$  of  $(\Omega - \Omega_*)$  in the near-ISCO limit. We can therefore write the first-order plunge self-force as

$$f_{\{1\}}^\mu(\Omega \rightarrow \Omega_*, \delta M^\pm) = f_{\{1\},0}^\mu(\delta M^\pm) + \sum_{n=2}^{\infty} f_{\{1\},n}^\mu(\delta M^\pm) (\Omega - \Omega_*)^{n/2}, \quad (3.31)$$

where  $f_{\{1\},n}^\mu$  and  $f_{\{1\},n}^\phi$  are independent of  $\delta M^\pm$  for  $n = 0, 2$ .<sup>7</sup>

Knowing the structure of the self-force, we can solve equation (3.26) in the near-ISCO limit. From equations (3.29) and (3.30) we get

$$\varepsilon r_{\{1\}}(\Omega \rightarrow \Omega_*, \delta M^\pm) = \lambda^5 \sum_{j \geq -3} r_{\{1\}}^{(5+j, j/2)} \lambda^j \Delta\Omega^{j/2}, \quad (3.32)$$

$$\varepsilon F_{\{1\}}^\Omega(\Omega \rightarrow \Omega_*, \delta M^\pm) = \lambda^5 \sum_{j \geq -2} F_{\{1\}}^{(5+j, j/2)} \lambda^j \Delta\Omega^{j/2}. \quad (3.33)$$

From the transition to plunge we have that  $r_{[0]}^{(2, -3/2)} = 0$  and  $r_{[1]}^{(3, -1)} = 0$ ; see appendix A.1. Through the asymptotic match this also implies that  $r_{\{1\}}^{(2, -3/2)} = r_{\{1\}}^{(2, -1)} = 0$ , and the series in equation (3.32) therefore starts at  $\Delta\Omega^{-1/2}$ . Solving equation (3.26) iteratively in the limit  $\Omega \rightarrow \Omega_*$  determines all these coefficients, with the exception of  $r_{\{1\}}^{(5,0)}$  in equation (3.32). We fix this coefficient by enforcing the match with the transition-to-plunge solution:  $r_{\{1\}}^{(5,0)} = r_{[3]}^{(5,0)} = -54M^2 f_{\{1\},0}^\mu$ . The coefficients appearing in the near-ISCO solutions (3.32) and (3.33) are given explicitly in appendix A.2. The asymptotic solutions (3.32) and (3.33) can then be used to obtain initial conditions to solve equation (3.26) (once the plunge self-force has been computed) and obtain 1PG solutions that asymptotically match with the transition to plunge close to the ISCO.

In summary, the asymptotic matching has (i) restricted the form of the plunge expansion by restricting the powers of  $(\Omega - \Omega_*)$  in the plunge's near-ISCO behaviour, and (ii) fixed the value of a single

<sup>7</sup> For  $\mu = t, \phi$ , linear-in- $\delta M^\pm$  perturbations towards Kerr appear at first order in equation (2.4) always accompanied by a factor  $u^r = \partial_\Omega r_{\{0\}} F_{\{0\}}^\Omega U_{\{0\}}$ . Close to the ISCO,  $u^r$  is  $\mathcal{O}[(\Omega - \Omega_*)^{3/2}]$  or higher, as follows from equations (3.17) and (3.18). The coefficients in equation (3.31) therefore start to depend on  $\delta M^\pm$  from  $n = 3$ . This is consistent with the results of [51] (i.e. that  $f_{[5]}^\mu$  and  $f_{[7]A}^\mu$  in that paper do not depend on  $\delta M^\pm$  for  $\mu = t, \phi$ ), which becomes manifest from the match of the self-force in equation (A.23) below.

**Table 2.** Matching conditions between plunge and transition to plunge for the asymptotic coefficients ( $r_{\{n\}}^{(p,q)}$ ,  $r_{[n]}^{(p,q)}$ ,  $F_{\{n\}}^{(p,q)}$ , and  $F_{[n]}^{(p,q)}$ ) of  $\lambda^p(\Delta\Omega)^q$  in the solutions for the orbital radius  $r_p$  and the rate of change  $d\Omega/dt$ . The coefficients are labelled with their PLT ( $\{n\}$ ) or PG order ( $\{n\}$ ) in addition to the powers of  $\lambda = \varepsilon^{1/5}$  and of  $\Delta\Omega = (\Omega - \Omega_*)/\lambda^2$  in the asymptotic expansions (3.27) and (3.29). Here we only show nonzero coefficients.

	0PG	1PG	...
0PLT	$r_{\{0\}}^{(2,1)} = r_{[0]}^{(2,1)}$ $F_{\{0\}}^{(3,3/2)} = F_{[0]}^{(3,3/2)}$	— $F_{\{1\}}^{(3,-1)} = F_{[0]}^{(3,-1)}$	...
1PLT	—	—	...
2PLT	$r_{\{0\}}^{(4,2)} = r_{[2]}^{(4,2)}$ $F_{\{0\}}^{(5,5/2)} = F_{[2]}^{(5,5/2)}$	$r_{\{1\}}^{(4,-1/2)} = r_{[2]}^{(4,-1/2)}$ $F_{\{1\}}^{(5,0)} = F_{[2]}^{(5,0)}$	...
3PLT	—	$r_{\{1\}}^{(5,0)} = r_{[3]}^{(5,0)}$ $F_{\{1\}}^{(6,1/2)} = F_{[3]}^{(6,1/2)}$	...
4PLT	$r_{\{0\}}^{(6,3)} = r_{[4]}^{(6,3)}$ $F_{\{0\}}^{(7,7/2)} = F_{[4]}^{(7,7/2)}$	$r_{\{1\}}^{(6,1/2)} = r_{[4]}^{(6,1/2)}$ $F_{\{1\}}^{(7,1)} = F_{[4]}^{(7,1)}$	...
5PLT	—	$r_{\{1\}}^{(7,1)} = r_{[5]}^{(7,1)}$ $F_{\{1\}}^{(8,3/2)} = F_{[5]}^{(8,3/2)}$	...
6PLT	$r_{\{0\}}^{(8,4)} = r_{[6]}^{(8,4)}$ $F_{\{0\}}^{(9,9/2)} = F_{[6]}^{(9,9/2)}$	$r_{\{1\}}^{(8,3/2)} = r_{[6]}^{(8,3/2)}$ $F_{\{1\}}^{(9,2)} = F_{[6]}^{(9,2)}$	...
7PLT	—	$r_{\{1\}}^{(9,2)} = r_{[7]}^{(9,2)}$ $F_{\{1\}}^{(10,5/2)} = F_{[7]}^{(10,5/2)}$	...
⋮	⋮	⋮	⋮

near-ISCO coefficient,  $r_{\{1\}}^{(5,0)}$ . This minimal information from the matching (together with the orbital equations of motion, field equations, and boundary conditions) serves to completely fix the plunge solution.

Although the plunge solution requires only a small amount of information from the transition regime, we can use the other term-by-term matching conditions as consistency checks. The conditions for equations (3.27) and (3.29) to match term by term are obtained by equating the powers of  $\lambda$  and  $\Delta\Omega$ :

$$p + n = 5i + j, \quad m = j. \tag{3.34}$$

Therefore,

- the near-ISCO expansion of the 0PG forcing term  $F_{\{0\}}^\Omega$  ( $p = 3, i = 0$ ) is matched by the terms with  $m = 3 + n$  in equation (3.27), that is, a term  $\sim \Delta\Omega^{3/2}$  from 0PLT, a (vanishing) term  $\sim \Delta\Omega^2$  from 1PLT, a term  $\sim \Delta\Omega^{5/2}$  from 2PLT, ...
- the near-ISCO expansion of the 0PG forcing term  $F_{\{1\}}^\Omega$  ( $p = 3, i = 1$ ) is matched by the terms with  $m = -2 + n$  in equation (3.27), that is, a term  $\sim \Delta\Omega^{-1}$  from 0PLT, a (vanishing) term  $\sim \Delta\Omega^{-1/2}$  from 1PLT, a constant term  $\sim \Delta\Omega^0$  from 2PLT, ...

and analogously for the orbital radius. This structure of the asymptotic match between the transition-to-plunge and plunge orbital motions is summarized in table 2.

We have explicitly verified the matching conditions in table 2 for all terms involved in the match between the transition to plunge to 7PLT order and the plunge to 1PG order, using the coefficients listed in appendices A.1 and A.2. In order to do so successfully, we required the match of the self-force. In practice, one proceeds order by order, obtaining matching conditions for both the orbital motion and the self-force. The matching conditions for the self-force are presented in appendix A.3.

### 3.3. Expansion at fixed orbital radius

In this section we derive an alternative to the formulation of section 3.1. Instead of parametrizing the orbit using the orbital frequency  $\Omega$ , we can recast the equations describing the plunge in terms of the orbital radius  $r_p$ , which is monotonically decreasing during the plunge—in contrast to  $\Omega$ , which encounters a maximum at the light ring (see figure 2). We refer to the two approaches as the fixed- $\Omega$  and the fixed- $r_p$  formulation, respectively. In appendix B we show how to transform between the formulations. Numerous other quantities would serve equally well as global coordinates for the plunge; an example is the orbital frequency with respect to advanced time,  $\Omega_v := d\phi_p/dv$ . However, we find the plunge dynamics is particularly simple when written in terms of  $r_p$ .

If using  $r_p$  as a phase-space coordinate, we expand all orbital quantities in powers of  $\varepsilon$  holding the coordinates  $(\phi_p, r_p, \delta M^\pm)$  fixed. The orbital frequency  $\Omega$  and the redshift  $U$  are therefore expanded as

$$\Omega(r_p, \delta M^\pm, \varepsilon) = \Omega_{\{0\}}(r_p) + \sum_{n=1}^{\infty} \varepsilon^n \Omega_{\{n\}}(r_p, \delta M^\pm), \quad (3.35)$$

$$U(r_p, \delta M^\pm, \varepsilon) = U_{\{0\}}(r_p) + \sum_{n=1}^{\infty} \varepsilon^n U_{\{n\}}(r_p, \delta M^\pm). \quad (3.36)$$

The expansion of the rate of change of the orbital radius is given by

$$\dot{r}_p(r_p, \delta M^\pm, \varepsilon) = F_{\{0\}}^r(r_p) + \sum_{n=1}^{\infty} \varepsilon^n F_{\{n\}}^r(r_p, \delta M^\pm), \quad (3.37)$$

while the corrections to the background mass and spin evolve as

$$\delta \dot{M}^\pm(r_p, \delta M^\pm, \varepsilon) = \sum_{n=1}^{\infty} \varepsilon^n F_{\{n\}}^\pm(r_p, \delta M^\pm). \quad (3.38)$$

The expansion of the self-force finally reads

$$f^\mu(r_p, \delta M^\pm, \varepsilon) = \sum_{n=1}^{\infty} \varepsilon^n f_{\{n\}}^\mu(r_p, \delta M^\pm). \quad (3.39)$$

As we have done for the fixed- $\Omega$  formulation, we now perform the post-geodesic expansion of the equation of motion (3.5) and the normalization of the four-velocity (3.4), and obtain equations describing the plunging motion at each  $n$ PG order.

#### 3.3.1. Geodesic order

We follow a procedure analogous to the one used for the fixed- $\Omega$  formulation. At leading order, we then find the geodesic quantities

$$\Omega_{\{0\}}(r_p) = \sqrt{\frac{3}{2}} \frac{3M}{r_p^2} f(r_p), \quad U_{\{0\}}(r_p) = \frac{2\sqrt{2}}{3f(r_p)}. \quad (3.40)$$

The leading-order rate of change of the orbital radius,  $\dot{r}_{\{0\}}$ , is defined as

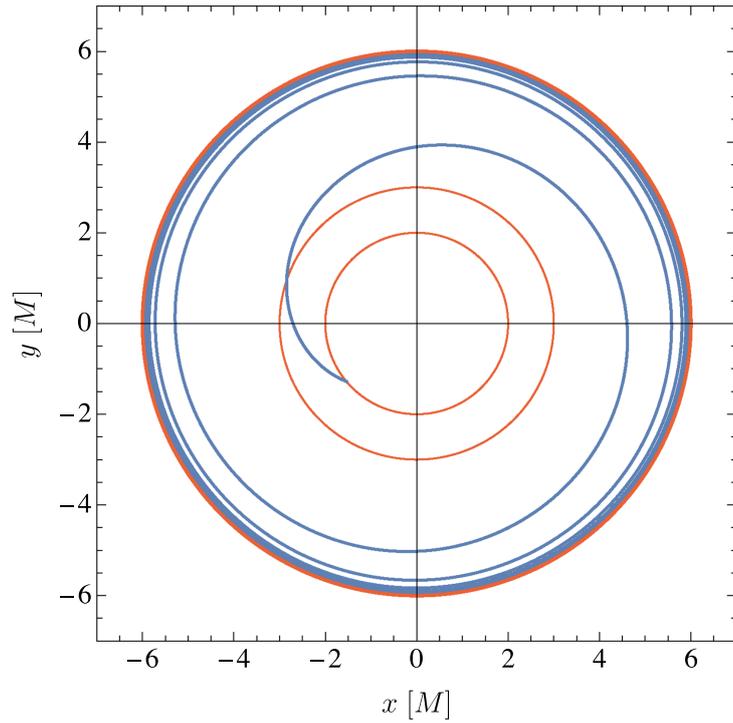
$$\dot{r}_{\{0\}}(r_p) := F_{\{0\}}^r(r_p) = -\frac{1}{2^{3/2}} (6M/r_p - 1)^{3/2} f(r_p), \quad (3.41)$$

and the leading-order radial acceleration,  $\ddot{r}_{\{0\}}$ , is then given by

$$\ddot{r}_{\{0\}}(r_p) := \dot{r}_{\{0\}} \partial_r \dot{r}_{\{0\}} = \frac{M(30M/r_p - 11)(6M/r_p - 1)^2}{8r_p^2} f(r_p). \quad (3.42)$$

We also obtain expressions for the coordinate time and the azimuthal angle at geodesic order:

$$t_G(r_p) = \int^{r_p} \frac{dr'_p}{\dot{r}_{\{0\}}(r'_p)} = \frac{2\sqrt{2}(r_p - 24M)}{(6M/r_p - 1)^{1/2}} - 44\sqrt{2}M \arctan \left[ (6M/r_p - 1)^{1/2} \right] + 4M \operatorname{arctanh} \left[ \frac{1}{\sqrt{2}} (6M/r_p - 1)^{1/2} \right] + t_0, \quad (3.43)$$



**Figure 3.** The trajectory of the geodesic plunge obtained from equation (3.47) (blue curve). In this plot we set  $\phi_0 = 0$ . The red lines at  $r = 2M, 3M, 6M$  mark the event horizon, light ring and ISCO, respectively.

$$\phi_G(r_p) = \int^{r_p} \Omega_{\{0\}}(r'_p) \frac{dr'_p}{\dot{r}_{\{0\}}(r'_p)} = -\frac{2\sqrt{3}}{(6M/r_p - 1)^{1/2}} + \phi_0, \quad (3.44)$$

where  $t_0$  and  $\phi_0$  are integration constants. The asymptotic behaviours close to the horizon and to the ISCO are given by

$$t_G(r_p \rightarrow 2M) = -2M \log\left(\frac{r_p}{2M} - 1\right) + \mathcal{O}\left[(r_p - 2M)^0\right], \quad (3.45)$$

$$t_G(r_p \rightarrow 6M) = -72\sqrt{3}M^{3/2}(6M - r_p)^{-1/2} + \mathcal{O}\left[(6M - r_p)^{1/2}\right]. \quad (3.46)$$

By inverting equation (3.44), we obtain a simple relationship for the plunging trajectory at geodesic order:

$$r_p(\phi_G) = \frac{6M}{1 + 12/(\phi_G - \phi_0)^2}. \quad (3.47)$$

This is a known result at least since [82]. The plunging trajectory is displayed in figure 3.

Our expressions at geodesic order match those in [57, 58].

### 3.3.2. First post-geodesic order

At first post-geodesic order we obtain the following nested differential equations for  $\Omega_{\{1\}}$  and  $F_{\{1\}}^{r_p}$  from the time and radial components of equation (3.5):

$$\partial_r \Omega_{\{1\}} + \frac{2(r_p - 3M)}{r_p(r_p - 2M)} \Omega_{\{1\}} = -\frac{\sqrt{6}}{8M} f_{\{1\}}^r - \frac{\sqrt{3}(r_p - 2M)(2r_p^3 - 27M^2r_p + 54M^3)}{4Mr_p^{5/2}(6M - r_p)^{3/2}} f_{\{1\}}^t, \quad (3.48a)$$

$$\begin{aligned} \partial_r F_{\{1\}}^{r_p} - \frac{M(5r_p + 6M)}{r_p(r_p - 2M)(6M - r_p)} F_{\{1\}}^{r_p} = & -\frac{12\sqrt{3}M(r_p - 2M)}{r_p^{1/2}(6M - r_p)^{3/2}} \Omega_{\{1\}} - \frac{9(r_p - 2M)^2}{8r_p^2} f_{\{1\}}^t \\ & - \frac{9\sqrt{2}r_p^{1/2}(r_p - 2M)}{4(6M - r_p)^{3/2}} f_{\{1\}}^r. \end{aligned} \quad (3.48b)$$

We notice that these equations are simpler and more compact than the ones in equation (3.26) obtained from the fixed- $\Omega$  formulation. More importantly, equation (3.48) admits a smooth evolution across the light ring.

Solving equation (3.48) requires initial conditions close to the ISCO. We start by considering the near-ISCO behaviour of the plunge self-force:

$$f_{\{1\}}^\mu (r_p \rightarrow 6M, \delta M^\pm) = f_{\{1\},0}^\mu (\delta M^\pm) + \sum_{n=2}^{\infty} f_{\{1\},n}^\mu (\delta M^\pm) (6M - r_p)^{n/2}. \quad (3.49)$$

This follows from equation (2.4) and the field equations (5.7) below: the first-order field contains a piece that is a smooth function of  $r_p$  at the ISCO and a piece that is not smooth  $\sim \dot{r}_{\{0\}} \sim (6M - r_p)^{3/2}$ . Equation (2.4) does not alter this structure since it introduces at most one time derivative, which yields terms of the form  $\sim \dot{r}_{\{0\}} \sim (6M - r_p)^{3/2}$  and with higher half-integer powers. The coefficients in equation (3.49) can be obtained via asymptotic matching with the transition-to-plunge regime in an analogous manner to what is shown in appendix A.3. Using equation (3.49), we then get the following asymptotic solutions from the near-ISCO limit of equation (3.48) (these solutions are obtained by enforcing the match with the late-time transition to plunge, similarly to the procedure described in section (3.2)):

$$\Omega_{\{1\}} (r_p \rightarrow 6M, \delta M^\pm) = -\frac{9\sqrt{2}M^{1/2}f_{\{1\},0}^t}{(6M - r_p)^{1/2}} - \frac{3}{4}\sqrt{\frac{3}{2}}f_{\{1\},0}^r + \mathcal{O}(6M - r_p)^{1/2}, \quad (3.50)$$

$$F_{\{1\}}^r (r_p \rightarrow 6M, \delta M^\pm) = -\frac{864M^2f_{\{1\},0}^t}{(6M - r_p)} + 24M(f_{\{1\},0}^t + 12Mf_{\{1\},2}^t) + \mathcal{O}(6M - r_p)^{1/2}. \quad (3.51)$$

These solutions can be used to obtain initial conditions to solve equation (3.48) once the plunge self-force has been computed.

#### 4. Plunge: field equations at first and second order

We now turn to deriving the field equations for the metric perturbation (2.15). We first consider the simple case in which the global time  $s$  reduces to  $t$  along the particle's trajectory, as assumed in section 2. We then describe how to lift that restriction.

The essential idea, regardless of which regime we consider and which choice of  $s$  we make, is to convert derivatives with respect to time into differential operators on phase space. When acting on a function of  $(\phi_p(s, \varepsilon), J^a(s, \varepsilon), x^i)$ , assuming  $s = t$  on the trajectory, we apply the chain rule together with  $d\phi_p/ds = \Omega$  and  $dJ^a/ds = F^a(J^b, \varepsilon)$ :

$$\frac{\partial}{\partial s} = \Omega \frac{\partial}{\partial \phi_p} + F^a(J^b, \varepsilon) \frac{\partial}{\partial J^a}. \quad (4.1)$$

We then expand this in the form appropriate to each regime. In the inspiral,

$$F^a \frac{\partial}{\partial J^a} = \varepsilon \left( F_{(0)}^\Omega \frac{\partial}{\partial \Omega} + F_{(1)}^\pm \frac{\partial}{\partial \delta M^\pm} \right) + \mathcal{O}(\varepsilon^2), \quad (4.2)$$

where  $F_{(1)}^\pm \partial_{\delta M^\pm} := F_{(1)}^+ \partial_{\delta M^+} + F_{(1)}^- \partial_{\delta M^-}$ . In the transition to plunge, with an obvious change of notation,

$$F^a \frac{\partial}{\partial \Delta J^a} = \varepsilon^{1/5} \left( F_{[0]}^{\Delta\Omega} + \varepsilon^{2/5} F_{[2]}^{\Delta\Omega} + \varepsilon^{3/5} F_{[3]}^{\Delta\Omega} + \varepsilon^{4/5} F_{[4]}^{\Delta\Omega} \right) \frac{\partial}{\partial \Delta\Omega} + \varepsilon F_{[3]}^\pm \frac{\partial}{\partial \delta M^\pm} + \mathcal{O}(\varepsilon^{6/5}). \quad (4.3)$$

Finally, in the plunge,

$$F^a \frac{\partial}{\partial J^a} = F_{\{0\}}^\Omega \frac{\partial}{\partial \Omega} + \varepsilon \left( F_{\{1\}}^\Omega \frac{\partial}{\partial \Omega} + F_{\{1\}}^\pm \frac{\partial}{\partial \delta M^\pm} \right) + \mathcal{O}(\varepsilon^2). \quad (4.4)$$

Focusing on the plunge, we can then write time derivatives as

$$\frac{\partial}{\partial s} = \Omega \frac{\partial}{\partial \phi_p} + F_{\{0\}}^\Omega (\Omega) \frac{\partial}{\partial \Omega} + \varepsilon F_{\{1\}}^a (J^b) \frac{\partial}{\partial J^a} + \mathcal{O}(\varepsilon^2), \quad (4.5)$$

assuming we work in phase-space coordinates  $(\phi_p, \Omega, \delta M^\pm)$ . If instead we work with  $(\phi_p, r_p, \delta M^\pm)$ , then  $\Omega$  is expanded, and time derivatives become

$$\frac{\partial}{\partial s} = \Omega_{\{0\}}(r_p) \frac{\partial}{\partial \phi_p} + F_{\{0\}}^r(r_p) \frac{\partial}{\partial r_p} + \varepsilon \left[ \Omega_{\{1\}}(J^a) \frac{\partial}{\partial \phi_p} + F_{\{1\}}^a(J^b) \frac{\partial}{\partial J^a} \right] + \mathcal{O}(\varepsilon^2). \quad (4.6)$$

When acting on a perturbation of the form (2.15), which has been expanded in discrete Fourier modes, we also use

$$\frac{\partial}{\partial \phi_p} = -im. \quad (4.7)$$

If  $s$  does *not* reduce to  $t$  along the particle's trajectory, then the above treatment must be generalized. Two options present themselves. First, one can formulate the orbital equations of motion directly in terms of  $s$ , defining an orbital frequency  $\Omega_s := d\phi_p/ds$  and deriving equations for the forcing functions appearing in  $dJ^a/ds$ . Alternatively, one can continue to use  $t$  as the parameter along the trajectory and account for how  $t$  changes with  $s$  when applying the chain rule for  $\partial/\partial s$ . In the next section, we take the latter approach. We define  $t_p(s)$  as the value of  $t$  on the particle at global time  $s$  (i.e. at the point where the slice of constant  $s$  intersects the trajectory). The definition  $s = t - \kappa(x)$  then implies  $t_p(s) = s + \kappa(x_p(t_p(s)))$ . Differentiating with respect to  $s$  and rearranging, we obtain

$$\frac{dt_p}{ds} = \frac{1}{1 - f^{-1}(r_p) H(r_p) \dot{r}_p}, \quad (4.8)$$

with  $H := d\kappa/dx$ ,  $r_p = r_p(t_p(s))$ , and  $\dot{r}_p = \dot{r}_p(t_p(s))$ . Equation (4.1) then becomes

$$\frac{\partial}{\partial s} = \frac{dt_p}{ds} \left[ \Omega \frac{\partial}{\partial \phi_p} + F^a(J^b, \varepsilon) \frac{\partial}{\partial J^a} \right]. \quad (4.9)$$

This formula holds in all regimes. Equation (4.8) is additionally expanded for small  $\varepsilon$ , using the expansion appropriate to the regime (inspiral, transition to plunge, or plunge) and to the choice of phase-space coordinates ( $\Omega$  or  $r_p$ ).

This treatment of time derivatives implies an expansion

$$\delta G_{\alpha\beta} [h^{\{n\}}] = \delta G_{\alpha\beta}^{\{0\}} [h^{\{n\}}] + \varepsilon \delta G_{\alpha\beta}^{\{1\}} [h^{\{n\}}] + \varepsilon^2 \delta G_{\alpha\beta}^{\{2\}} [h^{\{n\}}] + \mathcal{O}(\varepsilon^3). \quad (4.10)$$

Here  $h_{\alpha\beta}^{\{n\}} = \sum_{m=-\infty}^{\infty} h_{\alpha\beta}^{\{n\},m}(J^a, x^i) e^{-im\phi_p}$ , and the expansion in powers of  $\varepsilon$  is at fixed  $(\phi_p, J^a, x^i)$ . The leading term,  $\delta G_{\alpha\beta}^{\{0\}} [h^{\{n\}}]$  comes from omitting all order- $\varepsilon$  and higher terms in  $\partial/\partial s$ . Subleading terms arise from the subleading terms in  $\partial/\partial s$ . Analogously,

$$\delta^2 G_{\alpha\beta} [h^{\{1\}}, h^{\{1\}}] = \delta^2 G_{\alpha\beta}^{\{0\}} [h^{\{1\}}, h^{\{1\}}] + \varepsilon \delta^2 G_{\alpha\beta}^{\{1\}} [h^{\{1\}}, h^{\{1\}}] + \mathcal{O}(\varepsilon^2). \quad (4.11)$$

The stress-energy tensor (2.2) is likewise expanded at fixed phase-space coordinates:

$$T_{\alpha\beta} = m_p \frac{\tilde{u}_\alpha \tilde{u}_\beta}{\tilde{u}^t} \frac{\delta^3(x^i - x_p^i(\phi_p, J^a, \varepsilon))}{\sqrt{-\tilde{g}}} = \varepsilon T_{\alpha\beta}^{\{1\}} + \varepsilon^2 T_{\alpha\beta}^{\{2\}} + \mathcal{O}(\varepsilon^3). \quad (4.12)$$

Here  $T_{\alpha\beta}^{\{n\}}$  is a function of  $(\phi_p, J^a, x^i)$ , given at lowest order by

$$T_{\alpha\beta}^{\{1\}} = M g_{\alpha\mu} g_{\beta\nu} \dot{x}_{\{0\}}^\mu \dot{x}_{\{0\}}^\nu U_{\{0\}} \frac{\delta(r - r_{\{0\}})}{r_{\{0\}}^2} \delta(\theta - \pi/2) \delta(\phi - \phi_p), \quad (4.13)$$

with  $\dot{x}_{\{0\}}^\mu = \left(1, \frac{\partial r_{\{0\}}}{\partial \Omega} F_{\{0\}}^\Omega, 0, \Omega\right)$  in a fixed- $\Omega$  expansion or by  $\dot{x}_{\{0\}}^\mu = \left(1, F_{\{0\}}^r, 0, \Omega_{\{0\}}\right)$  in a fixed- $r_p$  expansion.

Once we have converted time derivatives to phase-space operators in this way, and suitably expanded the stress-energy tensor, we can then equate coefficients of powers of  $\varepsilon$  in the expansion of the Einstein equation (2.1), treating  $(\phi_p, J^a, x^i)$  as independent variables. This leads to a sequence of field equations for each  $h_{\alpha\beta}^{\{n\}}$ ,

$$\delta G_{\alpha\beta}^{\{0\}} [h^{\{1\}}] = 8\pi T_{\alpha\beta}^{\{1\}}, \quad (4.14a)$$

$$\delta G_{\alpha\beta}^{\{0\}} [h^{\{2\}}] = 8\pi T_{\alpha\beta}^{\{2\}} - \delta^2 G_{\alpha\beta}^{\{0\}} [h^{\{1\}}, h^{\{1\}}] - \delta G_{\alpha\beta}^{\{1\}} [h^{\{1\}}]. \quad (4.14b)$$

These are equations on a seven-dimensional product manifold made up of phase space (with coordinates  $\phi_p$  and  $J^a$ ) and space (with coordinates  $x^i$ ). However, the dimensionality is quickly reduced. By virtue of the background's stationarity, the Fourier basis functions  $e^{-im\phi_p}$  are eigenfunctions of  $\delta G_{\alpha\beta}^{\{0\}}$ , meaning the discrete Fourier modes decouple in  $\delta G_{\alpha\beta}^{\{0\}} [h^{\{n\}}]$ , reducing the problem to six-dimensional partial differential equations for each of the mode coefficients  $h_{\alpha\beta}^{\{n\},m}$ . Moreover, by virtue of the background's spherical symmetry, the angular dependence can be separated by expanding the metric perturbations in a basis of tensor or spin-weighted spherical harmonics [83, 84]. This reduces the equations to four-dimensional partial differential equations in  $J^a$  and  $r$  for each  $\ell m$  mode coefficient. Finally, we note that derivatives with respect to  $\delta M^\pm$  do not appear on the left-hand side of the field equations because they are suppressed by a power of  $\varepsilon$  in the chain rule (4.5). This means they only enter equation (4.14) in the source term  $\delta G_{\alpha\beta}^{\{1\}} [h^{\{1\}}]$ . Hence, the field equations are reduced to two-dimensional partial differential equations in  $\Omega$  (or  $r_p$ ) and  $r$ .

In the next section, we work through a concrete example of deriving mode-expanded phase-space field equations. We also describe how to recast them in the more familiar form of frequency-domain RWZ equations.

## 5. First-order RWZ waveforms

We restrict our attention to the first-order field equations for  $h_{\alpha\beta}^{\{1\}}$ . Rather than working directly with the Einstein equation (4.14a), we develop the associated RWZ equations for even- and odd-parity master functions. We first present the equations in a generic hyperboloidal slicing  $s$  and then restrict the analysis to our frequency-domain implementation in  $t$  slicing, obtained from transforming the phase-space equations to the frequency domain.

Our analysis in the remainder of the paper is restricted to 0PG order. For simplicity, we omit labels  $\{n\}$  on the field variables, sources, and linear operators, with the understanding that all quantities and operations are restricted to leading order.

### 5.1. RWZ equations and sources

We consider the RWZ equations describing the odd- and even-parity perturbations in Schwarzschild spacetime [85, 86],

$$(\partial_x^2 - \partial_t^2 - V_\ell^{e/o}(r)) \Psi_{\ell m}^{e/o}(t, r) = \tilde{S}_{\ell m}^{e/o}(t, r). \quad (5.1)$$

Following [87], in the odd-parity sector, we use the Cunningham-Price-Moncrief master function [88], while in the even-parity sector we use the Zerilli-Moncrief master function [89, 90]. The even- and odd-parity potentials are given by [87]

$$V_\ell^e(r) := \frac{f(r)}{r^2 \Lambda_\ell^2} \left[ 2\gamma_\ell^2 \left( \gamma_\ell + 1 + \frac{3M}{r} \right) + \frac{18M^2}{r^2} \left( \gamma_\ell + \frac{M}{r} \right) \right], \quad (5.2a)$$

$$V_\ell^o(r) := \frac{f(r)}{r^2} \left[ \ell(\ell + 1) - \frac{6M}{r} \right], \quad (5.2b)$$

where  $\gamma_\ell := (\ell + 2)(\ell - 1)/2$  and  $\Lambda_\ell(r) := \gamma_\ell + 3M/r$ . We write the master functions and the sources as

$$\Psi_{\ell m}^{e/o}(t, r) = R_{\ell m}^{e/o}(r_p(t_p(s)), r) e^{-im\phi_p(t_p(s))}, \quad \tilde{S}_{\ell m}^{e/o}(t, r) = S_{\ell m}^{e/o}(r_p(t_p(s)), r) e^{-im\phi_p(t_p(s))}, \quad (5.3)$$

where we recall that  $t_p$  is the value of  $t$  where the slice of constant  $s$  intersects the worldline,  $t_p = s + \kappa(x(r_p(s)))$ . The master functions are related to the tensor spherical-harmonic modes of  $h_{\alpha\beta}^{\{1\}}$  through a linear operation, in which we neglect subleading terms in the chain rules (4.5) and (4.9); likewise, the sources are constructed through a linear operation on  $T_{\alpha\beta}^{\{1\}}$ , again neglecting subleading terms in the chain rule.

Our primary goal in this paper is to calculate the GW strain, which is expressed in terms of the two GW polarizations as the limit  $r \rightarrow \infty$  of the expression

$$r(h_+ - ih_\times) = rh_{\mu\nu} \bar{m}^\mu \bar{m}^\nu = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{\ell} r h_{\bar{m}\bar{m}}^{\ell m} {}_{-2}Y_{\ell m}(\theta, \phi). \quad (5.4)$$

Here  $\bar{m}^\mu = \frac{1}{\sqrt{2r}}(0, 0, 1, -i \csc \theta)$  and  ${}_{-2}Y_{\ell m}$  is a spin-weighted spherical harmonic. The limit is taken at fixed  $(\phi_p, J^a)$ , assuming slices of global time  $s$  smoothly connect to future null infinity. We define  $h_{\ell m} := \lim_{r \rightarrow \infty} (r h_{\bar{m}^\mu}^{\ell m})$  and write the asymptotic  $\ell m$  mode of the waveform as [28]

$$h_{\ell m} = \varepsilon \frac{\sqrt{D_\ell}}{2} (\Psi_{\ell m}^e - i \Psi_{\ell m}^o) = \varepsilon \frac{\sqrt{D_\ell}}{2} (R_{\ell m}^e - i R_{\ell m}^o) e^{-im\phi_p} := H_{\ell m} e^{-im\phi_p} := |H_{\ell m}| e^{-i\Phi_{\ell m}}, \quad (5.5)$$

where  $D_\ell := (\ell - 1)\ell(\ell + 1)(\ell + 2)$  and  $\Phi_{\ell m} = m\phi_p - \arg(H_{\ell m})$ . Only  $m \geq 0$  modes need to be computed as the  $m < 0$  modes are deduced from the usual symmetry relation  $h_{\ell, -m} = (-1)^m h_{\ell m}^*$ .

We allow a generic choice of time  $s$ , which might not reduce to  $t$  on the particle trajectory. By plugging equation (5.3) into equation (5.1) and applying the chain rule (4.9), we put the RWZ equations in the following form:

$$\begin{aligned} & (\partial_x^2 - V_\ell^{e/o}) R_{\ell m}^{e/o} - 2H \frac{dt_p}{ds} (\dot{r}_{\{0\}} \partial_x \partial_{r_p} - im \Omega_{\{0\}} \partial_x) R_{\ell m}^{e/o} - \frac{dH}{dx} \frac{dt_p}{ds} (\dot{r}_{\{0\}} \partial_{r_p} - im \Omega_{\{0\}}) R_{\ell m}^{e/o} \\ & + (1 - H^2) \left[ im \Omega_{\{0\}} \frac{d^2 t_p}{ds^2} + (m^2 \Omega_{\{0\}}^2 + im \partial_{r_p} \Omega_{\{0\}} \dot{r}_{\{0\}}) \left( \frac{dt_p}{ds} \right)^2 \right. \\ & \left. - \dot{r}_{\{0\}} \frac{d^2 t_p}{ds^2} \partial_{r_p} - \dot{r}_{\{0\}}^2 \left( \frac{dt_p}{ds} \right)^2 \partial_{r_p}^2 - (\ddot{r}_{\{0\}} - 2im \Omega_{\{0\}} \dot{r}_{\{0\}}) \left( \frac{dt_p}{ds} \right)^2 \partial_{r_p} \right] R_{\ell m}^{e/o} = S_{\ell m}^{e/o}. \end{aligned} \quad (5.6)$$

The quantities  $dt_p/ds$  and  $d^2 t_p/ds^2$  are functions of  $r_p$ , as given in equation (4.8) (and the  $s$  derivative thereof). The geodesic quantities  $\Omega_{\{0\}}(r_p)$ ,  $\dot{r}_{\{0\}}(r_p)$  and  $\ddot{r}_{\{0\}}(r_p)$  are given in equations (3.40)–(3.42), respectively.

While the original, time-domain RWZ equations are hyperbolic, our phase-space equations (5.6) have an unusual character. They are hyperbolic for all  $2M < r_p < 6M$ , as is easily confirmed by calculating the discriminant, which is proportional to  $(\dot{r}_{\{0\}} \frac{dt_p}{ds})^2 > 0$  for any  $H$ . However, this discriminant vanishes in the limit to  $r_p = 6M$ , since  $\dot{r}_{\{0\}}$  vanishes in that limit. One might conclude that the equations are consequently parabolic at  $r_p = 6M$ , but  $\dot{r}_{\{0\}}(6M) = 0$  implies they actually reduce to radial ordinary differential equations there. Fundamentally, this singular behaviour is a consequence of the fact that the plunging geodesic asymptotes to  $r_p = 6M$  in the infinite past but never actually reaches  $6M$ ; the physical domain for the plunge solution is the open region  $2M < r_p < 6M$ . To obtain the physically correct solution in that domain, one must ensure that the solution appropriately matches to the transition-to-plunge solution; we discuss this further in the next section.

In the remainder of this section and the practical implementation of first-order waveforms in this paper, we will consider  $t$  slicing,  $\kappa(x) = H(x) = 0$ . Appendix D provides the parallel development of the field equations in generic slicing, which are important in establishing the validity of our approach. With  $t$  slicing, the RWZ equations become

$$\begin{aligned} & \left[ \partial_x^2 + (m^2 \Omega_{\{0\}}^2 + im \partial_{r_p} \Omega_{\{0\}} \dot{r}_{\{0\}}) - \dot{r}_{\{0\}}^2 \partial_{r_p}^2 \right. \\ & \left. - (\ddot{r}_{\{0\}} - 2im \Omega_{\{0\}} \dot{r}_{\{0\}}) \partial_{r_p} - V_\ell^{e/o}(r) \right] R_{\ell m}^{e/o}(r_p, r) = S_{\ell m}^{e/o}(r_p, r). \end{aligned} \quad (5.7)$$

The source appearing on the right-hand side of this equation is that of the plunging point-particle. We construct this source in the even and odd sectors using the formalism described in appendix C of [87]. We write

$$\begin{aligned} \tilde{S}_{\ell m}^e(t, r) & := \frac{1}{(\gamma_\ell + 1) \Lambda_\ell} \left[ r^2 f (f^2 \partial_r Q_{\ell m}^{tt} - \partial_r Q_{\ell m}^{rr}) + r (\Lambda_\ell - f) Q_{\ell m}^{rr} + r f^2 Q_{\ell m}^b \right. \\ & \left. - \frac{f^2}{r \Lambda_\ell} (\gamma_\ell (\gamma_\ell - 1) r^2 + (4\gamma_\ell - 9) Mr + 15M^2) Q_{\ell m}^{tt} \right] + \frac{2f}{\Lambda_\ell} Q_{\ell m}^r - \frac{f}{r} Q_{\ell m}^\#, \end{aligned} \quad (5.8a)$$

$$\tilde{S}_{\ell m}^o(t, r) := \frac{rf}{\gamma_\ell} \left( \frac{1}{f} \partial_t P_{\ell m}^r + f \partial_r P_{\ell m}^t + \frac{2M}{r^2} P_{\ell m}^t \right), \quad (5.8b)$$

with

$$P_{\ell m}^a(t, r) := \frac{16\pi r^2}{\ell(\ell+1)} \int T_{\{1\}}^{aB}(t, r, \theta, \phi) X_B^{\ell m*}(\theta, \phi) \sin\theta d\theta d\phi, \tag{5.9a}$$

$$Q_{\ell m}^a(t, r) := \frac{16\pi r^2}{\ell(\ell+1)} \int T_{\{1\}}^{aB}(t, r, \theta, \phi) Y_B^{\ell m*}(\theta, \phi) \sin\theta d\theta d\phi, \tag{5.9b}$$

$$Q_{\ell m}^{ab}(t, r) := 8\pi \int T_{\{1\}}^{ab}(t, r, \theta, \phi) Y^{\ell m*}(\theta, \phi) \sin\theta d\theta d\phi, \tag{5.9c}$$

$$Q_{\ell m}^b(t, r) := 8\pi r^2 \int T_{\{1\}}^{AB}(t, r, \theta, \phi) \Omega_{AB} Y^{\ell m*}(\theta, \phi) \sin\theta d\theta d\phi, \tag{5.9d}$$

$$Q_{\ell m}^\#(t, r) := 32\pi r^4 \frac{(\ell-2)!}{(\ell+2)!} \int T_{\{1\}}^{AB}(t, r, \theta, \phi) Y_{AB}^{\ell m*}(\theta, \phi) \sin\theta d\theta d\phi. \tag{5.9e}$$

In these expressions, lowercase Latin letters stand for the coordinates  $\{t, r\}$ , while uppercase Latin letters indicate the coordinates  $\{\theta, \phi\}$ . The scalar ( $Y^{\ell m}$ ), vector ( $Y_A^{\ell m}$  and  $X_A^{\ell m}$ ) and tensor ( $Y_{AB}^{\ell m}$  and  $X_{AB}^{\ell m}$ ) spherical harmonics are defined in appendix C, and  $\Omega_{AB} = \text{diag}(1, \sin^2\theta)$  is the metric on the unit 2-sphere.

We evaluate the integrals in equation (5.9) using the point-particle stress-energy tensor (4.13). The differential operator  $\partial_t$  in equation (5.8) should therefore be understood as the operator  $\dot{r}_{\{0\}} \partial_{r_p} + \Omega_{\{0\}} \partial_{\phi_p}$ . Recalling equation (5.3), we can then write the sources (5.8) as

$$S_{\ell m}^{e/o}(r_p, r) = A_{\ell m}^{e/o}(r_p) \delta(r - r_p) + B_{\ell m}^{e/o}(r_p) \partial_r \delta(r - r_p), \tag{5.10}$$

where

$$\begin{aligned} A_{\ell m}^c(r_p) := & -\frac{2\sqrt{2}\pi M(r_p - 2M)^2}{(\gamma_\ell + 1)r_p^7 \Lambda_\ell(r_p)^2} \left[ 324M^4 + 36M^3(3 + 8\gamma_\ell)r_p + 9M^2(1 + 4\gamma_\ell^2)r_p^2 \right. \\ & \left. + 2M(1 + 9\gamma_\ell)r_p^3 + 3\gamma_\ell(1 + \gamma_\ell)r_p^4 \right] Y^{\ell m}(\pi/2, 0) \\ & - \frac{16\sqrt{6}\pi M^2(r_p - 2M)^2(6M - r_p)^{3/2}}{\ell(\ell+1)r_p^{11/2} \Lambda_\ell(r_p)} Y_\phi^{\ell m*}(\pi/2, 0) \\ & - \frac{288\sqrt{2}\pi M^3(\ell-2)!(r_p - 2M)^2}{(\ell+2)!r_p^5} Y_{\phi\phi}^{\ell m*}(\pi/2, 0), \end{aligned} \tag{5.11a}$$

$$\begin{aligned} A_{\ell m}^o(r_p) := & -\frac{4\sqrt{3}\pi M^2(r_p - 2M)^2}{\ell(\ell+1)\gamma_\ell r_p^7} \left[ 756M^3 - 6\sqrt{3}iMm r_p^{1/2}(6M - r_p)^{3/2} \right. \\ & \left. - 216M^2 r_p + 45M r_p^2 + 7r_p^3 \right] X_\phi^{\ell m*}(\pi/2, 0), \end{aligned} \tag{5.11b}$$

$$B_{\ell m}^c(r_p) := \frac{6\sqrt{2}\pi M(r_p - 2M)^3(r_p^2 + 12M^2)}{(\gamma_\ell + 1)r_p^5 \Lambda_\ell(r_p)} Y^{\ell m}(\pi/2, 0), \tag{5.11c}$$

$$B_{\ell m}^o(r_p) := \frac{36\sqrt{3}\pi M^2(r_p - 2M)^3(r_p^2 + 12M^2)}{\ell(\ell+1)\gamma_\ell r_p^6} X_\phi^{\ell m*}(\pi/2, 0). \tag{5.11d}$$

Here  $Y_\phi^{\ell m*}(\pi/2, 0) = -imY^{\ell m}(\pi/2, 0)$  and  $Y_{\phi\phi}^{\ell m}(\pi/2, 0) = \left(-m^2 + \frac{\ell(\ell+1)}{2}\right) Y^{\ell m}(\pi/2, 0)$ . The even- and odd-parity sectors are sourced by respectively even and odd  $\ell + m$  modes since

$$S_{\ell m}^c(r_p, r) = 0 \quad \text{for } \ell + m \text{ odd}, \tag{5.12a}$$

$$S_{\ell m}^o(r_p, r) = 0 \quad \text{for } \ell + m \text{ even}. \tag{5.12b}$$

### 5.2. Punctured RWZ equations

As highlighted above, the plunge field equations become singular at  $r_p = 6M$ . Our physical boundary condition is that at early times, in a neighbourhood of the singular surface  $r_p = 6M$ , our plunge solution must asymptotically match the transition-to-plunge solution.

This asymptotic matching was discussed at the level of the orbital dynamics in section 3.2. For the metric perturbations (2.13) and (2.15), the matching condition applies at the level of Fourier coefficients  $h_{\alpha\beta}^{[n],m}$  and  $h_{\alpha\beta}^{\{n\},m}$ . At 0PG order, the condition is that if (i)  $h_{\alpha\beta}^{\{1\},m}(\Omega, x^i)$  is expanded for small  $(\Omega - \Omega_*)$ ,

and (ii)  $h_{\alpha\beta}^{[n],m}(\Delta\Omega, x^i) = h_{\alpha\beta}^{[n],m}((\Omega - \Omega_*)/\varepsilon^{2/5}, x^i)$  is expanded for small  $\varepsilon$  at fixed  $\Omega$ , then (iii) the small- $(\Omega - \Omega_*)$  expansion of  $\varepsilon h_{\alpha\beta}^{\{1\},m}(\Omega, x^i)$  must agree, term by term, with the linear-in- $\varepsilon$  terms in the re-expansion of  $\sum_{n=5}^{\infty} \varepsilon^{n/5} h_{\alpha\beta}^{[n],m}$ . Here and in many expressions below, we suppress dependence on  $\delta M^\pm$  for simplicity.

From equation (2.13), we see that the re-expansion of the transition-to-plunge metric takes the form

$$h_{\alpha\beta} = \sum_{m=-\infty}^{\infty} \sum_{n=5}^{\infty} \sum_{k=5-n}^{\infty} \varepsilon^{n/5} \frac{\varepsilon^{k/5}}{(\Omega - \Omega_*)^{k/2}} h_{\alpha\beta}^{[n,k],m}(\delta M^\pm, x^i) e^{-im\phi_p}. \tag{5.13}$$

For  $\Omega > \Omega_*$ , this re-expansion for small  $\varepsilon$  at fixed  $\Omega$  is equivalent to an expansion in the limit  $\Delta\Omega \rightarrow +\infty$  (i.e. the ‘late transition to plunge’). The expansion contains a power series in  $\Delta\Omega^{1/2}$  (or, equivalently,  $(\Omega - \Omega_*)^{1/2}/\varepsilon^{1/5}$ ) rather than  $\Delta\Omega$  due to the structure of the orbital forcing functions in the transition to plunge<sup>8</sup>. Here we have restricted the range of  $k$  using the fact that this must match the plunge expansion, which begins at linear order in  $\varepsilon$ ; hence,  $(k + n)/5 \geq 1$ . Moreover,  $(k + n)/5$  must be an integer, equal to the order  $\{n\}$  in the plunge metric, implying  $h_{\alpha\beta}^{[n,k],m}$  vanishes except when  $k = (5\{n\} - n)$  for integers  $\{n\} \geq 1$ . Hence, the terms that must match the near-ISCO expansion of  $h_{\alpha\beta}^{\{1\}}$  are  $1/(\Omega - \Omega_*)^{(5-n)/2}$ : a constant term  $\propto (\Omega - \Omega_*)^0$  from  $h_{\alpha\beta}^{[5],m}$ , a linear term  $\propto (\Omega - \Omega_*)$  from  $h_{\alpha\beta}^{[7],m}$ , a term  $\propto (\Omega - \Omega_*)^{3/2}$  from  $h_{\alpha\beta}^{[8],m}$ , and so on.

From the above analysis, we see that the leading term in the near-ISCO expansion of  $h_{\alpha\beta}^{\{1\},m}$  must be a constant, equal to  $h_{\alpha\beta}^{[5,0],m}$ . In fact, the analysis in [51] shows that the first two terms in the transition-to-plunge regime are exactly equal to the first two terms in the near-ISCO expansion of the first-order inspiral solution:

$$\varepsilon h_{\alpha\beta}^{[5],m} + \varepsilon^{7/5} h_{\alpha\beta}^{[7],m} = \varepsilon \left[ h_{\alpha\beta}^{\{1\},m}(\Omega_*, x^i) + (\Omega - \Omega_*) \partial_\Omega h_{\alpha\beta}^{\{1\},m}(\Omega_*, x^i) \right]. \tag{5.14}$$

The first two terms in the near-ISCO expansion of  $h_{\alpha\beta}^{\{1\},m}$  must agree with these two terms from the inspiral solution. This immediately extends to the RWZ master variables as functions of  $r_p$ :

$$R_{\ell m}^{e/o}(r_p, r) = R_{\ell m}^{e/o(1)}(6M, r) + (r_p - 6M) \partial_{r_p} R_{\ell m}^{e/o(1)}(6M, r) + \mathcal{O} \left[ (r_p - 6M)^2 \right], \tag{5.15}$$

where  $R_{\ell m}^{e/o(1)}(r_p, r)$  are the first-order master functions in the inspiral.

To enforce this near-ISCO matching to the transition-to-plunge solution, we adopt a puncture scheme [70]. We emphasize the puncture in this case is distinct from the one associated with the point particle singularity discussed in section 2.1. Instead, as explained in [70], we consider a puncture scheme as a more general method of enforcing the correct singular boundary conditions that arise from matched asymptotic expansions when extending an approximate solutions outside its domain of validity; in this case, the boundary conditions we enforce are the singularity conditions at the ISCO, which arise from extending the PG expansion to its unphysical early-time limit.

Concretely, in a region near the ISCO, we write the physical field as the sum of two terms,

$$R_{\ell m}^{e/o}(r_p, r) = R_{\ell m}^{e/o\mathcal{R}}(r_p, r) + R_{\ell m}^{e/o\mathcal{P}}(r_p, r). \tag{5.16}$$

The residual field  $R_{\ell m}^{e/o\mathcal{R}}$  has vanishing boundary conditions at the ISCO,  $R_{\ell m}^{e/o\mathcal{R}}(6M, r) = 0$ . The puncture field  $R_{\ell m}^{e/o\mathcal{P}}$  consists of the near-ISCO solution (5.15) truncated at some order in  $(r_p - 6M)$ , and it ‘lives’ until an orbital radius  $r_{\mathcal{P}}$  such that  $2M < r_{\mathcal{P}} < 6M$ . As an example, here we truncate at leading order, meaning

$$R_{\ell m}^{e/o\mathcal{P}}(r_p, r) := R_{\ell m}^{e/o(1)}(6M, r) \theta(r_p - r_{\mathcal{P}}). \tag{5.17}$$

<sup>8</sup> The reason for this can be understood as follows: in the transition-to-plunge expansion, the Fourier coefficients  $h_{\alpha\beta}^{[n],m}$  can always be written as a sum of terms that are factored into a  $\Delta\Omega$ -dependent and a  $\Delta\Omega$ -independent piece [51]. The  $\Delta\Omega$ -dependent factors consist in integer powers of  $\Delta\Omega$  itself, the forcing terms  $F_{[n]}^{\Delta\Omega}$  and their  $\Delta\Omega$  derivatives. For example, in the notation of [51], the 4PLT metric perturbation is given by  $h_{\alpha\beta}^{[9],m}(\Delta\Omega, x^i) = \Delta\Omega^2 h_{\alpha\beta}^{[9]A,m}(x^i) + F_{[0]}^{\Delta\Omega} \partial_{\Delta\Omega} F_{[0]}^{\Delta\Omega} h_{\alpha\beta}^{[9]B,m}(x^i)$ . In the large- $\Delta\Omega$  limit, the forcing terms  $F_{[n]}^{\Delta\Omega}$  admit asymptotic solutions in half-integer powers of  $\Delta\Omega$ , as can be seen from equation (3.27) or table 2.

The inspiral field  $R_{\ell m}^{e/o(1)}(6M, r)$  satisfies equation (5.7) evaluated at  $r_p = 6M$ . To understand this, recall that  $\dot{r}_{\{0\}}(6M) = 0 = \partial_{r_p} \dot{r}_{\{0\}}(6M)$  (from equation (3.41)), implying that the limit of equation (5.7) to  $r_p = 6M$  is simply the RWZ equation for a particle on a circular orbit at the ISCO:

$$\left[ \partial_x^2 + m^2 \Omega_{\{0\}}^2 \Big|_* - V_\ell^{e/o}(r) \right] R_{\ell m}^{e/o(1)}(6M, r) = S_{\ell m}^{e/o}(6M, r), \quad (5.18)$$

where  $|_*$  indicates evaluation at  $r_p = 6M$ . We now plug equation (5.16) into equation (5.7), move the puncture fields to the right-hand side and use equation (5.18). We then obtain the following equation for the residual field:

$$\begin{aligned} & \left[ \partial_x^2 + \left( m^2 \Omega_{\{0\}}^2 + im \partial_{r_p} \Omega_{\{0\}} \dot{r}_{\{0\}} \right) - \dot{r}_{\{0\}}^2 \partial_{r_p}^2 \right. \\ & \left. - \left( \ddot{r}_{\{0\}} - 2im \Omega_{\{0\}} \dot{r}_{\{0\}} \right) \partial_{r_p} - V_\ell^{e/o}(r) \right] R_{\ell m}^{e/o \mathcal{R}}(r_p, r) = S_{\ell m}^{e/o \text{ eff}}(r_p, r). \end{aligned} \quad (5.19)$$

The effective source,  $S_{\ell m}^{e/o \text{ eff}}$ , can be split into two pieces: the ordinary point particle source in the region  $r_p < r_{\mathcal{P}}^-$  and a more complicated, extended source in the punctured region  $6M \geq r_p \geq r_{\mathcal{P}}^-$ , where we use the notation  $r_{\mathcal{P}}^- = r_{\mathcal{P}} - 0^+$  to indicate that the interval  $[r_{\mathcal{P}}^-, 6M]$  encloses the point  $r_p = r_{\mathcal{P}}$ . Explicitly, we write

$$S_{\ell m}^{e/o \text{ eff}}(r_p, r) = S_{\ell m}^{e/o \text{ pp}}(r_p, r) + S_{\ell m}^{e/o \text{ ext}}(r_p, r), \quad (5.20)$$

where  $S_{\ell m}^{e/o \text{ pp}}$  is the source outside the punctured region,

$$S_{\ell m}^{e/o \text{ pp}}(r_p, r) := S_{\ell m}^{e/o}(r_p, r) \theta(r_{\mathcal{P}} - r_p), \quad (5.21)$$

and  $S_{\ell m}^{e/o \text{ ext}}$  is the source inside the punctured region,

$$\begin{aligned} S_{\ell m}^{e/o \text{ ext}}(r_p, r) & := \theta(r_p - r_{\mathcal{P}}) \left[ S_{\ell m}^{e/o}(r_p, r) - S_{\ell m}^{e/o}(6M, r) \right] \\ & - \theta(r_p - r_{\mathcal{P}}) \left( m^2 \Omega_{\{0\}}^2 - m^2 \Omega_{\{0\}}^2 \Big|_* + im \partial_{r_p} \Omega_{\{0\}} \dot{r}_{\{0\}} \right) R_{\ell m}^{e/o(1)}(6M, r) \\ & - \delta(r_p - r_{\mathcal{P}}) \dot{r}_{\{0\}} \Big|_{\mathcal{P}} \left( 2im \Omega_{\{0\}} \Big|_{\mathcal{P}} + \partial_{r_p} \dot{r}_{\{0\}} \Big|_{\mathcal{P}} \right) R_{\ell m}^{e/o(1)}(6M, r) \\ & + \partial_{r_p} \delta(r_p - r_{\mathcal{P}}) \dot{r}_{\{0\}}^2 \Big|_{\mathcal{P}} R_{\ell m}^{e/o(1)}(6M, r). \end{aligned} \quad (5.22)$$

We use  $|_{\mathcal{P}}$  to indicate quantities evaluated at  $r_p = r_{\mathcal{P}}$ . In arriving at this result we have used  $\dot{r}_{\{0\}}^2(r_p) \partial_{r_p} \delta(r_p - r_{\mathcal{P}}) = \dot{r}_{\{0\}}^2(r_{\mathcal{P}}) \partial_{r_p} \delta(r_p - r_{\mathcal{P}}) - \partial_{r_p} \dot{r}_{\{0\}}^2(r_{\mathcal{P}}) \delta(r_p - r_{\mathcal{P}})$ .

The essential feature of the extended source is that it vanishes as  $r_p \rightarrow 6M$  (i.e. in the infinite past), consistent with our initial condition that the residual field vanishes there; the total solution (residual field plus puncture) then reduces to the transition-to-plunge solution, as desired. We also explain in the next section how this behaviour of the effective source eliminates ill-defined integrals that would otherwise appear.

Since the puncture vanishes for  $r_p < r_{\mathcal{P}}$ , the residual field reduces to the physical field in that region:

$$R_{\ell m}^{e/o \mathcal{R}}(r_p, r) = R_{\ell m}^{e/o}(r_p, r) \quad \text{for } r_p < r_{\mathcal{P}}. \quad (5.23)$$

We also note that the Dirac  $\delta$  and  $\delta'$  terms in equation (5.22) can be discarded and replaced with the junction condition that

$$\lim_{r_p \rightarrow r_{\mathcal{P}}^-} R_{\ell m}^{e/o}(r_p, r) = \lim_{r_p \rightarrow r_{\mathcal{P}}^+} \left[ R_{\ell m}^{e/o \mathcal{R}}(r_p, r) + R_{\ell m}^{e/o \mathcal{P}}(r_p, r) \right]. \quad (5.24)$$

Moreover, although we said that the punctured region  $[r_{\mathcal{P}}^-, 6M]$  is ‘near the ISCO’, the physical solution (5.16) must be independent of  $r_{\mathcal{P}}$ , so long as  $2M < r_{\mathcal{P}} < 6M$ . This independence follows from the same type of arguments as in [70]. Finally, we comment that we have only worked with a leading-order puncture in this section, but a higher-order (in  $(r_p - 6M)$ ) puncture could be required. It is possible for the puncture to be sufficiently high order to obtain a convergent solution for the residual field, without being high enough order to obtain the *correct* solution; see again [70]. We leave investigation of this, and derivation of the required order of the puncture, to future work.

### 5.3. Frequency-domain equations

In order to solve equation (5.19) for the residual field we work in the frequency domain. This allows us to use well-known Green-function techniques to solve the problem.

We start by defining the following forward and inverse transforms:

$$\hat{g}(\omega, r) = \int_{2M}^{6M} \frac{dr_p}{\dot{r}_{\{0\}}(r_p)} g(r_p, r) e^{i\omega t_G(r_p) - im\phi_G(r_p)}, \tag{5.25a}$$

$$g(r_p, r) = -\frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega \hat{g}(\omega, r) e^{-i\omega t_G(r_p) + im\phi_G(r_p)}, \tag{5.25b}$$

with  $t_G$  and  $\phi_G$  defined in equations (3.43) and (3.44), respectively. For simplicity we have set  $t_0 = \phi_0 = 0$ , noting that  $t_0$  and  $\phi_0$  cancel between the forward and inverse transforms. We can verify that the above transforms are self-consistent in the range of interest  $2M \leq r_p \leq 6M$ :

$$\begin{aligned} g(r_p, r) &= -\frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega e^{-i\omega t_G(r_p) + im\phi_G(r_p)} \int_{2M}^{6M} \frac{dr'_p}{\dot{r}_{\{0\}}(r'_p)} g(r'_p, r) e^{i\omega t_G(r'_p) - im\phi_G(r'_p)} \\ &= \frac{1}{2\pi} \int_{2M}^{6M} \frac{dr'_p}{|\dot{r}_{\{0\}}(r'_p)|} g(r'_p, r) e^{im(\phi_G(r_p) - \phi_G(r'_p))} \int_{-\infty}^{+\infty} d\omega e^{i\omega(t_G(r'_p) - t_G(r_p))} \\ &= \int_{2M}^{6M} \frac{dr'_p}{|\dot{r}_{\{0\}}(r'_p)|} g(r'_p, r) e^{im(\phi_G(r_p) - \phi_G(r'_p))} |\dot{r}_{\{0\}}(r'_p)| \delta(r'_p - r_p) = g(r_p, r), \end{aligned} \tag{5.26}$$

where we have used  $\dot{r}_{\{0\}} = -|\dot{r}_{\{0\}}|$  since  $\dot{r}_{\{0\}} \leq 0$  between the ISCO and the event horizon.

We now take the transform of equation (5.19). This transform is delicate because we arrive at ill-defined boundary terms if we apply the transform naively and integrate by parts to move derivatives onto the exponential. If we naively discard those boundary terms, we obtain

$$(\partial_x^2 + \omega^2 - V_\ell^{e/o}(r)) \hat{R}_{\ell m}^{e/o \mathcal{R}}(\omega, r) = \hat{S}_{\ell m}^{e/o \text{eff}}(\omega, r), \tag{5.27}$$

where  $\hat{S}_{\ell m}^{e/o \text{eff}}$  is the transform of  $S_{\ell m}^{e/o \text{eff}}$ . To properly justify this equation, we treat each side of equation (5.19) as a tempered distribution and adopt the standard definition  $\langle \varphi, \mathcal{F}[f] \rangle := \langle \mathcal{F}[\varphi], f \rangle$ , where  $f$  is any tempered distribution,  $\mathcal{F}$  is our forward transform,  $\varphi$  is any element of the Schwartz space, and  $\langle \varphi, f \rangle$  is the integral over the real line (for integrable  $f$ ) or the action of  $f$  on  $\varphi$  (for distributional  $f$ ). Writing equation (5.19) schematically as  $(\partial_x^2 + \mathcal{D})R_{\ell m} = S_{\ell m}^{\text{eff}}$ , we apply the forward transform to obtain  $\langle \mathcal{F}[\varphi], (\partial_x^2 + \mathcal{D})R_{\ell m} \rangle = \langle \varphi, \mathcal{F}[S_{\ell m}^{\text{eff}}] \rangle$ . On the left, we apply the standard definition  $\langle \varphi, \mathcal{D}f \rangle := \langle \mathcal{D}^\dagger \varphi, f \rangle$  for any linear operator  $\mathcal{D}$ , where  $\mathcal{D}^\dagger$  is the adjoint of  $\mathcal{D}$ . Defining the frequency-domain operator  $\mathcal{D}_\omega^\dagger$  via  $\mathcal{D}^\dagger \mathcal{F}[\varphi] = \mathcal{F}[\mathcal{D}_\omega^\dagger \varphi]$ , we obtain  $\langle \mathcal{F}[\varphi], \mathcal{D}R_{\ell m} \rangle = \langle \mathcal{F}[\mathcal{D}_\omega^\dagger \varphi], R_{\ell m} \rangle = \langle \mathcal{D}_\omega^\dagger \varphi, \mathcal{F}[R_{\ell m}] \rangle = \langle \varphi, \mathcal{D}_\omega \mathcal{F}[R_{\ell m}] \rangle$ . Therefore,  $(\partial_x^2 + \mathcal{D}_\omega) \mathcal{F}[R_{\ell m}] = \mathcal{F}[S_{\ell m}^{\text{eff}}]$  as a distribution, which is simply equation (5.27).

Equation (5.27) has the standard form of the frequency-domain RWZ equation. Our method of reaching this form applies equally well at higher PG orders. At all orders, we use  $t_G$  and  $\phi_G$  in the transforms, rather than  $t_p$  and  $\phi_p$ , since we require known functions of  $r_p$ ;  $t_p$  and  $\phi_p$  are never determined until the online waveform-generation stage.

### 5.4. Inhomogeneous solutions

We solve equation (5.27) using the standard Green-function method (or equivalently, variation of parameters), building a retarded Green function from a basis of homogeneous solutions.

In the limits  $r \rightarrow +\infty$  and  $r \rightarrow 2M$  (or, equivalently,  $x \rightarrow \pm\infty$ ), the homogeneous solutions to equation (5.27) are complex exponentials of the form  $e^{\pm i\omega x}$ . We consider two independent homogeneous solutions to equation (5.27), the so-called ‘in’ and ‘up’ solutions, with the following asymptotic behaviours:

$$\hat{R}_\ell^{e/o \text{ in}}(\omega, x) \sim \begin{cases} A_\ell^{e/o \text{ in}}(\omega) e^{-i\omega x} + A_\ell^{e/o \text{ out}}(\omega) e^{+i\omega x} & \text{as } r \rightarrow +\infty (x \rightarrow +\infty), \\ e^{-i\omega x} & \text{as } r \rightarrow 2M (x \rightarrow -\infty), \end{cases} \tag{5.28a}$$

$$\hat{R}_\ell^{e/o \text{ up}}(\omega, x) \sim \begin{cases} e^{i\omega x} & \text{as } r \rightarrow +\infty (x \rightarrow +\infty), \\ B_\ell^{e/o \text{ in}}(\omega) e^{-i\omega x} + B_\ell^{e/o \text{ out}}(\omega) e^{+i\omega x} & \text{as } r \rightarrow 2M (x \rightarrow -\infty). \end{cases} \tag{5.28b}$$

Here and below we freely write functions of  $r$  as functions of  $x(r)$ . According to the Chandrasekhar–Detweiler transformation between the Regge–Wheeler and Zerilli–Moncrief solutions [91], the homogeneous even and odd ‘in’ solution are related by

$$(D_\ell - 12iM\omega) \hat{R}_\ell^{e\text{ in}}(\omega, x) = \left[ D_\ell + \frac{72M^2 f(r)}{(\ell-1)(\ell+2)r^2 + 6Mr} + 12M \frac{\partial}{\partial x} \right] \hat{R}_\ell^{o\text{ in}}(\omega, x). \quad (5.29)$$

As a consequence, the coefficients  $A_\ell^{e/o\text{ in}}(\omega)$  and  $A_\ell^{e/o\text{ out}}(\omega)$  obey the relations

$$A_\ell^{e\text{ in}}(\omega) = A_\ell^{o\text{ in}}(\omega), \quad A_\ell^{e\text{ out}}(\omega) = \frac{D_\ell + 12iM\omega}{D_\ell - 12iM\omega} A_\ell^{o\text{ out}}(\omega). \quad (5.30)$$

From here on we will therefore omit the label e/o on  $A_\ell^{e/o\text{ in}}(\omega)$ . As a consequence of the RWZ equation we have  $\hat{R}_\ell^{e/o\text{ in}}(-\omega^*, x) = \hat{R}_\ell^{e/o\text{ in}^*}(\omega, x)$ . This implies

$$A_\ell^{\text{in}}(-\omega^*) = A_\ell^{\text{in}^*}(\omega), \quad A_\ell^{e/o\text{ out}}(-\omega^*) = A_\ell^{e/o\text{ out}^*}(\omega). \quad (5.31)$$

Using the two independent homogeneous solutions (5.28) we can construct the following Green function:

$$\hat{G}_\ell^{e/o}(\omega, x, x') = \frac{1}{W_\ell(\omega)} \left[ \theta(x-x') \hat{R}_\ell^{e/o\text{ in}}(\omega, x') \hat{R}_\ell^{e/o\text{ up}}(\omega, x) + \theta(x'-x) \hat{R}_\ell^{e/o\text{ in}}(\omega, x) \hat{R}_\ell^{e/o\text{ up}}(\omega, x') \right], \quad (5.32)$$

where  $W_\ell$  is the Wronskian

$$W_\ell(\omega) = \hat{R}_\ell^{e/o\text{ in}}(\omega, x) \partial_x \hat{R}_\ell^{e/o\text{ up}}(\omega, x) - \hat{R}_\ell^{e/o\text{ up}}(\omega, x) \partial_x \hat{R}_\ell^{e/o\text{ in}}(\omega, x), \quad (5.33)$$

which we anticipate is identical for the even and odd sector. This Green function satisfies

$$(\partial_x^2 + \omega^2 - V_\ell^{e/o}) \hat{G}_\ell^{e/o}(\omega, x, x') = \delta(x-x'). \quad (5.34)$$

It is straightforward to verify that  $W_\ell$  is independent of the field point where it is computed (since  $\partial_x W_\ell = 0$ ). We can therefore compute it in the limit  $r \rightarrow +\infty$  using the asymptotic solutions (5.28), yielding

$$W_\ell(\omega) = 2i\omega A_\ell^{\text{in}}(\omega). \quad (5.35)$$

This justifies  $W_\ell$  having no e/o label. Evaluating the Wronskian in the limit  $r \rightarrow 2M$ , we deduce  $B_\ell^{e/o\text{ out}}(\omega) = A_\ell^{\text{in}}(\omega)$ . The inhomogeneous solution to equation (5.27) can then be obtained as

$$\begin{aligned} \hat{R}_{\ell m}^{e/o\mathcal{R}}(\omega, r) &= \int_{-\infty}^{+\infty} dx' \hat{G}_\ell^{e/o}(\omega, x, x') \hat{S}_{\ell m}^{e/o\text{ eff}}(\omega, x') \\ &= \frac{\hat{R}_\ell^{e/o\text{ up}}(\omega, r)}{W_\ell(\omega)} \int_{2M}^r \frac{dr'}{f(r')} \hat{R}_\ell^{e/o\text{ in}}(\omega, r') \hat{S}_{\ell m}^{e/o\text{ eff}}(\omega, r') \\ &\quad + \frac{\hat{R}_\ell^{e/o\text{ in}}(\omega, r)}{W_\ell(\omega)} \int_r^{+\infty} \frac{dr'}{f(r')} \hat{R}_\ell^{e/o\text{ up}}(\omega, r') \hat{S}_{\ell m}^{e/o\text{ eff}}(\omega, r'). \end{aligned} \quad (5.36)$$

As written, the integrals over the (transform of the) extended effective source (5.21) fail to converge due to the source’s behaviour at  $r = 2M$  and  $r \rightarrow \infty$ ; this is an artefact of  $t$  slicing, which we address in appendix D.

The goal of our implementation in this paper is to construct the 0PG asymptotic waveform. Moreover, since we are interested in the merger and ringdown and not in very early times when the particle is near the ISCO, we restrict our attention to the region  $r_p < r_{\mathcal{P}}$ , outside the puncture window. In this region the physical and residual fields are identical, and we can omit the superscript  $\mathcal{R}$  on the field; see equation (5.23). To obtain the waveform for  $r_p < r_{\mathcal{P}}$ , we now focus on the only source in that

region: the point-particle portion of the effective source,  $S_{\ell m}^{\text{e/o PP}}$ . At large distances the field generated by this source,  $\hat{R}_{\ell m}^{\text{e/o}}|_{\text{pp}\infty}(\omega) := \hat{R}_{\ell m}^{\text{e/o}}|_{\text{pp}}(\omega, r \rightarrow \infty)$ , is given by

$$\begin{aligned} \hat{R}_{\ell m}^{\text{e/o}}|_{\text{pp}\infty}(\omega) &= \frac{e^{i\omega x}}{W_{\ell}(\omega)} \int_{2M}^{\infty} \frac{dr'}{f(r')} \hat{R}_{\ell}^{\text{e/o in}}(\omega, r') \hat{S}_{\ell m}^{\text{e/o PP}}(\omega, r') \\ &= \frac{e^{i\omega x}}{W_{\ell}(\omega)} \int_{2M}^{\infty} \frac{dr'}{f(r')} \hat{R}_{\ell}^{\text{e/o in}}(\omega, r') \int_{2M}^{r_{\mathcal{P}}} \frac{dr_p}{\dot{r}_{\{0\}}(r_p)} e^{i\omega t_G(r_p) - im\phi_G(r_p)} S_{\ell m}^{\text{e/o PP}}(r_p, r'). \end{aligned} \quad (5.37)$$

Here we have used the asymptotic behaviour of the ‘up’ solution (5.28). We evaluate the  $r_p$  integral using the properties of the Dirac delta function, recalling that  $\delta'(r' - r_p) = -\partial_{r_p} \delta(r' - r_p)$ . We are then left with

$$\hat{R}_{\ell m}^{\text{e/o}}|_{\text{pp}\infty}(\omega) = \frac{e^{i\omega x}}{2i\omega A_{\ell}^{\text{in}}(\omega)} \int_{2M}^{r_{\mathcal{P}}} \frac{dr'}{\dot{r}_{\{0\}}(r')} e^{i\omega t_G(r') - im\phi_G(r')} K_{\ell m}^{\text{e/o}}(\omega, r'), \quad (5.38)$$

where

$$K_{\ell m}^{\text{e/o}}(\omega, r) = \frac{A_{\ell m}^{\text{e/o}}(r) \hat{R}_{\ell}^{\text{e/o in}}(\omega, r)}{f(r)} - B_{\ell m}^{\text{e/o}}(r) \frac{d}{dr} \left( \frac{\hat{R}_{\ell}^{\text{e/o in}}(\omega, r)}{f(r)} \right). \quad (5.39)$$

The solution on phase space is obtained after taking the inverse transform (5.25b):

$$\begin{aligned} \Psi_{\ell m}^{\text{e/o}}|_{\text{pp}\infty}(\phi_p, r_p) &= R_{\ell m}^{\text{e/o}}|_{\text{pp}\infty}(r_p) e^{-im\phi_p} \\ &= -\frac{1}{2\pi} \left[ \int_{-\infty}^{+\infty} d\omega e^{-i\omega[t_G(r_p) - x]} \frac{C_{\ell m}^{\text{e/o}}(\omega)}{2i\omega A_{\ell}^{\text{in}}(\omega)} \right] e^{im[\phi_G(r_p) - \phi_p]}, \end{aligned} \quad (5.40)$$

where

$$C_{\ell m}^{\text{e/o}}(\omega) := \int_{2M}^{r_{\mathcal{P}}} \frac{dr'}{\dot{r}_{\{0\}}(r')} e^{i\omega t_G(r') - im\phi_G(r')} K_{\ell m}^{\text{e/o}}(\omega, r'). \quad (5.41)$$

This becomes a time-domain waveform when we substitute the solutions  $\phi_p(t)$  and  $r_p(t)$  to the orbital equations of motion,  $\Psi_{\ell m}^{\text{e/o}}|_{\text{pp}\infty}(t) = \Psi_{\ell m}^{\text{e/o}}|_{\text{pp}\infty}(\phi_p(t), r_p(t))$ . At geodesic order, the difference  $\phi_G(r_p(t)) - \phi_p(t)$  is a constant. At 1PG and higher orders, equation (5.40) still represents the leading-order waveform, but  $\phi_p(t)$  and  $r_p(t)$  become  $\varepsilon$  dependent, and  $\phi_p(t)$  no longer cancels  $\phi_G(r_p(t))$ ; this incomplete cancellation will be true even at leading, 0PG order if  $\phi_p(t)$  is obtained from an equation of motion that is hybridized with the transition-to-plunge dynamics.

We can now point to the singular behaviour cured by the puncture. If there were no puncture, the integral (5.41) would run from  $2M$  to  $6M$ . That integral would not converge, due to the fact that  $\dot{r}_{\{0\}}(r') \sim (6M - r')^{3/2}$  near the ISCO. Obtaining the final physical solution requires adding the piece of the field sourced by the effective source (5.22) in the puncture region, call it  $R_{\ell m}^{\text{e/o}\mathcal{R}}|_{\text{ext}}$ . However, as shown in appendix E, a large part of  $R_{\ell m}^{\text{e/o}\mathcal{R}}|_{\text{ext}}$  vanishes in the limit  $r_{\mathcal{P}} \rightarrow 6M$ , and the remainder makes a small contribution to the final waveform for values of  $r_p$  sufficiently far below  $r_{\mathcal{P}}$ ; this is to be expected because any small interval of  $r_p$  near  $6M$  corresponds to an asymptotically large interval of coordinate time. As a first approximation, we therefore place  $r_{\mathcal{P}}$  very near the ISCO, at  $r_{\mathcal{P}} = 5.999M$ , and we neglect  $R_{\ell m}^{\text{e/o}\mathcal{R}}|_{\text{ext}}$ . We expect that a complete implementation of the puncture scheme will become more important at higher PG orders, as the early-time behaviour becomes more singular at each successive order.

As a final remark in this section, we highlight that equation (5.40) does not precisely yield the waveform at future null infinity. This is because we work with  $t$  as our time coordinate. Our waveform therefore represents the coefficient of  $1/r$  in a large- $r$  expansion of  $h_{\ell m}^{\{1\}}$  as a function of  $t$  along a timelike surface at large  $r$ . A consequence is the explicit appearance of the tortoise coordinate  $x$  in equation (5.40), which makes our waveform phase depend on the value of  $r$  at which the waveform is extracted. Such a dependence is degenerate with a choice of initial time, meaning it has no physical consequence in the present paper. A more complete treatment will use hyperboloidal slicing or the sharp null slicing displayed in figure 1 to unambiguously obtain the waveform at future null infinity.

### 5.5. SPA

We can immediately extract qualitative information from the overall structure of the waveform (5.40). The waveform comes with a total phase factor  $e^{-i\omega t_G(r_p) + im\phi_G(r_p) - im\phi_p}$ . At early times, the particle is near the ISCO on a very nearly circular orbit. We therefore expect the waveform to be strongly peaked around the orbital frequency  $\omega = m\Omega$  at these early times, such that the first two terms in the exponential cancel, leaving an overall phase factor  $e^{-im\phi_p}$ ; in words, the waveform phase closely mimics the orbital phase. On the other hand, at very late times, when the particle has plunged deep behind the effective potential and is near the black hole horizon, we expect the waveform to be dominated by QNM ringdown frequencies. In that case,  $\omega t_G(r_p)$  is unrelated to  $\phi_G(r_p)$ ; instead, the final two terms in the exponential cancel, leaving  $e^{-i\omega t_G(r_p)}$ .

In this section we apply a SPA that makes the above reasoning about the early-time behaviour more precise. We develop the complementary late-time approximation in the next section.

We can rewrite the  $r_p$ -domain solution (the coefficient of  $e^{-im\phi_p}$  in equation (5.40)) as

$$R_{\ell m}^{e/o}|_{\text{pp}\infty}(r_p) = \int_{-\infty}^{+\infty} d\omega \int_{2M}^{r_p} dr' g(\omega, r') e^{i\varphi(\omega, r', r_p)}, \quad (5.42)$$

where we have defined

$$g(\omega, r') := -\frac{1}{2\pi \dot{r}_{\{0\}}(r')} \frac{K_{\ell m}^{e/o}(\omega, r')}{2i\omega A_{\ell}^{\text{in}}(\omega)}, \quad (5.43)$$

$$\varphi(\omega, r', r_p) := \omega [t_G(r') - t_G(r_p)] - m [\phi_G(r') - \phi_G(r_p)]. \quad (5.44)$$

For simplicity, here we have set  $x=0$ . We notice that equation (5.42) has the form of a two-dimensional integral against a rapidly oscillating complex exponential, especially as the particle is close to the ISCO; recall that  $t_G(r_p \rightarrow 6M) \rightarrow -\infty$ . We can therefore evaluate the integral using a leading-order SPA,

$$R_{\ell m}^{e/o}|_{\text{pp}\infty}(r_p) \approx \frac{2\pi g(\omega, r')}{|\det(\text{Hess}(\varphi))|^{1/2}} e^{i\varphi(\omega, r', r_p) + i\pi\sigma/4} \Big|_{(\omega, r')=\text{stationary point}}, \quad (5.45)$$

where the stationary point  $(\omega, r') = (m\Omega_{\{0\}}(r_p), r_p)$  is obtained from the condition  $\nabla\varphi = (\partial_\omega\varphi, \partial_{r'}\varphi) = 0$ .  $\text{Hess}(\varphi)$  is the Hessian of  $\varphi$ . Its signature,  $\sigma$ , vanishes, and its determinant evaluates to  $|\det(\text{Hess}(\varphi))|^{1/2} = 1/|\dot{r}_{\{0\}}(r_p)| = -1/\dot{r}_{\{0\}}(r_p)$  since  $\dot{r}_{\{0\}}$  is negative; see equation (3.41). The leading-order SPA to equation (5.42) then gives the final solution

$$\Psi_{\ell m}^{e/o}|_{\text{pp}\infty}(\phi_p, r_p) \approx \left[ \frac{K_{\ell m}^{e/o}(\omega, r')}{2i\omega A_{\ell}^{\text{in}}(\omega)} \right]_{\omega=m\Omega_{\{0\}}(r_p), r'=r_p} e^{-im\phi_p}. \quad (5.46)$$

Equation (5.46) is an ‘instantaneous’ representation of the waveform, in which the state of the binary system is instantaneously transmitted to infinity. It relies on the waveform frequency and amplitude changing slowly relative to the waveform phase, just as in the inspiral and transition-to-plunge regimes. In section 6.2 we show that the SPA is a good approximation almost up to merger.

### 5.6. QNM sum

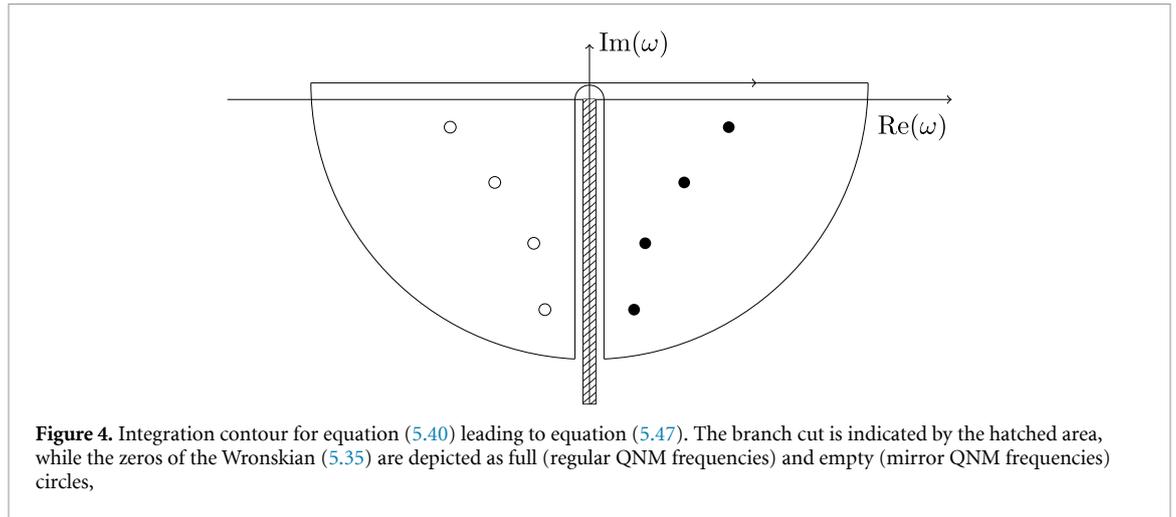
At late times, we expect our OPG solution to be well described by a sum of QNM modes. In this section we describe how to uniquely extract the amplitudes of these modes.

We can evaluate the real-line  $\omega$  integral in equation (5.40) by expressing it in terms of residues at poles in the complex  $\omega$  plane. Figure 4 shows the analyticity properties of the Green function, which has a branch cut along the negative imaginary axis and poles in the lower half plane; the poles, at zeros of the Wronskian (5.35), correspond to the QNM frequencies. Following the standard procedure [92], we define a contour that runs along the real axis and closes in the lower half plane, with a detour around the branch cut. By virtue of the residue theorem, the integral along the real line in equation (5.40) is equal to the sum of residues of the enclosed poles, minus the integral along the high-frequency arc and the integral around the branch cut:

$$\Psi_{\ell m}^{e/o}|_{\text{pp}\infty} = \Psi_{\ell m}^{e/o}|_{\text{QNM}} - \Psi_{\ell m}^{e/o}|_{\text{bc}} - \Psi_{\ell m}^{e/o}|_{\text{arc}}. \quad (5.47)$$

Here we focus on the QNM contribution and leave assessing the impact of the branch-cut and arc pieces on the waveform to future work. Using the residue theorem, we then obtain

$$\Psi_{\ell m}^{e/o}|_{\text{QNM}} = i \sum_{n=0}^{\infty} \text{Res} \left[ \frac{C_{\ell m}^{e/o}(\omega)}{2i\omega A_{\ell}^{\text{in}}(\omega)} e^{-i\omega[t_G(r_p) - x]} \right]_{\omega=\omega_{\ell n}, -\omega_{\ell n}^*} e^{im[\phi_G(r_p) - \phi_p]}. \quad (5.48)$$



A minus sign due to the clockwise orientation of the contour cancels the overall minus sign in equation (5.40).  $\omega_{\ell n}$  with  $n = 0, 1, 2, \dots$  are the QNM frequencies, for which  $A_\ell^{\text{in}}(\omega_{\ell n}) = 0$ . The fundamental mode is indicated by  $n = 0$ , while  $n \geq 1$  label the  $n$ th overtone. Note that for all ‘regular’ frequencies  $\omega_{\ell n}$  (living in the quadrant where  $\text{Re}(\omega_{\ell n}) > 0$ ), we also need to include ‘mirror’ frequencies  $-\omega_{\ell n}^*$ , since the QNM frequency spectrum is symmetric with respect to the imaginary axis.

Computing the residues explicitly and multiplying and dividing by  $A_\ell^{\text{e/o out}}(\omega_{\ell n})$ , we can write

$$\Psi_{\ell m}^{\text{e/o}}|_{\text{QNM}} = \sum_{n=0}^{\infty} \left[ E_{\ell m}^{\text{e/o}}(\omega) e^{-i\omega(t_G(r_p) - x)} \right]_{\omega=\omega_{\ell n}, -\omega_{\ell n}^*} e^{im[\phi_G(r_p) - \phi_p]}, \tag{5.49}$$

where  $E_{\ell m}^{\text{e/o}}(\omega_{\ell n}) := B_\ell^{\text{e/o}}(\omega_{\ell n}) D_{\ell m}^{\text{e/o}}(\omega_{\ell n})$  are the so-called quasinormal excitation coefficients, given by the product of the quasinormal excitation factors  $B_\ell^{\text{e/o}}(\omega_{\ell n})$  [93–95] and the coefficients  $D_{\ell m}^{\text{e/o}}(\omega_{\ell n})$  defined as

$$B_\ell^{\text{e/o}}(\omega_{\ell n}) := \frac{1}{2\omega_{\ell n}} \frac{A_\ell^{\text{e/o out}}(\omega_{\ell n})}{\frac{dA_\ell^{\text{in}}}{d\omega}(\omega_{\ell n})}, \quad D_{\ell m}^{\text{e/o}}(\omega_{\ell n}) := \frac{C_{\ell m}^{\text{e/o}}(\omega_{\ell n})}{A_\ell^{\text{e/o out}}(\omega_{\ell n})}. \tag{5.50}$$

While the excitation factors only depend on the Schwarzschild geometry, the coefficients  $D_{\ell m}^{\text{e/o}}(\omega_{\ell n})$  depend on the nature of the perturbation, that is, in our case, on the plunging point-particle source. As a consequence of equation (5.30), the even and odd excitation factors are related by

$$B_\ell^{\text{o}}(\omega_{\ell n}) = \frac{D_\ell + 12iM\omega_{\ell n}}{D_\ell - 12iM\omega_{\ell n}} B_\ell^{\text{e}}(\omega_{\ell n}). \tag{5.51}$$

Furthermore, the properties (5.31) imply that

$$B_\ell^{\text{e/o}}(-\omega_{\ell n}^*) = [B_\ell^{\text{e/o}}(\omega_{\ell n})]^*. \tag{5.52}$$

The integrand in  $C_{\ell m}^{\text{e/o}}(\omega_{\ell n})$  (5.41) is exponentially convergent at the upper limit when  $r_p \approx 6M$ , while at the horizon it behaves as  $(r' - 2M)^{-4iM\omega_{\ell n}}$ . The integral therefore converges when  $\text{Im}(\omega_{\ell n}) > -1/(4M)$ , which only holds for the fundamental QNM frequencies. In order to compute the excitation coefficients for the overtones we consider the following regularization procedure [57, 96]:

$$C_{\ell m}^{\text{e/o}}|_{\text{reg}}(\omega) := \int_{2M}^{r_p} dr' \left[ \frac{e^{i\omega t_G(r') - im\phi_G(r')}}{\dot{r}_{\{0\}}(r')} K_{\ell m}^{\text{e/o}}(\omega, r') - q_{\ell m}^{\text{e/o}}(\omega, r') \right] + Q_{\ell m}^{\text{e/o}}(\omega, r_p). \tag{5.53}$$

Here we have introduced an auxiliary function  $q_{\ell m}^{\text{e/o}}(\omega, r')$ ,

$$q_{\ell m}^{\text{e/o}}(\omega, r') = (r' - 2M)^{-4iM\omega} \left[ q_{0\ell m}^{\text{e/o}} + q_{1\ell m}^{\text{e/o}}(\omega)(r' - 2M) + q_{2\ell m}^{\text{e/o}}(\omega)(r' - 2M)^2 + \dots \right], \tag{5.54}$$

where the coefficients  $q_{i\ell m}^{\text{e/o}}$  with  $i = 0, 1, 2, \dots$  are obtained from the near-horizon expansion of the integrand in equation (5.41). Subtracting  $q_{\ell m}^{\text{e/o}}$  from the integrand ensures that the  $r'$  integral is finite. This regularized integral is then added to the antiderivative of  $q_{\ell m}^{\text{e/o}}$ ,  $Q_{\ell m}^{\text{e/o}}(\omega, r) = \int^r dr' q_{\ell m}^{\text{e/o}}(\omega, r')$  (neglecting

any constant arising from the integration), evaluated at the integral's upper boundary,  $r = r_p$ . We justify this regularization procedure in appendix F.

Like for the SPA, we explore the accuracy of the QNM sum in section 6.2.

## 6. Numerical implementation and comparisons

In this section we present our numerical implementation to obtain first-order plunge waveforms from the formalism presented in section 5.4, discussing and validating our implementation choices. We also outline the procedure to obtain SPA and QNM waveforms following sections 5.5 and 5.6, and assess how well they approximate the full waveform in their respective regimes of validity. Finally, we compare the OPG waveforms to NR simulations. Our numerical implementation is freely accessible in the ancillary material [97].

### 6.1. Implementation and validation

We start by describing our numerical implementation to obtain first-order (OPG) waveforms from equation (5.40), with  $r_p = 5.999M$ . Computing the waveforms consists in two major steps: we first tabulate the quantity  $C_{\ell m}^{e/o}(\omega)$  (5.41) on a grid of  $\omega$  values. The final waveform is then obtained as a function of  $t_G(r_p) - x$  by evaluating equation (5.40) using a discrete Fourier transform. We almost exclusively present the time-domain waveform obtained using the geodesic orbital dynamics,  $r_p(t) = r_{\{0\}}(t)$  and  $\phi_p(t) = \phi_G(r_p(t))$ . In this case, using the freedom to absorb  $x$  into a choice of initial time, we are able to rename ' $t_G(r_p) - x$ ' simply as 'geodesic time'  $t_G$ . Our complete time-domain waveform (5.40) is then

$$\Psi_{\ell m}^{e/o}|_{\text{pp}\infty}(t_G) = -\frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega e^{-i\omega t_G} \frac{C_{\ell m}^{e/o}(\omega)}{2i\omega A_{\ell}^{\text{in}}(\omega)}; \quad (6.1)$$

our early-time SPA waveform (5.46) is

$$\Psi_{\ell m}^{e/o}|_{\text{pp}\infty}(t_G) \approx \left[ \frac{K_{\ell m}^{e/o}(\omega, r_{\{0\}}(t_G))}{2i\omega A_{\ell}^{\text{in}}(\omega)} \right]_{\omega=m\Omega_{\{0\}}(t_G)} e^{-im\phi_G(t_G)}; \quad (6.2)$$

and our late-time QNM waveform (5.49) is

$$\Psi_{\ell m}^{e/o}|_{\text{QNM}}(t_G) = \sum_{n=0}^{\infty} [E_{\ell m}^{e/o}(\omega) e^{-i\omega t_G}]_{\omega=\omega_{\ell n}, -\omega_{\ell n}^*}. \quad (6.3)$$

All numerical results are obtained setting  $M = 1$ .

We first construct the odd-parity 'in' solution, which goes into  $C_{\ell m}^{e/o}(\omega)$  (as defined in equation (5.41)) through the function  $K_{\ell m}^o$ . We use Mathematica's built-in HeunC function (available in Mathematica version 12.1 or higher) as

$$\hat{R}_{\ell}^{\text{in}}(\omega, r) = \frac{2M}{r} e^{-i(4M-r)\omega} \left( \frac{r}{2M} - 1 \right)^{-2iM\omega} \text{HeunC}[\ell + \ell^2 - 2 - 8iM\omega - 16M^2\omega^2, -4M\omega(4M\omega - i), 1 - 4iM\omega, -3, -4iM\omega, 1 - \frac{r}{2M}]. \quad (6.4)$$

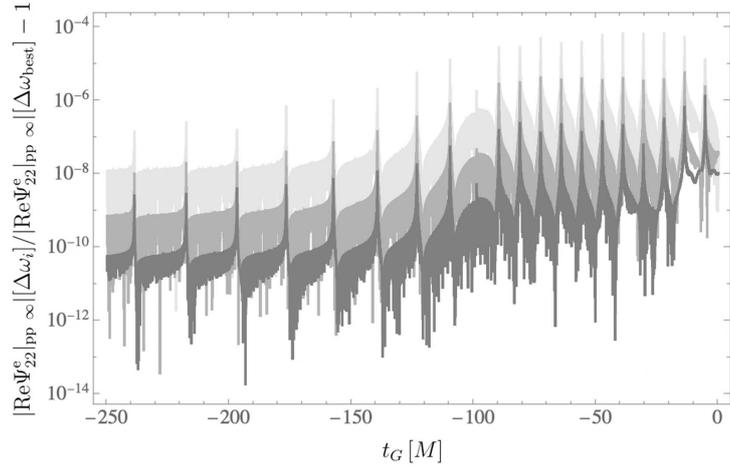
The even-parity 'in' solution is then obtained from the odd-parity one using the relation (5.29).

We have numerically checked that the definition (6.4) matches with the current implementations of the 'in' solution in the ReggeWheeler package within the Black Hole Perturbation Toolkit [98] (`ReggeWheelerRadial[2, \ell, \omega, Method -> "HeunC"] ["In"] [r]`). Our 'in' solution differs from the one defined in [96] by an overall factor,  $\hat{R}_{\ell}^{\text{in}}(\omega, r) = e^{-2iM\omega} \hat{R}_{\ell}^{\text{in}}(\omega, r)|_{\text{Leaver}}$ . We get high-precision values for  $A_{\ell}^{\text{in}}(\omega)$  in the denominator of equation (6.1) using the ReggeWheeler package.

Most of the computation time is spent in the radial numerical integration at a given value of  $\omega$ , which defines  $C_{\ell m}^{e/o}(\omega)$ . Through trial and error, we settled on the following integration strategy:

$$\int_2^{5.999} = \int_2^{2.1} |_{\text{GK}} + \int_{2.1}^3 |_{\text{GK}} + \int_3^4 |_{\text{Levin}} + \int_4^5 |_{\text{Levin}} + \int_5^{5.999} |_{\text{Levin}}. \quad (6.5)$$

The individual integrals are performed either using the Gauss–Kronrod rule (GK) with 50 points or the Levin rule (Levin) with 50 points as options for Mathematica's `NIntegrate`. We set a timeout of 5000 seconds for each such integral. We found that for all  $2 \leq \ell \leq 12$ ,  $-\ell \leq m \leq \ell$  ( $m \neq 0$ ), and  $\omega$  in the range  $[-4, 4]$  these integrals were performed successfully.



**Figure 5.** Relative difference between the absolute value of the real part of the waveforms  $|\text{Re}\Psi_{22}^{e/o}|_{pp\infty}[\Delta\omega_{\text{best}}]$  for the  $(\ell, m) = (2, 2)$  mode with  $\omega_{\text{cutoff}} = 4M^{-1}$  and sampling interval  $\Delta\omega = \Delta\omega_{\text{best}} := 1/(8000M)$  as compared with the corresponding waveform  $|\text{Re}\Psi_{22}^{e/o}|_{pp\infty}[\Delta\omega_i]$  with sampling  $\Delta\omega_1 = 1/(1000M)$  (light gray),  $\Delta\omega_2 = 1/(2000M)$  (gray) and  $\Delta\omega_3 = 1/(4000M)$  (dark gray).

We tabulate  $C_{\ell m}^{e/o}(\omega)$  on an evenly spaced grid of  $\omega$  values with spacing  $\Delta\omega = 10^{-3}M^{-1}$  in the range  $\omega \in [-\omega_{\text{cutoff}}, \omega_{\text{cutoff}}]$  with  $\omega_{\text{cutoff}} = 4M^{-1}$ . In order to attenuate the effects of the frequency cutoff on the Fourier transform, we used the low-pass filter described in equation (16) of [99] with parameters  $(x_0, w, q, s) = (-4, 0.2, -3.9, 1)$  in the negative  $\omega$  range, and a symmetrically defined low-pass filter in the positive  $\omega$  range. With this filter, there are only marginally small high-frequency residual oscillations that remain visible in the deep QNM regime; see figure 10. As a benchmark of the implementation, the maximal value for  $\text{Re}\Psi_{\ell m}^{e/o}(t_G)$  is reached for the  $(\ell, m) = (2, 2)$  mode at  $t_{\text{peak}} \approx -104.1M$  where  $\Psi_{22}^{e/o}(t_{\text{peak}}) \approx (0.5877 + 0.0137i)M$ .

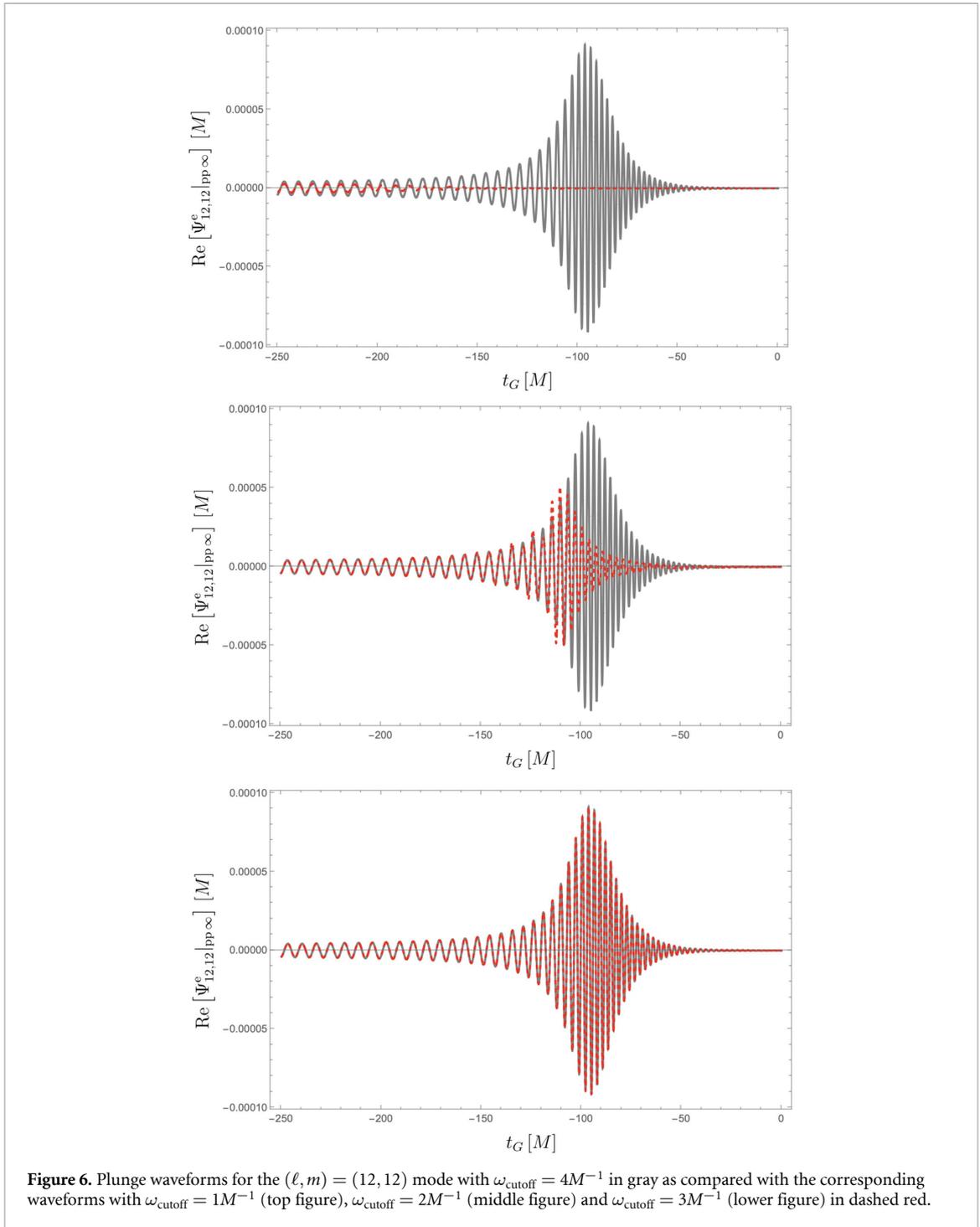
We have investigated how the choice of  $\Delta\omega$  impacts the waveforms. We found that using a spacing  $\Delta\omega = 1/(500M)$  leads to qualitatively incorrect waveforms because the peak of the  $(\ell, m) = (2, 2)$  mode is not captured. In order to test the accuracy of the waveforms with our choice of  $\Delta\omega = 1/(1000M)$ , we have computed the relative numerical difference between waveforms with different sampling intervals and a benchmark waveform with sampling interval  $\Delta\omega = 1/(8000M)$ . The result is displayed in figure 5 for the  $(\ell, m) = (2, 2)$  mode: the waveform with  $\Delta\omega = 1/(1000M)$  has a relative precision of  $10^{-7}$  with respect to the waveform with  $\Delta\omega = 1/(8000M)$ . The precision scales globally according  $\Delta\omega$ : shortening the interval by a factor 2 increases the accuracy by a factor 10.

In order to study the frequency-cutoff dependence, we computed the relative error of the real amplitude as a function of  $\omega_{\text{cutoff}}$  as

$$\mathcal{E}_{\ell m}(\omega'_{\text{cutoff}}) = \frac{1}{\text{Re}\Psi_{22}^{e/o}(t_G)|_{\omega_{\text{cutoff}}=4/M}} \max_{t_G} \left( \text{Re}\Psi_{\ell m}^{e/o}(t_G)|_{\omega_{\text{cutoff}}=\omega'_{\text{cutoff}}} - \text{Re}\Psi_{\ell m}^{e/o}(t_G)|_{\omega_{\text{cutoff}}=4/M} \right). \quad (6.6)$$

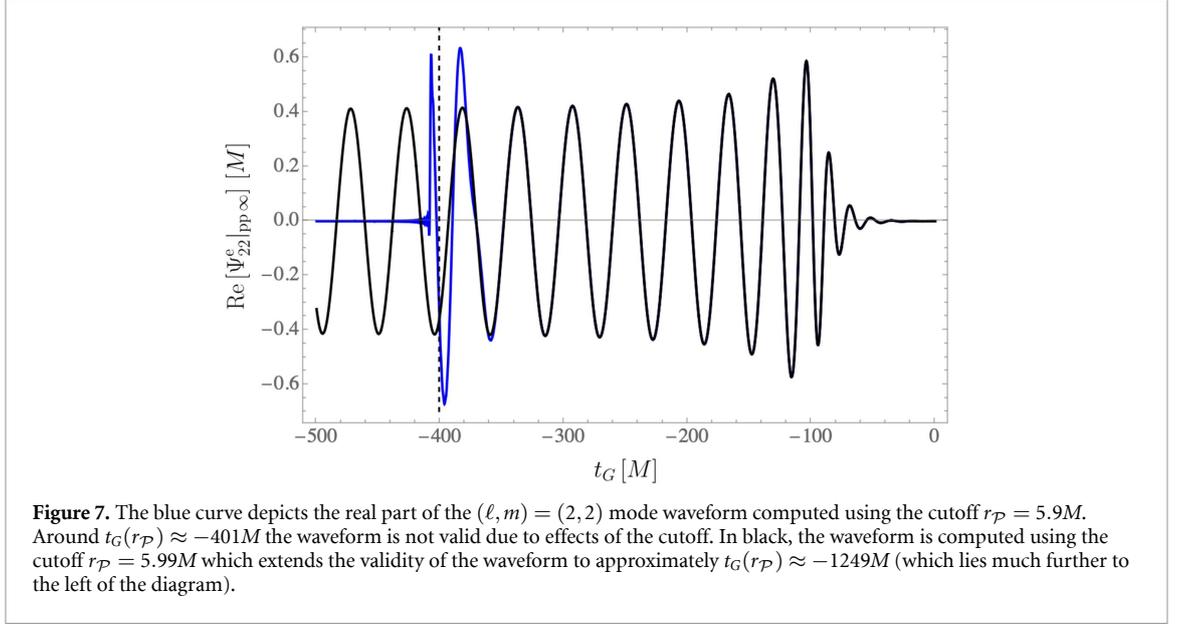
We numerically obtained that  $\mathcal{E}_{\ell m}(\omega_{\text{cutoff}} = 2M^{-1}) < 10^{-4}$  for all  $2 \leq \ell \leq 12$  and  $-\ell \leq m \leq \ell$ ,  $m \neq 0$  modes except  $(\ell, m) = (11, 11)$  where  $\mathcal{E}_{1111}(2M^{-1}) \approx 2.1 \times 10^{-4}$ . In that sense, it is barely sufficient to set the frequency cutoff to  $\omega_{\text{cutoff}} = 2M^{-1}$  in order to obtain accurate waveforms with a  $10^{-4}$  precision. The cutoff  $\omega_{\text{cutoff}} = 1M^{-1}$  only allows us to have waveform with a precision of  $10^{-2}$ : the maximal value of  $\mathcal{E}_{\ell m}(1)$  is reached for the  $(\ell, m) = (5, 5)$  mode with  $\mathcal{E}_{55}(1M^{-1}) \approx 6.3 \times 10^{-3}$ . The analysis of the error at the cutoff  $\omega_{\text{cutoff}} = 3M^{-1}$  allows us to provide an estimate of the accuracy of our waveforms which are defined for  $\omega_{\text{cutoff}} = 4M^{-1}$ . We find  $\mathcal{E}_{\ell m}(\omega_{\text{cutoff}} = 3M^{-1}) < 10^{-6}$  for all  $2 \leq \ell \leq 12$  and  $-\ell \leq m \leq \ell$ ,  $m \neq 0$  modes except  $(\ell, m) = (4, 4)$  where  $\mathcal{E}_{44}(3M^{-1}) \approx 1.1 \times 10^{-6}$ . In that sense, setting the cutoff to  $4M^{-1}$  leads to a loss of precision less than  $10^{-6}$ . The higher the value of  $\ell$ , the higher the cutoff should be in order to qualitatively capture the peak of the waveform. This is illustrated in figure 6. Setting the cutoffs at  $1M^{-1}$  or  $2M^{-1}$  does not allow us to correctly capture the  $(\ell, m) = (12, 12)$  waveform, while it is correctly captured using a cutoff at  $3M^{-1}$ . We learn that even if the precision is  $10^{-4}$  at cutoff value  $2M^{-1}$ , we need a larger cutoff frequency in order to qualitatively capture the peak of the waveforms at high  $\ell$ . After this analysis we conclude that our waveforms are qualitatively and quantitatively accurate upon setting  $\omega_{\text{cutoff}} = 4M^{-1}$ .

We also investigated the effects of changing the radial cutoff  $r_{\mathcal{P}}$ . In figure 7 we plot the waveform corresponding to  $r_{\mathcal{P}} = 5.9M$  and  $r_{\mathcal{P}} = 5.99M$ . In the case  $r_{\mathcal{P}} = 5.9M$ , we observe that the waveform fails



to capture the correct behaviour around the time corresponding to the cutoff  $t_G(r_p) \approx -401M$ , due to our omission of the field generated by the effective source in the punctured region. Putting a higher cutoff allows us to better capture the early-time behaviour since the geodesic gets closer and closer to the ISCO located at  $r_p = 6M$  as  $t_G \rightarrow -\infty$ . For  $r_p = 5.99M$ , the waveforms fail around  $t_G = -1249M$ , while for  $r_p = 5.999$  they fail around  $t_G = -3944M$ . In the range  $-750M < t_G < 100M$ , the change in the waveform upon changing the cutoff from  $r_p = 5.999M$  to  $r_p = 5.99M$  is of the order of  $10^{-8}M$  and therefore totally negligible. However, the smaller the mass ratio, the greater the time the particle spends near the ISCO, and for some sufficiently small mass ratio we might need to place the cutoff even closer to the ISCO. On the other hand, this is likely a moot point because at 1PLT order we require the first-order field for all  $2M < r_p < 6M$  as input to the second-order source, calling for a more complete implementation of the puncture scheme.

Next, SPA waveforms are obtained from equation (6.2). Here no frequency integral is required. The waveform amplitudes (the coefficients of  $e^{-im\phi_G(t_G)}$ ) are tabulated on a grid of  $r_p$  values. In practice,  $t_G$



**Table 3.** Schwarzschild QNM frequencies  $M\omega_{\ell n}$  up to the third overtone for each  $2 \leq \ell \leq 7$ .

	$n=0$	$n=1$	$n=2$	$n=3$
$\ell = 2$	$0.37367 - 0.08896i$	$0.34671 - 0.27391i$	$0.30105 - 0.47828i$	$0.25150 - 0.70515i$
$\ell = 3$	$0.59944 - 0.09270i$	$0.58264 - 0.28130i$	$0.55168 - 0.47909i$	$0.51196 - 0.69034i$
$\ell = 4$	$0.80918 - 0.09416i$	$0.79663 - 0.28433i$	$0.77271 - 0.47991i$	$0.73984 - 0.68392i$
$\ell = 5$	$1.01230 - 0.09487i$	$1.00222 - 0.28582i$	$0.98270 - 0.48033i$	$0.95500 - 0.68056i$
$\ell = 6$	$1.21201 - 0.09527i$	$1.20357 - 0.28665i$	$1.18707 - 0.48056i$	$1.16327 - 0.67859i$
$\ell = 7$	$1.40974 - 0.09551i$	$1.40247 - 0.28716i$	$1.38818 - 0.48071i$	$1.36736 - 0.67735i$

is more simply expressed as a function of  $r_p$  than vice versa, and we plot the waveforms parametrically using  $t_G = t_G(r_p)$  and  $\phi_G = \phi_G(r_p)$ .

Finally, constructing the QNM sum (6.3) requires high-precision QNM frequencies  $\omega_{\ell n}$ , the excitation factors  $B_{\ell}^{e/o}$ , the coefficients  $C_{\ell m}^{e/o}$ , and  $A_{\ell}^{e/o \text{ out}}(\omega_{\ell n})$ . We start with the QNM frequencies from [92, 100, 101] and further increase the digits of accuracy using Leaver's continued fraction method [102]. The results for  $\ell = 2, 3, \dots, 7$  and  $n = 0, 1, 2, 3$  are presented in table 3.

We take the excitation factors from the tabulated values in [94, 100]. The coefficients  $C_{\ell m}^{e/o}$  are computed from equation (5.41) for the fundamental mode and equation (5.53) for the overtones. We have kept the first five terms in equation (5.54) (we have verified that increasing the number of terms to ten leads to a relative difference of  $\sim 10^{-6}$  or smaller in  $C_{\ell m}^{e/o}$ ). We have tabulated all excitation coefficients  $E_{\ell m}^{e/o}(\omega_{\ell n}) = B_{\ell}^{e/o}(\omega_{\ell n}) D_{\ell m}^{e/o}(\omega_{\ell n})$  for  $\ell = 2$  up to the third overtone in tables 4 and 5. The excitation coefficients for higher  $\ell$  are provided in the ancillary material [97]. Finally, we compute the asymptotic amplitudes  $A_{\ell}^{e/o \text{ out}}(\omega_{\ell n})$  using the procedure described in [96].

At this point, we can compare our excitation coefficients with the ones obtained by Hadar and Kol (HK) in tables 1 and 2 of [57], and by Folacci and Ould El Hadj (FO) in table II of [58]. In order to do so we first need to find dictionaries between the different conventions that are being used. Hadar and Kol fix the constant  $t_0$  in equation (3.43) such that  $t_G(2.2M) = 0$ , which leads to the relation  $E_{\ell m}^{e/o}(\omega_{\ell n}) = R_{n\ell m}|_{\text{HK}} e^{i\omega_{\ell n} t_G(2.2M)}$  between the excitation coefficients. We then find agreement with the results of [57] up to an overall minus sign with relative errors of  $\sim 10^{-4}$  or smaller (and a few isolated cases have relative errors of  $10^{-3}$ ). We suspect the sign disagreement to stem from different conventions used to construct the Green function in [57]. It is however of secondary importance for the aim of building full IMR waveforms: the orbital phase during the plunge is known up to a constant, which will be determined by the orbital phase at the end of the transition to plunge. Using the conversion  $\delta'(t - t_G) = -(\dot{r}_{\{0\}}(r_p))^2 \delta'(r - r_p) + (\dots) \delta(r - r_p)$ , we compare the terms proportional to  $\delta'(r - r_p)$  in our sources (5.10) and those in equations (14) and (15) of [58]. We find that the sources differ by an overall factor,  $S_{\ell m}^{e/o} = c_{\text{FO}} S_{\ell m}^{e/o}|_{\text{FO}}$  with  $c_{\text{FO}} = 2\sqrt{2\pi}$ . This leads to  $E_{\ell m}^{e/o}(\omega_{\ell n}) = c_{\text{FO}} \left( -\sqrt{2\pi} C_{\ell m n}^{(e/o)}|_{\text{FO}} \right)$  and  $E_{\ell m}^{e/o}(-\omega_{\ell n}^*) = c_{\text{FO}} \left( -\sqrt{2\pi} D_{\ell m n}^{(e/o)}|_{\text{FO}} \right)$ . Up to an overall factor of  $1/\sqrt{2\pi}$  and an overall minus sign, we

**Table 4.** Excitation coefficients  $E_{\ell m}^{e/o}(\omega_{\ell n})$  for  $\ell = 2$  up to  $n = 3$ .

	$n = 0$	$n = 1$	$n = 2$	$n = 3$
$m = 2$	$-1.14668 \times 10^{-4}$ $+5.72179 \times 10^{-5}i$	$-1.12799 \times 10^{-12}$ $+4.13842 \times 10^{-13}i$	$-8.03493 \times 10^{-22}$ $-1.42463 \times 10^{-21}i$	$2.52610 \times 10^{-31}$ $-1.32057 \times 10^{-31}i$
$m = 1$	$6.53075 \times 10^{-5}$ $-1.87158 \times 10^{-5}i$	$1.33791 \times 10^{-12}$ $+1.37608 \times 10^{-13}i$	$-9.43025 \times 10^{-22}$ $+2.48916 \times 10^{-21}i$	$-3.96283 \times 10^{-31}$ $-3.66314 \times 10^{-31}i$
$m = 0$	$2.06438 \times 10^{-5}$ $+8.84288 \times 10^{-6}i$	$4.91061 \times 10^{-13}$ $+5.75437 \times 10^{-13}i$	$-2.12623 \times 10^{-21}$ $+4.28862 \times 10^{-22}i$	$2.36628 \times 10^{-31}$ $-4.82273 \times 10^{-31}i$
$m = -1$	$-8.83249 \times 10^{-8}$ $-4.89523 \times 10^{-6}i$	$8.87604 \times 10^{-14}$ $-2.20023 \times 10^{-13}i$	$5.31923 \times 10^{-22}$ $+7.96482 \times 10^{-22}i$	$-2.92157 \times 10^{-31}$ $-7.312004 \times 10^{-32}i$
$m = -2$	$6.18182 \times 10^{-7}$ $-8.08701 \times 10^{-8}i$	$3.95622 \times 10^{-14}$ $+5.77111 \times 10^{-15}i$	$-1.49044 \times 10^{-22}$ $+1.63288 \times 10^{-22}i$	$1.01126 \times 10^{-32}$ $-8.95405 \times 10^{-32}i$

**Table 5.** Excitation coefficients  $E_{\ell m}^{e/o}(-\omega_{\ell n}^*)$  for  $\ell = 2$  up to  $n = 3$ .

	$n = 0$	$n = 1$	$n = 2$	$n = 3$
$m = 2$	$6.18182 \times 10^{-7}$ $+8.08701 \times 10^{-8}i$	$3.95622 \times 10^{-14}$ $-5.77111 \times 10^{-15}i$	$-1.49044 \times 10^{-22}$ $-1.63288 \times 10^{-22}i$	$1.01126 \times 10^{-32}$ $+8.95405 \times 10^{-32}i$
$m = 1$	$8.83249 \times 10^{-8}$ $-4.89523 \times 10^{-6}i$	$-8.87604 \times 10^{-14}$ $-2.20023 \times 10^{-13}i$	$-5.31923 \times 10^{-22}$ $+7.96482 \times 10^{-22}i$	$2.92157 \times 10^{-31}$ $-7.31200 \times 10^{-32}i$
$m = 0$	$2.06438 \times 10^{-5}$ $-8.84288 \times 10^{-6}i$	$4.91061 \times 10^{-13}$ $-5.75437 \times 10^{-13}i$	$-2.12623 \times 10^{-21}$ $-4.28862 \times 10^{-22}i$	$2.36628 \times 10^{-31}$ $+4.82273 \times 10^{-31}i$
$m = -1$	$-6.53075 \times 10^{-5}$ $-1.87158 \times 10^{-5}i$	$-1.33791 \times 10^{-12}$ $+1.37608 \times 10^{-13}i$	$9.43025 \times 10^{-22}$ $+2.48916 \times 10^{-21}i$	$3.96283 \times 10^{-31}$ $-3.66314 \times 10^{-31}i$
$m = -2$	$-1.14668 \times 10^{-4}$ $-5.72179 \times 10^{-5}i$	$-1.12799 \times 10^{-12}$ $-4.13842 \times 10^{-13}i$	$-8.03493 \times 10^{-22}$ $+1.42463 \times 10^{-21}i$	$2.52610 \times 10^{-31}$ $+1.32057 \times 10^{-31}i$

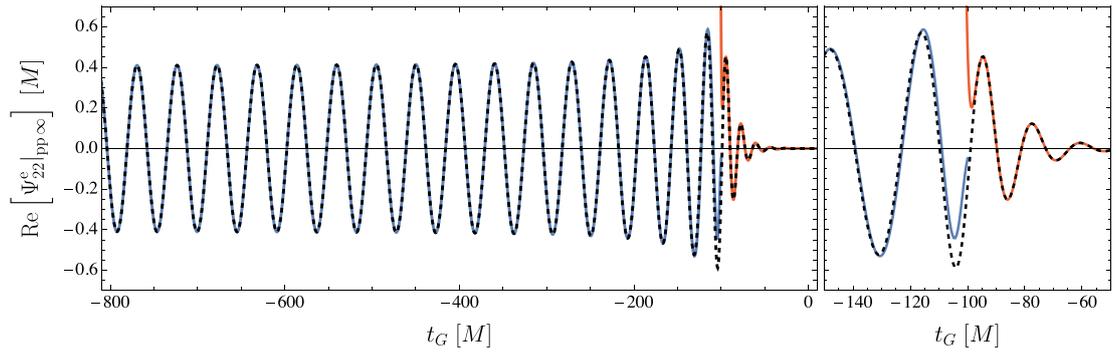
find the first equality to be satisfied with relative errors of  $\sim 10^{-4}$  or smaller (except for the  $\ell = m = 2$  excitation coefficient, where we suspect a typo in the sign of the imaginary part in [58]). We also find the equality for the mirror modes to have larger relative errors especially for  $\ell \geq 4$ , which originate from disagreements of the same magnitude already at the level of the QNM frequencies. We have investigated the source of the disagreement by an overall factor of  $1/\sqrt{2\pi}$ : in [58] the inverse Fourier transform is defined in equations (21) and (22) with the symmetric normalization  $1/\sqrt{2\pi}$ . Equations (15) and (25) in [58] are however consistent with a forward transform of the form  $S_{\omega \ell m}^{(e/o)}(r) = \int_{-\infty}^{+\infty} dt S_{\ell m}^{(e/o)}(t, r) e^{i\omega t}$ , missing a factor  $1/\sqrt{2\pi}$ . Again, we have not found the source of the minus sign disagreement.

### 6.2. Comparison with the SPA and the QNM sum

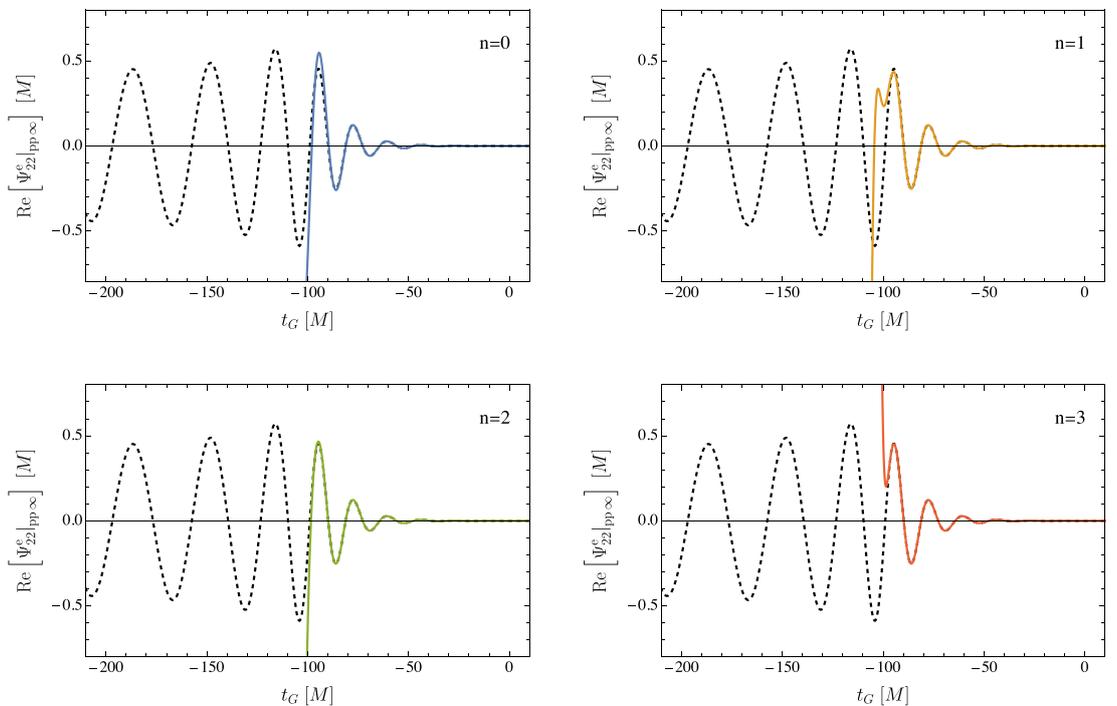
In this subsection we perform internal consistency checks and compare the OPG waveforms computed following the procedure described in sections 5.4 and 6.1 with the SPA and the QNM sum discussed in sections 5.5 and 5.6, respectively.

In figure 8 we compare the SPA (5.46) (blue curve) with the OPG waveform (dashed black curve) for the  $(\ell, m) = (2, 2)$  mode. The waveforms agree very well at early times, with relative errors below 1% up to four cycles before the peak amplitude. Around merger the two waveforms start to differ significantly, with the SPA being unable to capture the late-time QNM behaviour. Indeed, by construction, the SPA is expected to work as long as the integrand in equation (5.42) contains a rapidly oscillating factor, which occurs as the particle is close to the ISCO and corresponds to the early portion of the waveform. This is also consistent with the geodesic plunge trajectory displayed in figure 3, where the particle spends a large number of orbits close to the ISCO before falling from  $r \approx 5.7M$  to the event horizon within  $\sim 2$  orbital cycles.

As shown by the comparison with the red curve in figure 8, at late times the OPG waveform is well described by the QNM sum (5.49). Figure 9 shows the comparison between the OPG waveform for the  $(\ell, m) = (2, 2)$  mode and QNM sums that include a different number of overtones (up to the first three),



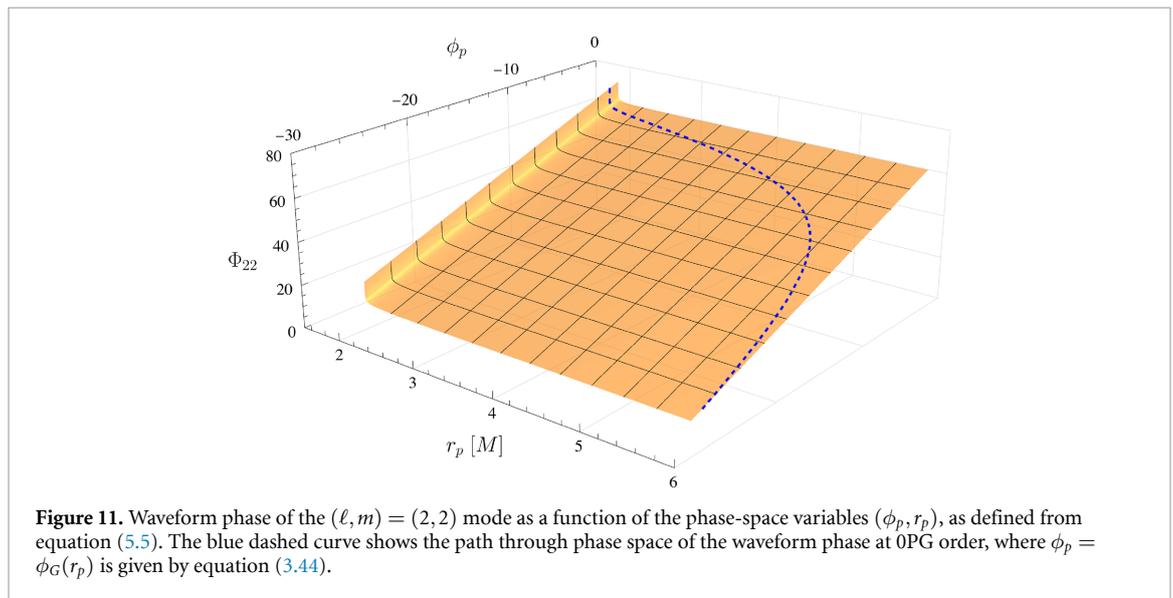
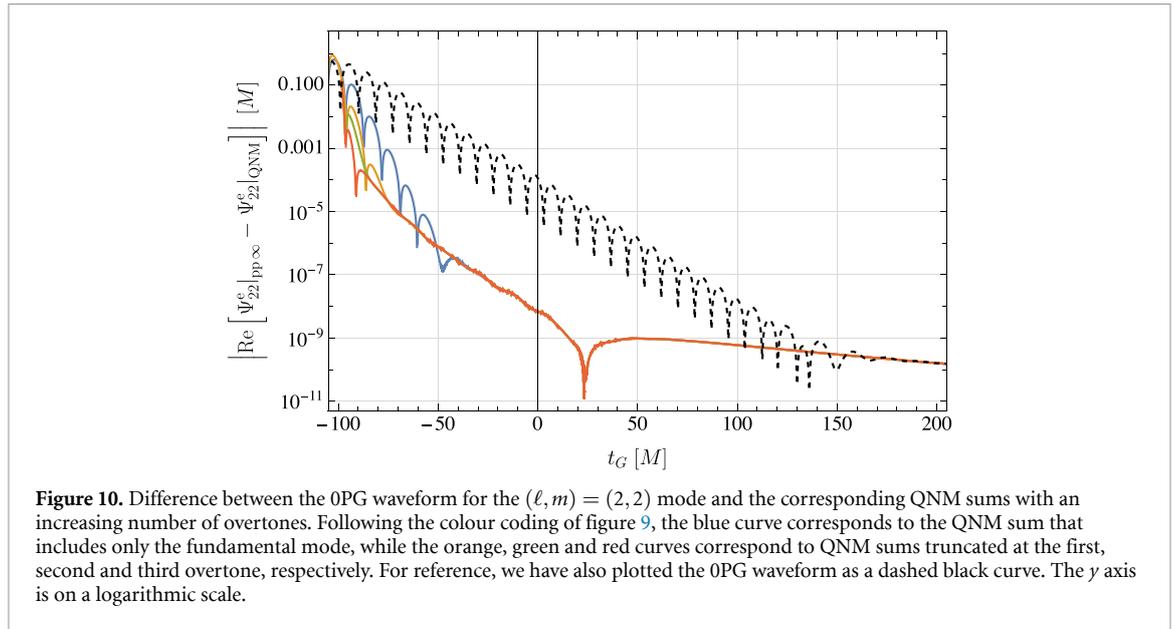
**Figure 8.** The  $(\ell, m) = (2, 2)$  mode of the OPG waveform (black dashed curve) compared to the corresponding SPA (5.46) (blue curve) and QNM (5.49) (red curve) waveforms. The QNM sum includes overtones through  $n = 3$ . We have truncated the SPA and QNM waveforms once they start to significantly differ from the OPG waveform. The plot on the right zooms in on the region around the peak amplitude.



**Figure 9.** OPG waveforms for the  $(\ell, m) = (2, 2)$  mode (dashed black curves) compared with the corresponding QNM sums that include the fundamental mode only ( $n = 0$ ) and an increasing number of overtones up to  $n = 3$ . Note that we are only plotting the QNM waveforms up to a time where they start to significantly differ from the OPG waveforms.

while in figure 10 we plot the difference between the waveform and the QNM sums. These plots confirm that the QNM approximation correctly reproduces the late-time behaviour of the waveform, with increasing accuracy as higher overtones are included. Although including a larger number of overtones improves the accuracy of the QNM sum at late times, doing so does not extend its accuracy to earlier times. Crucially, the regime of validity breaks down well before reaching the waveform's peak amplitude. We can see from the right panel of figure 8 that this breakdown occurs around  $\approx 10M$  after the peak, which is consistent with studies of numerical merger-ringdown waveforms [59–62]. Those studies found that numerically fitting QNM sums to NR data becomes stable and robust around  $10M$ – $20M$  after the waveform peak.

We hence conclude that our first-order waveforms are internally consistent and agree with expectations in the sense that the SPA correctly reproduces the periodic behaviour at early times, while the QNM sum describes the ringdown at late times. The combination of the two approximations accurately covers almost the full OPG waveform, only missing a few cycles around the peak amplitude.



With these results in mind, we return to the phase-space description of the waveform. The SPA makes explicit that the waveform is a simple function of phase-space coordinates until near the waveform’s peak amplitude, while the QNM sum, which involves frequencies unrelated to the orbit, suggests that the waveform becomes divorced from the phase-space trajectory after the peak amplitude. However, this intuition is somewhat misleading. Before generating the waveform as a function of time, we can directly examine its original form (5.40) as a function of the phase-space coordinates  $(\phi_p, r_p)$ . In figure 11 we plot the waveform phase as a function of  $\phi_p$  and  $r_p$ . Here we see there is no drama at the waveform peak, corresponding to  $r_p \approx 3M$ . The emitted waveform’s phase is a simple, smooth function of the particle’s orbital parameters until the particle gets extremely close to the event horizon; there, at  $r_p \approx 2.01M$ , the waveform phase rapidly dissociates from the state of the particle’s orbit.

In the figure we also plot the particle’s phase-space trajectory. From that trajectory, we see that at OPG order, the waveform phase gradually deviates from  $m\phi_p$  over the course of the plunge. When the trajectory approaches the horizon, it effectively freezes: approximately 0.7 radians after the peak, large changes in the waveform phase become associated with virtually no change in  $\phi_p$  or  $r_p$ . It is this marked change in behaviour, when the particle is far below the light ring, that coincides with the time ( $\approx 10$ – $15M$  after the peak) when the QNM sum begins to approximate the complete waveform.

At higher PG orders, the surface in figure 11 will be deformed due to terms  $h_{22}^{\{2\}}$  and higher in the waveform. The trajectory on the surface will also change due to 1PG and higher corrections to the orbital motion.

### 6.3. Comparison with NR

This subsection contains qualitative comparisons between our OPG waveforms and NR simulations from the SXS catalog [103, 104]. All SXS data we work with here are drawn from the public python package `sxs` [105].

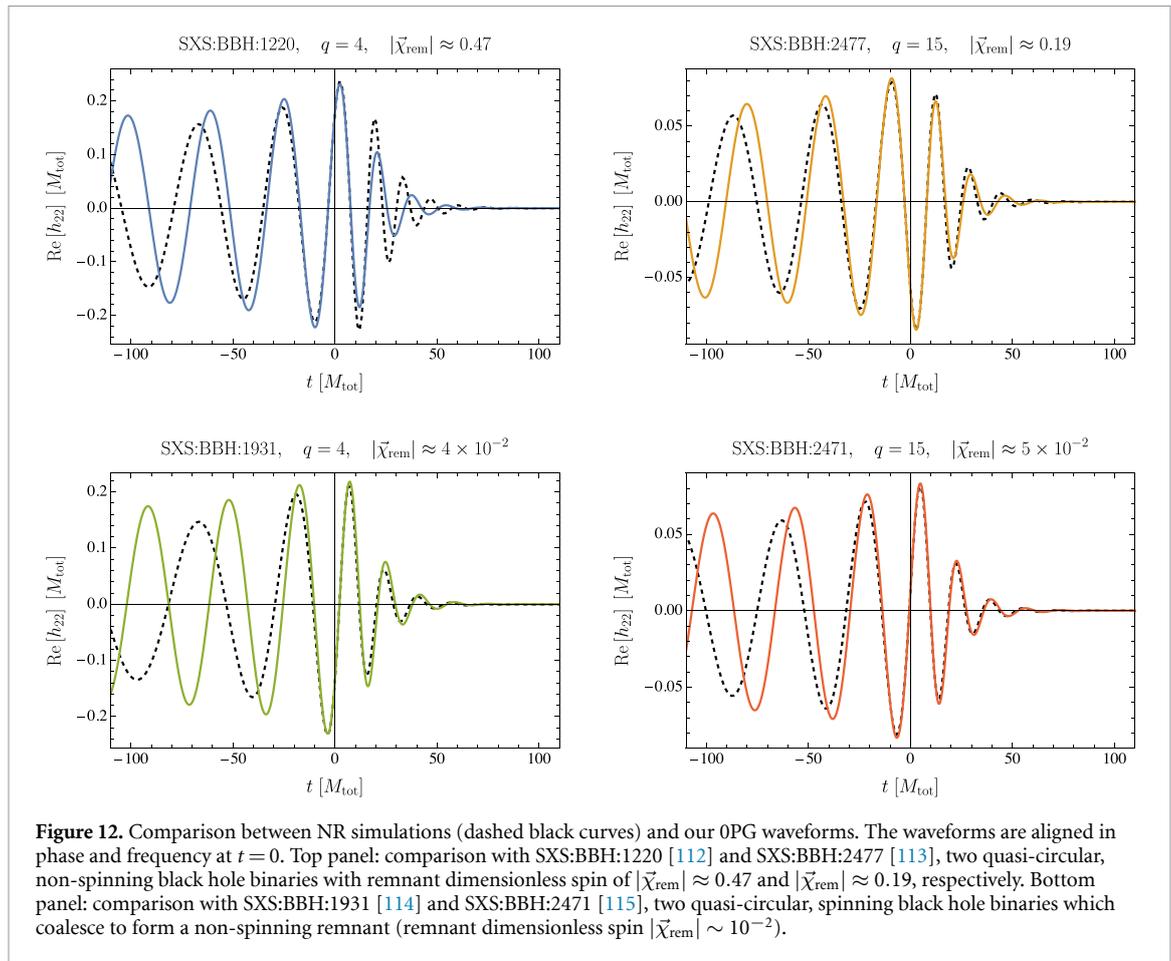
We start by re-expanding the relevant quantities for our waveform generation in powers of the symmetric mass ratio  $\nu := Mm_p/(M+m_p)^2 = \varepsilon/(1+\varepsilon)^2$  at fixed total mass  $M_{\text{tot}} := M+m_p = M(1+\varepsilon)$ . This yields the most accurate comparisons in the regime of comparable-mass binaries [21, 51, 55, 106–111]. Inverting the relation between  $\nu$  and  $\varepsilon$  gives  $\varepsilon = (1-2\nu - \sqrt{1-4\nu})/(2\nu)$ , which, in the small-mass-ratio expansion, leads to  $\varepsilon = \nu + 2\nu^2 + \mathcal{O}(\nu^3)$ . For our first-order waveforms this re-expansion simply amounts to substituting  $\varepsilon \rightarrow \nu$  and  $M \rightarrow M_{\text{tot}}$ . In order to compare the OPG and NR waveforms, we align them in time and phase at a given waveform frequency. Referring to equation (5.5), we define the  $\ell m$ -mode waveform phase and frequency from  $h_{\ell m} = |H_{\ell m}|e^{-i\Phi_{\ell m}}$  as  $\Phi_{\ell m} := m\phi_p - \arg(H_{\ell m})$  and  $\omega_{\ell m} := d\Phi_{\ell m}/dt$ .

After the above preliminaries, we can begin to assess the accuracy of our OPG waveforms, using SXS waveforms as our benchmark. The top panel of figure 12 compares the  $(\ell, m) = (2, 2)$  mode of our OPG waveforms with the NR simulations SXS:BBH:1220 [112] and SXS:BBH:2477 [113]. These are NR waveforms for quasi-circular, non-spinning black hole binaries with mass ratios  $q=4$  and  $q=15$ , respectively. For both simulations, the reference dimensionless spin on each of the two black holes is  $\sim 10^{-5}$  or smaller, and the eccentricity is  $\lesssim 10^{-4}$ . We align the waveforms in time and phase at  $\omega_{22} = 2\Omega_{\{0\}}(3M_{\text{tot}})$ , which corresponds to the peak amplitude  $|H_{22}|$  of the plunge waveform. We also shift the relevant part of the waveforms in time such that  $t=0$  corresponds to the alignment time. In the  $q=4$  case the OPG waveform is not able to accurately capture the late-time behaviour of the NR simulation, while the comparison improves drastically for  $q=15$ . The improvement with increasing  $q$  is expected: our OPG model omits 1PG and higher terms in the phase and amplitude, leading to relative errors that scale roughly linearly with  $\varepsilon$ .

Numerous 1PG and higher effects contribute to our model's error, but the error after the waveform's peak appears to be dominated by the fact that our OPG model does not account for the final, remnant black hole's spin. In the top panel of figure 12, the NR simulations we compare against have small spins on the individual black holes during the inspiral, but the remnant black holes have substantial dimensionless spins  $|\vec{\chi}_{\text{rem}}| \approx 0.47$  for SXS:BBH:1220 and  $|\vec{\chi}_{\text{rem}}| \approx 0.19$  for SXS:BBH:2477. The spin of the remnant is largely determined by the dimensionless angular momentum of the plunging particle, which at OPG order is given by its value at the ISCO,  $2\sqrt{3}\nu$ , equal to  $\approx 0.55$  for  $q=4$  and  $\approx 0.20$  for  $q=15$ , in rough agreement with the SXS values. This significant final spin leads to QNM ringdown frequencies that differ significantly from the Schwarzschild ones captured by our OPG model. In a 1PG model, these shifts in the  $(\ell, m) = (2, 2)$  ringdown frequencies should be accounted for primarily by the nonlinear source  $\delta^2 G_{\alpha\beta}^{\{0\}}[h^{\{1\}}, h^{\{1\}}]$ , through couplings between the  $m=2$  modes of  $h_{\alpha\beta}^{\{1\}}$  and the  $(\ell, m) = (1, 0)$  mode of  $h_{\alpha\beta}^{\{1\}}$ ; this is because the  $(1, 0)$  mode encodes the angular momentum content in  $h_{\alpha\beta}^{\{1\}}$ , and it necessarily asymptotes, at late times, to a perturbation toward the final, stationary Kerr metric of the remnant.

To investigate whether the remnant spin is really our main source of error in the post-peak regime, we now consider two alternative NR waveforms that have the same mass ratios as the ones used above but negligible remnant spin: the quasi-circular binary black hole simulations SXS:BBH:1931 [114] with  $q=4$  and SXS:BBH:2471 with  $q=15$  [115]. In these simulations the individual black holes are spinning prior to merger, but the remnant dimensionless spins are very small ( $|\vec{\chi}_{\text{rem}}| \sim 10^{-2}$ ). We compare these NR simulations to our OPG waveforms in the bottom panel of figure 12. At late times the agreement has improved since the Schwarzschild QNM sum is now a better approximation to the remnants' ringdown. Before the peak amplitude the comparison instead becomes worse since before merger the NR waveform describes the inspiral of two spinning black holes, which is not the scenario described by our waveforms. In any case, we do not expect the plunge waveform to be accurate far before the peak amplitude and expect the transition to plunge to take over before reaching the ISCO.

We can illustrate our accuracy in more detail by separately comparing the OPG waveform phases  $\Phi_{\ell m}$  (figures 13 and 15) and waveform amplitudes  $|h_{\ell m}|$  (figures 14 and 15) with NR. For a given mass ratio, we see that for  $t \gtrsim 0$  the 'zero-remnant-spin' waveforms (SXS:BBH:1931 and SXS:BBH:2471) compare better than 'zero-initial-spin' waveforms (SXS:BBH:1220 and SXS:BBH:2477), while the roles are reversed



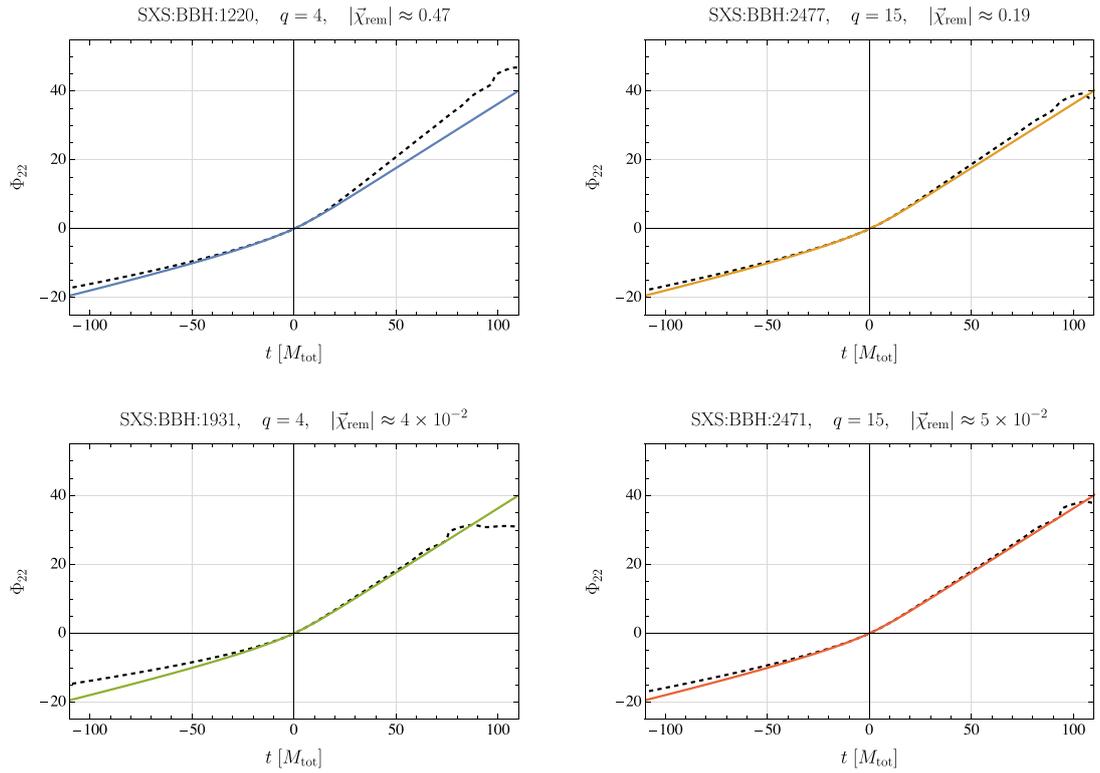
**Figure 12.** Comparison between NR simulations (dashed black curves) and our OPG waveforms. The waveforms are aligned in phase and frequency at  $t = 0$ . Top panel: comparison with SXS:BBH:1220 [112] and SXS:BBH:2477 [113], two quasi-circular, non-spinning black hole binaries with remnant dimensionless spin of  $|\bar{\chi}_{\text{rem}}| \approx 0.47$  and  $|\bar{\chi}_{\text{rem}}| \approx 0.19$ , respectively. Bottom panel: comparison with SXS:BBH:1931 [114] and SXS:BBH:2471 [115], two quasi-circular, spinning black hole binaries which coalesce to form a non-spinning remnant (remnant dimensionless spin  $|\bar{\chi}_{\text{rem}}| \sim 10^{-2}$ ).

for  $t \lesssim 0$ ; graphically, in figure 15, the green and red curves respectively lie below the blue and orange ones for  $t \gtrsim 0$  and vice versa for  $t \lesssim 0$ . It is also the case that the comparison improves with increasing  $q$ : in figure 15, the orange and red curves respectively lie below the blue and green ones for all  $t$ , modulo the parts of the plots dominated by the numerical noise. This also holds when considering the relative, rather than absolute, difference. In all cases except the one with large remnant spin, we observe that our OPG model matches the peak amplitude of the SXS waveform to a high degree of accuracy.

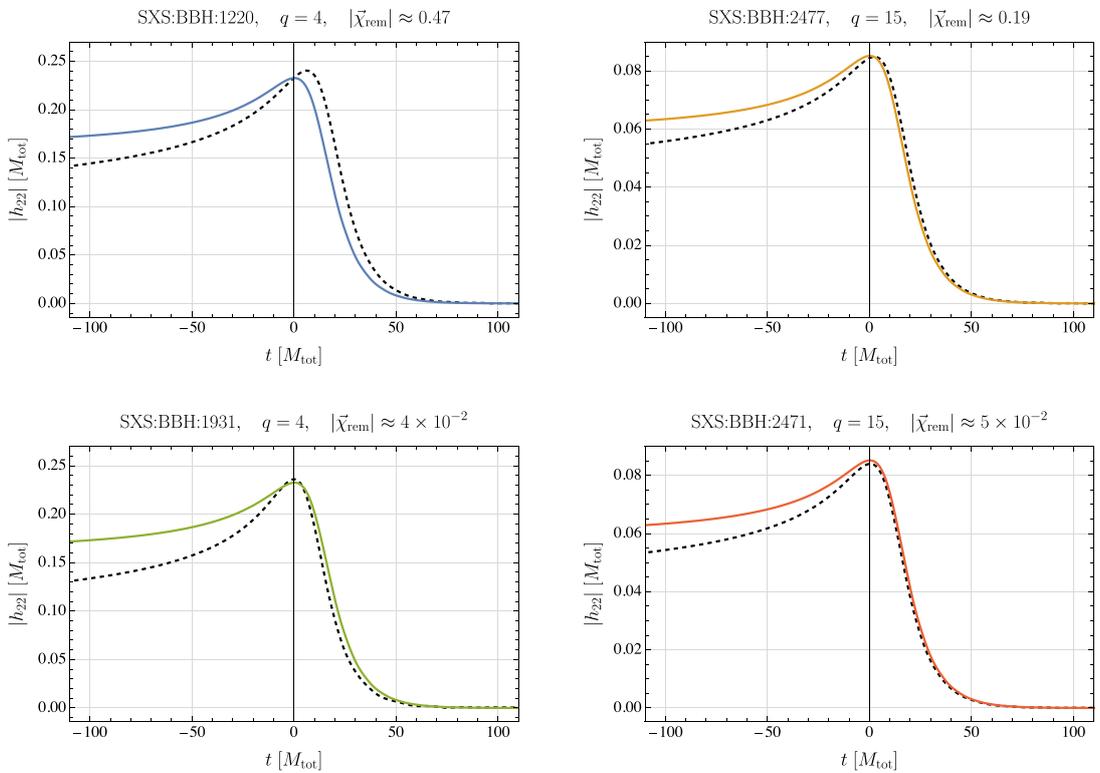
## 7. Discussion and conclusions

In this paper we have presented a framework for generating merger-ringdown waveforms within GSF theory. We have treated the final plunge, which gives rise to the merger-ringdown portion of the GW signal, using a post-geodesic expansion in a phase-space approach, mirroring our previous treatments of the inspiral and the transition to plunge [43, 51]. The resulting waveform is summarized in equations (2.14) and (2.15), with  $x^i = (r, \theta, \phi)$  and  $r \rightarrow \infty$  in equation (2.15). All the ingredients needed for waveform generation—the waveform amplitudes  $h_{\alpha\beta}^{\{n\},m}(J^a, r \rightarrow \infty)$  and the forcing functions  $F^a_{\{n\}}(J^b)$ —are pre-computed as functions of the binary’s mechanical variables  $J^a = (\Omega, \delta M^\pm)$ . Evolving through the binary’s mechanical phase space then yields the time-domain waveforms. This split between slow offline computations (solving field equations) and fast online ones (solving a very small number of ordinary differential equations) allows us to maintain the rapid waveform generation of inspiral and transition-to-plunge self-force models. We have outlined this framework through second order in  $\varepsilon$  for the waveform, or equivalently, 1PG order in the orbital dynamics. Similarly, while we have worked with  $t$  slicing in our calculations here, we have also emphasized that hyperboloidal slicing (or null slicing as in figure 1) would offer advantages that become increasingly important at post-geodesic orders.

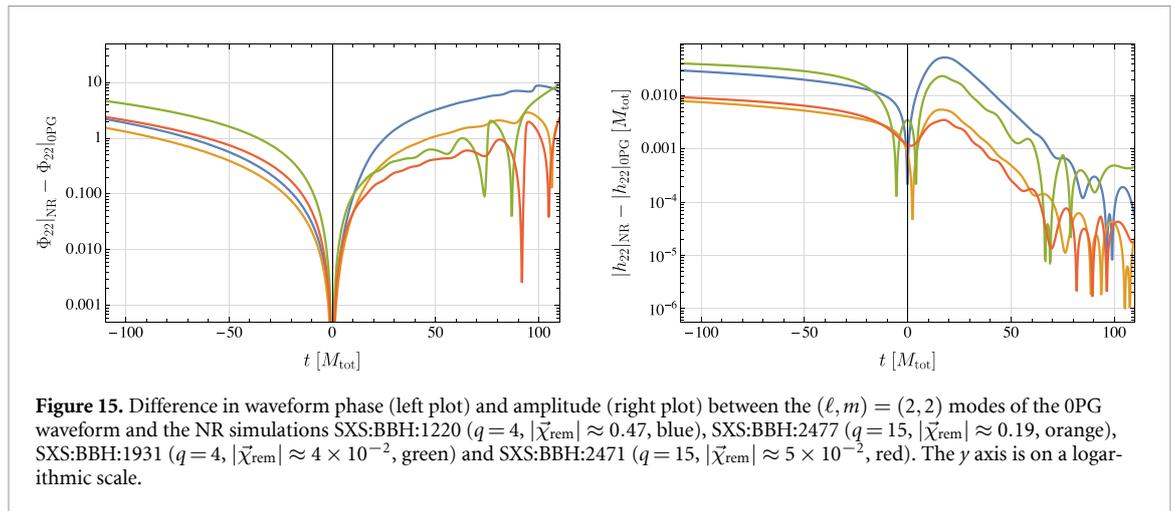
At early times, close to the ISCO, our plunge expansion asymptotically matches the late-time transition-to-plunge expansion of [51]. We have detailed this asymptotic matching up to 1PG and 7PLT order and shown how it provides critical input for the plunge solution, both in the orbital dynamics and in the field equations. Without the asymptotic matching, the plunge expansion is under-determined and involves singularities at the ISCO. In the field equations, the matching conditions can be rigorously



**Figure 13.** Comparison of the  $(\ell, m) = (2, 2)$  waveform phase  $\Phi_{\ell m}$  as a function of time between the OPG waveforms (coloured curves) and the NR simulations (dashed black curves) considered in figure 12. The colour coding follows the one of figure 12. The phase difference vanishes at the alignment time  $t = 0$  by construction.



**Figure 14.** Comparison of the  $(\ell, m) = (2, 2)$  waveform amplitude  $|h_{\ell m}|$  as a function of time between the OPG waveforms (coloured curves) and the NR simulations (dashed black curves) considered in figure 12. The colour coding follows the one of figure 12.



enforced through a puncture scheme, with an effective source appearing in a punctured region at early times, near the ISCO. For our implementation in this paper, we have safely neglected the effective source in the punctured region by focusing on times sufficiently far to the future of the puncture. However, calculations at 1PG order will likely require a more thorough treatment of the effective source, as the 0PG field is needed all the way back to the ISCO as input for the 1PG field equations.

Our framework involves field equations defined directly on the binary phase space, as opposed to the ordinary time domain, but we have shown how to transform these equations into a familiar frequency-domain form. Using frequency-domain methods, we have implemented our formalism at leading (0PG) order, solving RWZ equations to obtain the leading-order plunge waveform. Although this waveform has been obtained previously from the same frequency-domain equations [58], we have shown that there is still new information to be mined from it. Specifically, we have utilized it to illuminate fundamental features of merger and ringdown, and we have begun to assess its accuracy through comparisons with NR waveforms.

First, to probe the structure of the merger and ringdown, we have focused on two questions: how well is the waveform described by an ‘extended inspiral’ prior to the waveform’s peak, as in EOB [33], and how well is it described by a QNM ringdown after the peak? Our key findings are illustrated in figures 8 and 11.

In figure 8 we see that the early part of the waveform, until very close to the peak, is well approximated by a SPA that expresses the waveform phase and amplitude as instantaneous functions of the orbital phase and radius, akin to EOB’s extended-inspiral construction. In the future we will explore how much this can be improved by including higher-order terms in the SPA and whether our approach can be used to inform EOB’s ‘NQC’s’ [35].

Figure 8 also shows that at late times,  $\gtrsim 10M$  after the peak amplitude, the QNM ringdown becomes a good approximation to the full waveform. However, at earlier times the QNM sum diverges dramatically. This has bearing on a long line of studies. Many papers have investigated whether a QNM sum, including a variable number of overtones, is able to reproduce the GW strain even at the peak [59, 116–128], following earlier evidence that a QNM model with up to  $n = 3$  overtones can fit the ringdown even at times *preceding* the peak [129]. Such investigations are mainly based on fitting NR data using QNM models built from a sum of damped sinusoids (with QNM frequencies). We have instead computed the QNM sum directly from first principles (analogous to [130]’s study for a toy source term). Our results show that, at least in the case of a Schwarzschild primary and including the first three overtones, the QNM sum only becomes accurate significantly after the waveform’s peak amplitude. Our finding of *when* it becomes accurate is broadly consistent with the time at which QNM fits to NR ringdown waveforms become robust [59–62]. Figure 9 shows that this ringdown-onset time changes only marginally as we vary the number of overtones up to  $n = 3$ , but we note that [130]’s results suggest that the effective onset might appear at substantially earlier times if many more overtones were included.

A natural followup, in addition to including higher overtones, would be to further dissect the ringdown. We compute our full waveform through an inverse Fourier transform, involving an integral over all real frequencies. In section 5.6 we have reviewed how this real-line integral can be equated to the sum of three contributions with complex frequencies: the QNM sum, which we have computed including the first 3 overtones; an integral around a branch cut, associated with power-law tails (which are clearly

visible after  $t_G \approx 150M$  in our full waveform in figure 10); and an integral along a high-frequency arc, associated with a prompt response. We will explore the branch-cut and high-frequency arc contributions in future work, complementing recent studies such as [131].

At least at 0PG order, in figure 8 we can identify the genuine ‘merger’ regime as the brief interval when neither the SPA nor the QNM sum is accurate. However, figure 11 shows that the waveform remains closely tied to the orbital motion in this interval, even though it occurs after the particle has fallen behind the light ring. It therefore appears plausible that a higher-order SPA can be pushed to later times, past the waveform’s peak. On the other hand, the onset of the QNM regime is very sudden in figure 11, marking a clear dissociation of the waveform from the orbit and indicating the difficulty of pushing the QNM sum to earlier times.

We expect these types of analyses will further sharpen our understanding of the merger process, particularly when our model is extended to 1PG order, where nonlinear effects arise. Such analyses might additionally have utility in improving EOB and Phenom IMR waveform models.

Separate from the broad, fundamental questions about the merger regime, we have also qualitatively assessed the accuracy of the 0PG waveform, using NR waveforms from the SXS catalog as a benchmark. For sufficiently small mass ratios, our 0PG waveform’s error is necessarily dominated by its omission of 1PG terms, corresponding to  $\mathcal{O}(\nu^2)$  terms in the waveform amplitude and  $\mathcal{O}(\nu)$  terms in the phase. Our comparisons with NR are consistent with that expectation, and they suggest that a complete 1PG model could be highly accurate at mass ratios  $\varepsilon \approx 1/10$  (and perhaps even closer to  $\varepsilon = 1$ ). We have also singled out the most significant 1PG corrections: those associated with the nonzero spin of the final, remnant black hole. Such corrections will appear automatically in our framework at 1PG order, and they are potentially easier to calculate than the full litany of 1PG effects. Hence, as an intermediate step toward 1PG waveforms, there would be value in building an incomplete 1PG model, including only the 1PG terms associated with the remnant spin.

Our comparisons with NR have sufficed to establish that our error converges correctly with the mass ratio (and that the order- $\nu^2$  error term does not have a large coefficient), but this benchmarking is still somewhat limited by the sparseness of NR simulations for mass ratios more asymmetric than 1:10. In the future, advances in NR methods [132–134] might enable more detailed comparisons across a large range of mass ratios. However, there is also strong motivation for NR simulations that specifically target the final stages of the binary evolution using existing codes, in order to more quantitatively assess the accuracy of self-force-based treatments of the transition and plunge. Another point of comparison will be EOB treatments of the merger, extending the inspiral comparisons from [22, 135]. Finally, we plan to compare our method against the IMR calculations that are now possible within self-force theory [136] using the self-consistent approach [74], which smoothly evolves across the ISCO and through the plunge without performing additional expansions in each of these regimes.

Our next immediate goal, however, is to build an IMR model that enables the production of fast and accurate waveforms for asymmetric binaries. We aim to achieve this by combining the framework presented in this paper, together with our earlier work on the transition to plunge [51], with the post-adiabatic inspiral [21, 43]. In work presented in [137–139], we have built a fast IMR waveform model using the 1PA inspiral, the transition to plunge through 2PLT order and the geodesic plunge, with switching points between the three regimes. That model will be detailed in a followup paper. Future work will focus on improving the connection between the three regimes and carrying the transition-to-plunge and plunge expansions to higher order, with the aim of extending the sub-radian accuracy of post-adiabatic inspiral waveforms through merger and ringdown.

Further future work includes the generalization to plunges (both prograde and retrograde) into a Kerr black hole. The general framework described in this paper also applies when considering a spinning primary. For a particle plunging from the Kerr ISCO [140, 141], obtaining the first-order plunge waveform would then be achieved using the Teukolsky equation [142] rather than the RWZ equations employed in this paper. The generalization of the transition-to-plunge regime to the case of a spinning primary black hole is described in [143], and the extension to the plunge is underway.

The generalization to generic equatorial plunging orbits, which are reached when the inspiral admits eccentricity is also an open problem. On the one hand, the late-time behaviour of the adiabatic inspiral was obtained in the approach to the separatrix [144]. On the other hand, the role of critical plunging geodesics that exactly admit the separatrix values of energy and angular momentum was identified [145], and the corresponding time-domain Teukolsky waveforms were obtained and analysed [145]. Teukolsky waveforms on plunging geodesics guided by an EOB Hamiltonian were also recently derived [145–148]. Yet, a GSF framework that allows one to consistently assemble the inspiral and plunging regime has yet to be formulated to complete already defined models [149]. The formulation of such a framework is in progress.

The offline/online split in our approach also lends itself to modelling mergers beyond general relativity or including matter effects. Waveform ingredients (beyond-GR or matter corrections to amplitudes and forcing functions) can be computed offline, in a modular way, and then added to the online waveform generation, just as has been done in the inspiral (see, e.g. [150–152]). We will exploit this fact to explore black hole mergers in alternative theories of gravity [153] or in the presence of dark matter halos [154].

Finally, we note that, very recently, during the preparation of our paper, [81] also presented a fast, leading-order IMR waveform model that includes the same OPG waveform we have detailed here. Their IMR model involves stitching the time-domain OPA inspiral waveform to the time-domain OPG plunge waveform, skipping the transition regime in between. Since the transition regime has a frequency width of order  $\varepsilon^{2/5}$ , omitting it should have a small effect for sufficiently small mass ratios; however, we expect that this omission will incur significant errors for moderate mass ratios. Such stitching might also become impractical beyond leading order, as the post-adiabatic and post-geodesic expansions become increasingly singular at the ISCO. Since the OPA and OPG waveforms never have a common frequency, stitching them together in this way would additionally seem to require a discontinuity in the waveform frequency. However, the approach has a clear advantage in its simplicity, and it should suffice for many systems, such as EMRIs, where the merger and ringdown carry little SNR.

## Data availability statement

For the ease of reproducibility of our results, we provide several Mathematica notebooks in ancillary files [97].

The data that support the findings of this study are openly available at the following URL/DOI: <https://github.com/gcompere/Self-force-framework-for-merger-ringdown-waveforms>.

## Acknowledgment

We would like to thank Ben Leather for helpful discussions and collaboration on a closely related project, as well as for providing a Mathematica template for drawing Penrose diagrams. We are also grateful to Ayush Roy and Rodrigo Panosso Macedo for insights gleaned from another related project. A P thanks Jeremy Miller for valuable discussions and collaboration in the preliminary stages of this work, as well as Gregorio Carullo, Marc Casals, and Laura Sberna for conversations about QNM sums. L K and G C thank Mohamed Ould El Hadj for several interesting conversations about the implementation of [58]. Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region. A P and L K acknowledge the support of a UKRI Frontier Research Grant (as selected by the ERC) under the Horizon Europe Guarantee scheme (Grant Number EP/Y008251/1). A P additionally acknowledges the support of a Royal Society University Research Fellowship. G C is Research Director of the FNRS. He acknowledges the support of the Win4Project grant ETLOG of the Walloon Region for the Einstein Telescope. This work makes use of the Black Hole Perturbation Toolkit [98].

## Appendix A. Asymptotic match between transition to plunge and plunge

### A.1. Coefficients of the asymptotic late-time transition-to-plunge solutions

Up to 7PLT order, the first coefficients in the late-time solution of the orbital radius (3.27a) are given by

$$r_{[0]}^{(2,1)} = -24\sqrt{6}M^2, \quad r_{[0]}^{(2,1-5i/2)} = 0, \quad \forall i \geq 1, \quad (\text{A.1})$$

$$r_{[1]}^{(3,3/2-5i/2)} = 0, \quad \forall i \geq 0, \quad (\text{A.2})$$

$$r_{[2]}^{(4,2)} = 576M^3, \quad r_{[2]}^{(4,-1/2)} = -362^{1/4}3^{3/4}M^{3/2}f_{[5]}^t, \quad (\text{A.3})$$

$$r_{[3]}^{(5,5/2)} = 0, \quad r_{[3]}^{(5,0)} = -54M^2f_{[5]}^r, \quad r_{[3]}^{(5,-5i/2)} = 0, \quad \forall i \geq 1, \quad (\text{A.4})$$

$$r_{[4]}^{(6,3)} = -3072\sqrt{6}M^4, \quad r_{[4]}^{(6,1/2)} = -366^{1/4}M^{3/2} \left( 48\sqrt{2}Mf_{[5]}^t - \sqrt{3}f_{[7]A}^t \right), \quad (\text{A.5})$$

$$r_{[5]}^{(7,7/2)} = 0, \quad r_{[5]}^{(7,1)} = 6M \left( f_{[8]A}^t + 90\sqrt{6}M^2f_{[5]}^r \right), \quad (\text{A.6})$$

$$r_{[6]}^{(8,4)} = 92160M^5, \quad (A.7)$$

$$r_{[6]}^{(8,3/2)} = \left(\frac{2}{3}\right)^{3/4} M^{1/2} f_{[9]B}^t + 12 \cdot 6^{1/4} M^{3/2} \left[ \sqrt{3} f_{[9]A}^t + 24M \left( 65\sqrt{3} M f_{[5]}^t - 3\sqrt{2} f_{[7]A}^t \right) \right],$$

$$r_{[7]}^{(9,9/2)} = 0 \quad (A.8)$$

$$r_{[7]}^{(9,2)} = \frac{1}{72} \left[ \sqrt{6} f_{[10]B}^t + 3\sqrt{6} f_{[10]C}^t + 216 M f_{[10]D}^t - 432 M^2 \left( 11\sqrt{6} f_{[8]A}^t - 3\sqrt{6} f_{[10]E}^t - 36\sqrt{6} M f_{[7]A}^t + 4752 M^2 f_{[5]}^t \right) \right].$$

The first coefficients in the late-time solution of  $d\Omega/dt$  (3.27a) up to 7PLT order read

$$F_{[0]}^{(3,3/2)} = \frac{1}{2^{1/4} 3^{7/4} M^{1/2}}, \quad F_{[0]}^{(3,-1)} = \frac{f_{[5]}^t}{4M^2}, \quad (A.9)$$

$$F_{[1]}^{(4,2-5i/2)} = 0, \quad \forall i \geq 0, \quad (A.10)$$

$$F_{[2]}^{(5,5/2)} = 2 \left(\frac{2}{3}\right)^{1/4} M^{1/2}, \quad F_{[2]}^{(5,0)} = \frac{1}{12M^2} \left( 36\sqrt{6} M f_{[5]}^t - f_{[7]A}^t \right), \quad (A.11)$$

$$F_{[3]}^{(6,3)} = 0, \quad (A.12)$$

$$F_{[3]}^{(6,1/2)} = \frac{1}{96 \cdot 6^{3/4} M^{5/2}} \left[ 9M \left( 8\sqrt{3} M f_{[5]}^t - \sqrt{2} f_{[7]A}^t \right) - \sqrt{2} f_{[8]A}^t \right],$$

$$F_{[4]}^{(7,7/2)} = -8 \left(\frac{2}{3}\right)^{3/4} M^{3/2}, \quad (A.13)$$

$$F_{[4]}^{(7,1)} = \frac{1}{6480M^3} \left[ 108M \left( 1032M^2 f_{[5]}^t - 19\sqrt{6} M f_{[7]A}^t - f_{[9]A}^t \right) - 9\sqrt{6} M f_{[8]A}^t - \sqrt{6} f_{[9]B}^t \right],$$

$$F_{[5]}^{(8,4)} = 0,$$

$$F_{[5]}^{(8,3/2)} = -\frac{1}{31104 \cdot 6^{3/4} M^{7/2}} \left[ \sqrt{3} f_{[10]B}^t + 3\sqrt{3} f_{[10]C}^t + 36M \left( \sqrt{3} f_{[9]B}^t + 3\sqrt{2} f_{[10]D}^t + 6M \left( 9\sqrt{2} f_{[9]A}^t + 4\sqrt{3} f_{[8]A}^t + 6\sqrt{3} f_{[10]E}^t + 72\sqrt{3} M f_{[7]A}^t - 1512\sqrt{2} M^2 f_{[5]}^t \right) \right) \right], \quad (A.14)$$

$$F_{[6]}^{(9,9/2)} = -32 \cdot 2^{1/4} \cdot 3^{3/4} M^{5/2},$$

$$F_{[6]}^{(9,2)} = -\frac{1}{181440M^3} \left[ 5f_{[10]B}^t + 15f_{[10]C}^t + 8\sqrt{6} f_{[11]D}^t + 12\sqrt{6} f_{[11]E}^t + 36M \left( 5\sqrt{6} f_{[10]D}^t - 4 \left( f_{[9]B}^t - 3f_{[11]B}^t - 5f_{[11]C}^t - 9f_{[11]F}^t - 21M f_{[8]A}^t - 45M f_{[10]E}^t + 18\sqrt{6} M f_{[9]A}^t + 9324M^2 f_{[7]A}^t - 379008\sqrt{6} M^3 f_{[5]}^t \right) \right) \right], \quad (A.15)$$

$$F_{[7]}^{(10,5)} = 0,$$

$$F_{[7]}^{(10,5/2)} = -\frac{1}{62208 \cdot 6^{3/4} M^{7/2}} \left[ 2\sqrt{3} f_{[12]D}^t + 3\sqrt{3} f_{[12]L}^t + \sqrt{3} f_{[12]M}^t + 36\sqrt{3} M f_{[11]D}^t + 54\sqrt{3} M f_{[11]E}^t - 18\sqrt{2} M f_{[10]B}^t - 54\sqrt{2} M f_{[10]C}^t + 54\sqrt{2} M f_{[12]E}^t + 18\sqrt{2} M f_{[12]F}^t + 90\sqrt{2} M f_{[12]G}^t + 90\sqrt{2} M f_{[12]H}^t + 108\sqrt{2} M f_{[12]J}^t + 324\sqrt{2} M^2 f_{[9]B}^t + 972\sqrt{2} M^2 f_{[11]B}^t + 1620\sqrt{2} M^2 f_{[11]C}^t + 2916\sqrt{2} M^2 f_{[11]F}^t - 1296\sqrt{3} M^2 f_{[10]D}^t + 1296\sqrt{3} M^2 f_{[12]K}^t + 11664\sqrt{3} M^3 f_{[9]A}^t - 44064\sqrt{2} M^3 f_{[8]A}^t - 23328\sqrt{2} M^3 f_{[10]E}^t - 5184\sqrt{2} M^3 f_{[12]A}^t - 1003104\sqrt{2} M^4 f_{[7]A}^t + 8491392\sqrt{3} M^5 f_{[5]}^t \right]. \quad (A.16)$$

### A.2. Coefficients of the asymptotic near-ISCO plunge solutions

Up to 1PG order, the first coefficients in the near-ISCO solutions of the orbital radius (equations (3.29) and (3.32)) are given by

$$r_{\{0\}}^{(2,1)} = -24\sqrt{6}M^2, \quad r_{\{0\}}^{(2i+1,i+1/2)} = 0, \quad \forall i \geq 1 \tag{A.17}$$

$$r_{\{0\}}^{(4,2)} = 576M^3, \quad r_{\{0\}}^{(6,3)} = -3072\sqrt{6}M^4, \quad r_{\{0\}}^{(8,4)} = 92160M^5,$$

$$r_{\{1\}}^{(2,-3/2)} = r_{\{1\}}^{(3,-1)} = 0,$$

$$r_{\{1\}}^{(4,-1/2)} = -362^{1/4}3^{3/4}M^{3/2}f_{\{1\},0}^t,$$

$$r_{\{1\}}^{(5,0)} = -54M^2f_{\{1\},0}^r,$$

$$r_{\{1\}}^{(6,1/2)} = 362^{1/4}3^{3/4}M^{3/2}f_{\{1\},2}^t - 17282^{3/4}3^{1/4}M^{5/2}f_{\{1\},0}^t,$$

$$r_{\{1\}}^{(7,1)} = 540\sqrt{6}M^3f_{\{1\},0}^r + 182^{1/4}3^{3/4}M^{3/2}f_{\{1\},3}^t,$$

$$r_{\{1\}}^{(8,3/2)} = 187202^{1/4}3^{3/4}M^{7/2}f_{\{1\},0}^r - 8642^{3/4}3^{1/4}M^{5/2}f_{\{1\},2}^t$$

$$+ 122^{1/4}3^{3/4}M^{3/2}f_{\{1\},4}^t,$$

$$r_{\{1\}}^{(9,2)} = 216\sqrt{6}M^3f_{\{1\},2}^r - 28512M^4f_{\{1\},0}^r - 5942^{3/4}3^{1/4}M^{5/2}f_{\{1\},3}^t$$

$$+ 92^{1/4}3^{3/4}M^{3/2}f_{\{1\},5}^t. \tag{A.18}$$

The first coefficients in the near-ISCO solutions of the forcing terms (3.29) and (3.33) read

$$F_{\{0\}}^{(3,3/2)} = \frac{1}{2^{1/4}3^{7/4}M^{1/2}}, \quad F_{\{0\}}^{(2i,i)} = 0, \quad \forall i \geq 2,$$

$$F_{\{0\}}^{(5,5/2)} = 2\left(\frac{2}{3}\right)^{1/4}M^{1/2}, \quad F_{\{0\}}^{(7,7/2)} = -8\left(\frac{2}{3}\right)^{3/4}M^{3/2}, \tag{A.19}$$

$$F_{\{0\}}^{(9,9/2)} = -322^{1/4}3^{3/4}M^{5/2},$$

$$F_{\{1\}}^{(3,-1)} = \frac{f_{\{1\},0}^t}{4M^2},$$

$$F_{\{1\}}^{(4,-1/2)} = 0,$$

$$F_{\{1\}}^{(5,0)} = \frac{1}{12M^2} \left( 36\sqrt{6}Mf_{\{1\},0}^t - f_{\{1\},2}^t \right),$$

$$F_{\{1\}}^{(6,1/2)} = \frac{1}{64M^2} \left[ 6^{1/4}M^{1/2} \left( 8\sqrt{3}Mf_{\{1\},0}^r - \sqrt{2}f_{\{1\},2}^r \right) - 2f_{\{1\},3}^t \right],$$

$$F_{\{1\}}^{(7,1)} = -\frac{1}{240M^2} \left( 4f_{\{1\},4}^t + 32^{3/4}3^{1/4}M^{1/2}f_{\{1\},3}^r \right.$$

$$\left. + 76\sqrt{6}Mf_{\{1\},2}^t - 4128M^2f_{\{1\},0}^t \right),$$

$$F_{\{1\}}^{(8,3/2)} = \frac{1}{96M^2} \left[ 6^{1/4}M^{1/2} \left( 168\sqrt{2}M^2f_{\{1\},0}^r - 8\sqrt{3}Mf_{\{1\},2}^r \right) \right.$$

$$\left. - 46^{1/4}M^{1/2}f_{\{1\},3}^t - \sqrt{2}f_{\{1\},4}^r \right) - f_{\{1\},5}^t \Big],$$

$$F_{\{1\}}^{(9,2)} = -\frac{1}{560M^2} \left[ 4f_{\{1\},6}^t + 52^{3/4}3^{1/4}M^{1/2}f_{\{1\},5}^r + 4M \left( 72^{1/4}3^{3/4}M^{1/2}f_{\{1\},3}^r \right. \right.$$

$$\left. \left. - 2\sqrt{6}f_{\{1\},4}^t + 28M \left( 1504\sqrt{6}Mf_{\{1\},0}^t - 37f_{\{1\},2}^t \right) \right) \right],$$

$$F_{\{1\}}^{(10,5/2)} = -\frac{1}{384M^2} \left[ 2f_{\{1\},7}^t + 3M^{1/2} \left( 2^{3/4}3^{1/4}f_{\{1\},6}^r - 4 \left( \sqrt{6}M^{1/2}f_{\{1\},5}^t \right. \right. \right.$$

$$\left. \left. - 2^{1/4}3^{3/4}Mf_{\{1\},4}^r + 68M^{3/2}f_{\{1\},3}^t + 862^{3/4}3^{1/4}M^2f_{\{1\},2}^r \right. \right.$$

$$\left. \left. - 7282^{1/4}3^{3/4}M^3f_{\{1\},0}^r \right) \right]. \tag{A.20}$$

### A.3. Self-force matching conditions

In the transition-to-plunge regime, the self-force admits an expansion in integer powers of  $\lambda$  at fixed phase-space coordinates  $\Delta\Omega$  and  $\delta M^\pm$ ,  $f^\mu = \sum_{n=5}^\infty \lambda^n f_{[n]}^\mu(\Delta\Omega, \delta M^\pm)$ . The coefficients of  $\lambda^n$  in this expansion can be written as a sum of terms factored into a  $\Delta\Omega$ -dependent piece and a  $\Delta\Omega$ -independent

piece (labelled with capital Latin letters); see [51]. Using equation (3.27a) to expand the  $\Delta\Omega$ -dependent pieces, we find that the late-time behaviour of the transition-to-plunge self-force reads

$$\begin{aligned} f^\mu(\lambda, \Delta\Omega \rightarrow +\infty, \delta M^\pm) &= \lambda^5 f_{[5]}^\mu + \lambda^7 \Delta\Omega f_{[7]A}^\mu + \lambda^8 \left[ F_{[0]}^{(3,3/2)} \Delta\Omega^{3/2} + \mathcal{O}(\Delta\Omega^{-1}) \right] f_{[8]A}^\mu \\ &+ \lambda^9 \left[ \Delta\Omega^2 f_{[9]A}^\mu + \frac{3}{2} \left( F_{[0]}^{(3,3/2)} \right)^2 \Delta\Omega^2 f_{[9]B}^\mu + \mathcal{O}(\Delta\Omega^{-1/2}) \right] \\ &+ \mathcal{O}_{\Delta\Omega}(\lambda^{10}). \end{aligned} \quad (\text{A.21})$$

Here we use the symbol  $\mathcal{O}_{\Delta\Omega}$  to indicate subleading terms in the transition-to-plunge expansion and the symbol  $\mathcal{O}$  to indicate subleading terms in the late-time expansion of a given PLT order.

The near-ISCO solution of the self-force in the plunge regime follows from equation (3.31):

$$\begin{aligned} f^\mu(\varepsilon, \Omega \rightarrow \Omega_*, \delta M^\pm) &= \lambda^5 f_{\{1\},0}^\mu + \lambda^7 \Delta\Omega f_{\{1\},2}^\mu + \lambda^8 \Delta\Omega^{3/2} f_{\{1\},3}^\mu \\ &+ \lambda^9 \Delta\Omega^2 f_{\{1\},4}^\mu + \mathcal{O}(\lambda^{10} \Delta\Omega^{5/2}) + \mathcal{O}_\Omega(\varepsilon^2). \end{aligned} \quad (\text{A.22})$$

Here we use the symbol  $\mathcal{O}_\Omega$  to indicate subleading terms in the plunge expansion and the symbol  $\mathcal{O}$  to indicate subleading terms in the near-ISCO expansion of a given post-geodesic order.

Equating the coefficients of equal powers of  $\lambda$  and  $\Delta\Omega$  in equations (A.21) and (A.22), we obtain the following matching conditions for the self-force:

$$f_{\{1\},0}^\mu = f_{[5]}^\mu, \quad (\text{A.23a})$$

$$f_{\{1\},2}^\mu = f_{[7]A}^\mu, \quad (\text{A.23b})$$

$$f_{\{1\},3}^\mu = F_{[0]}^{(3,3/2)} f_{[8]A}^\mu, \quad (\text{A.23c})$$

$$f_{\{1\},4}^\mu = f_{[9]A}^\mu + \frac{3}{2} \left( F_{[0]}^{(3,3/2)} \right)^2 f_{[9]B}^\mu. \quad (\text{A.23d})$$

The subleading terms in the  $\Delta\Omega \rightarrow +\infty$  expansions at each order in  $\lambda$  in equation (A.21) will match with terms that originate from subleading orders in the post-geodesic expansion in equation (A.22). The coefficients in equation (A.23) can be further connected to inspiral quantities through the asymptotic match between the inspiral and transition-to-plunge regimes [51]. In particular, the terms  $f_{\{1\},0}^\mu$  and  $f_{\{1\},2}^\mu$  are given by the inspiral's first-order self-force and its  $\Omega$  derivative evaluated at the ISCO, respectively.

The asymptotic match presented in table 2 between transition to plunge (up to 7PLT order) and plunge (up to 1PG order) requires the self-force matching conditions in equations (A.23) in addition to the higher-order ones listed below:

$$f_{\{1\},5}^\mu = \frac{1}{324 2^{3/4} 3^{1/4} M^{3/2}} \left[ f_{[10]B}^\mu + 3f_{[10]C}^\mu + 36M \left( \sqrt{6} f_{[10]D}^\mu + 36M f_{[10]E}^\mu \right) \right], \quad (\text{A.24})$$

$$f_{\{1\},6}^\mu = \frac{f_{[11]B}^\mu}{3} + \frac{5}{9} f_{[11]C}^\mu + \frac{f_{[11]D}^\mu}{27\sqrt{6}M} + \frac{f_{[11]E}^\mu}{18\sqrt{6}M} + f_{[11]F}^\mu, \quad (\text{A.25})$$

$$\begin{aligned} f_{\{1\},7}^\mu &= -\frac{1}{648 2^{1/2} 3^{3/4} M^{3/2}} \left[ 10368M^3 f_{[12]A}^\mu - 2\sqrt{6} f_{[12]D}^\mu - 108M f_{[12]E}^\mu \right. \\ &- 36M f_{[12]F}^\mu - 180M f_{[12]G}^\mu - 180M f_{[12]H}^\mu - 216M f_{[12]J}^\mu \\ &\left. - 1296\sqrt{6}M^2 f_{[12]K}^\mu - 3\sqrt{6} f_{[12]L}^\mu - \sqrt{6} f_{[12]M}^\mu \right]. \end{aligned} \quad (\text{A.26})$$

## Appendix B. Transformation between the fixed- $\Omega$ and fixed- $r_p$ expansions

In this appendix we derive the transformation between the fixed- $\Omega$  formulation presented in section 3.1 and the fixed- $r_p$  formulation presented in section 3.3. As stressed in the body of the paper, the two formulations are related by a transformation on phase space, which is an inherent gauge freedom in the phase-space formalism.

We first recall the expansions of  $r_p$  (at fixed  $\Omega$ ) and  $\Omega$  (at fixed  $r_p$ ):

$$r_p = r_{\{0\}}(\Omega) + \sum_{n=1}^{\infty} \varepsilon^n r_{\{n\}}(\Omega, \delta M^{\pm}), \tag{B.1a}$$

$$\Omega = \Omega_{\{0\}}(r_p) + \sum_{n=1}^{\infty} \varepsilon^n \Omega_{\{n\}}(r_p, \delta M^{\pm}). \tag{B.1b}$$

Substituting equation (B.1a) into equation (B.1b) we derive the dictionary between the two expansions.

At zeroth order in  $\varepsilon$ , we obtain

$$\Omega = \Omega_{\{0\}}(r_{\{0\}}(\Omega)), \tag{B.2}$$

which is automatically obeyed since  $\Omega_{\{0\}}(\cdot)$  and  $r_{\{0\}}(\cdot)$  are inverse functions of one another, after choosing the correct branches for  $r_{\{0\}}$ ; see section 3.1. Taking a first and a second  $\Omega$  derivative of equation (B.2), we obtain

$$1 = r'_{\{0\}}(\Omega) \Omega'_{\{0\}}(r_{\{0\}}(\Omega)), \tag{B.3a}$$

$$0 = r''_{\{0\}}(\Omega) \Omega'_{\{0\}}(r_{\{0\}}(\Omega)) + (r'_{\{0\}}(\Omega))^2 \Omega''_{\{0\}}(r_{\{0\}}(\Omega)), \tag{B.3b}$$

where a prime denotes differentiation with respect to the argument.

At order  $\varepsilon$ , after using equation (B.3), we find from equation (B.1) that

$$r_{\{1\}}(\Omega, \delta M^{\pm}) = -r'_{\{0\}}(\Omega) \Omega_{\{1\}}(r_{\{0\}}(\Omega), \delta M^{\pm}). \tag{B.4}$$

Taking a first and a second  $\Omega$  derivative of this equation, we furthermore obtain

$$r'_{\{1\}}(\Omega, \delta M^{\pm}) = -\left(r'_{\{0\}}(\Omega)\right)^2 \Omega'_{\{1\}}(r_{\{0\}}(\Omega), \delta M^{\pm}) - r''_{\{0\}}(\Omega) \Omega_{\{1\}}(r_{\{0\}}(\Omega), \delta M^{\pm}), \tag{B.5a}$$

$$\begin{aligned} r''_{\{1\}}(\Omega, \delta M^{\pm}) = & -\left(r'_{\{0\}}(\Omega)\right)^3 \Omega''_{\{1\}}(r_{\{0\}}(\Omega), \delta M^{\pm}) \\ & - 3r'_{\{0\}}(\Omega) r''_{\{0\}}(\Omega) \Omega'_{\{1\}}(r_{\{0\}}(\Omega), \delta M^{\pm}) \\ & - r''_{\{0\}}(\Omega) \Omega_{\{1\}}(r_{\{0\}}(\Omega), \delta M^{\pm}), \end{aligned} \tag{B.5b}$$

where a prime denotes differentiation with respect to the first argument.

The self-forces in either formulation are related as

$$\sum_{n=1}^{\infty} \varepsilon^n f_{\{n\}}^{\mu} \Big|_{\text{fixed } \Omega}(\Omega, \delta M^{\pm}) = \sum_{n=1}^{\infty} \varepsilon^n f_{\{n\}}^{\mu} \Big|_{\text{fixed } r_p}(r_p, \delta M^{\pm}). \tag{B.6}$$

The relationship between the two self-forces can be obtained by plugging equation (B.1) into equation (B.6). We can express the resulting equations in terms of either  $r_p$  or  $\Omega$ . We choose to express all matching equations in terms of the orbital radius  $r_p$ . At leading order we simply have

$$f_{\{1\}}^{\mu} \Big|_{\text{fixed } r_p}(r_p, \delta M^{\pm}) = f_{\{1\}}^{\mu} \Big|_{\text{fixed } \Omega}(\Omega_{\{0\}}(r_p), \delta M^{\pm}). \tag{B.7}$$

In order to obtain the dictionary between the evolution equations, we start by considering the rate of change of the orbital frequency. In the fixed- $r_p$  formulation, from equations (B.1b) and (3.37) we have

$$\frac{d\Omega}{dt} = \frac{d\Omega}{dr_p} \frac{dr_p}{dt} = \left[ \Omega'_{\{0\}}(r_p) + \sum_{n=1}^{\infty} \varepsilon^n \Omega'_{\{n\}}(r_p, \delta M^{\pm}) \right] \left[ F_{\{0\}}^{r_p}(r_p) + \sum_{n=1}^{\infty} \varepsilon^n F_{\{n\}}^{r_p}(r_p, \delta M^{\pm}) \right]. \tag{B.8}$$

We then use equation (B.1b) to re-expand the evolution equation (3.8) in the fixed- $\Omega$  formulation,

$$\frac{d\Omega}{dt} = F_{\{0\}}^{\Omega}(\Omega_{\{0\}}(r_p)) + \varepsilon \left[ F_{\{1\}}^{\Omega}(\Omega_{\{0\}}(r_p), \delta M^{\pm}) + \partial_{\Omega} F_{\{0\}}^{\Omega}(\Omega_{\{0\}}(r_p)) \Omega_{\{1\}}(r_p, \delta M^{\pm}) \right] + \mathcal{O}(\varepsilon^2). \tag{B.9}$$

Comparing equations (B.8) and (B.9) order by order in  $\varepsilon$ , we obtain, up to 1PG order,

$$F_{\{0\}}^{r_p} (r_p) = \frac{F_{\{0\}}^\Omega (\Omega_{\{0\}} (r_p))}{\Omega'_{\{0\}} (r_p)}, \tag{B.10a}$$

$$F_{\{1\}}^{r_p} (r_p, \delta M^\pm) = \frac{1}{\Omega'_{\{0\}} (r_p)} \left[ F_{\{1\}}^\Omega (\Omega_{\{0\}} (r_p), \delta M^\pm) + \partial_\Omega F_{\{0\}}^\Omega (\Omega_{\{0\}} (r_p)) \Omega_{\{1\}} (r_p, \delta M^\pm) - F_{\{0\}}^{r_p} (r_p) \Omega'_{\{1\}} (r_p, \delta M^\pm) \right]. \tag{B.10b}$$

At geodesic order, the two formulations are mathematically equivalent, as can now be easily verified from equation (B.10a) using equations (3.25), (3.40) and (3.41). At 1PG order, the 1PG set of differential equations (3.26) and (3.48) can also be recovered one from the other; this is easily verified using the relations (B.3)–(B.5), (B.7) and (B.10).

### Appendix C. Vector and tensor spherical harmonics

We list the components of the vector and tensor spherical harmonics as defined in appendix A of [83]. The even- and odd-parity vector spherical harmonics read

$$Y_\theta^{\ell m} := \partial_\theta Y^{\ell m}, \quad Y_\phi^{\ell m} := \partial_\phi Y^{\ell m}, \tag{C.1}$$

$$X_\theta^{\ell m} := -\frac{1}{\sin \theta} \partial_\phi Y^{\ell m}, \quad X_\phi^{\ell m} := \sin \theta \partial_\theta Y^{\ell m}, \tag{C.2}$$

where  $Y^{\ell m} = Y^{\ell m}(\theta, \phi)$  are the standard scalar spherical harmonics. The even- and odd-parity tensor spherical harmonics are given by

$$Y_{\theta\theta}^{\ell m} := \left[ \partial_\theta^2 + \frac{1}{2} \ell (\ell + 1) \right] Y^{\ell m}, \tag{C.3}$$

$$Y_{\theta\phi}^{\ell m} := \left[ \partial_\theta \partial_\phi - \frac{\cos \theta}{\sin \theta} \partial_\phi \right] Y^{\ell m}, \tag{C.4}$$

$$Y_{\phi\phi}^{\ell m} := \left[ \partial_\phi^2 + \sin \theta \cos \theta \partial_\theta + \frac{1}{2} \ell (\ell + 1) \sin^2 \theta \right] Y^{\ell m}, \tag{C.5}$$

$$X_{\theta\theta}^{\ell m} := -\frac{1}{\sin \theta} \left[ \partial_\theta \partial_\phi - \frac{\cos \theta}{\sin \theta} \partial_\phi \right] Y^{\ell m}, \tag{C.6}$$

$$X_{\theta\phi}^{\ell m} := \frac{1}{2} \left[ \sin \theta \partial_\theta^2 - \frac{1}{\sin \theta} \partial_\phi^2 - \cos \theta \partial_\theta \right] Y^{\ell m}, \tag{C.7}$$

$$X_{\phi\phi}^{\ell m} := [\sin \theta \partial_\theta \partial_\phi - \cos \theta \partial_\phi] Y^{\ell m}. \tag{C.8}$$

### Appendix D. First-order Regge–Wheeler–Zerilli (RWZ) equations in generic slicing

In this appendix we extend the main steps in sections 5.1–5.4 to generic hyperboloidal slicing  $s$ . Here we require  $s$  to approach retarded time  $u$  at future null infinity and advanced time  $v$  at the horizon.

We highlight how this slicing leads to well-behaved solutions in our puncture scheme, and we explain how those solutions can be used to obtain the correct (though singular) solutions in  $t$  slicing.

#### D.1. Punctured RWZ equations

We start from the RWZ equations on phase space for a generic choice of time slicing  $s$ , as given in equation (5.6). Following the same procedure as in the main text we can split the field  $R_{\ell m}^{e/o}$  into a puncture  $R_{\ell m}^{e/o\mathcal{P}}$  and a residual field  $R_{\ell m}^{e/o\mathcal{R}}$  as

$$R_{\ell m}^{e/o} (r_p, r) = R_{\ell m}^{e/o\mathcal{R}} (r_p, r) + R_{\ell m}^{e/o\mathcal{P}} (r_p, r), \tag{D.1}$$

where

$$R_{\ell m}^{e/o\mathcal{P}} (r_p, r) := R_{\ell m}^{e/o(1)} (6M, r) \theta (r_p - r_{\mathcal{P}}). \tag{D.2}$$

The inspiral field  $R_{\ell m}^{e/o(1)}(6M, r)$ , which hereafter we also denote with the short notation  $R_{\ell m}^{e/o(1)}|_*$ , satisfies equation (5.6) evaluated at  $r_p = 6M$ :

$$\begin{aligned} & (\partial_x^2 - V_\ell^{e/o}(r)) R_{\ell m}^{e/o(1)}|_* + im \Omega_{\{0\}}|_* \left. \frac{dt_p}{ds} \right|_* \left( 2H\partial_x + \frac{dH}{dx} \right) R_{\ell m}^{e/o(1)}|_* \\ & + (1 - H^2) \left[ im \Omega_{\{0\}}|_* \left. \frac{d^2 t_p}{ds^2} \right|_* + m^2 \Omega_{\{0\}}^2|_* \left( \left. \frac{dt_p}{ds} \right|_* \right)^2 \right] R_{\ell m}^{e/o(1)}|_* = S_{\ell m}^{e/o}(6M, r) := S_{\ell m}^{e/o}|_*. \end{aligned} \quad (D.3)$$

The residual field then solves an equation structurally analogous to equation (5.6),

$$\begin{aligned} & (\partial_x^2 - V_\ell^{e/o}) R_{\ell m}^{e/o\mathcal{R}} - 2H \frac{dt_p}{ds} (\dot{r}_{\{0\}} \partial_x \partial_{r_p} - im \Omega_{\{0\}} \partial_x) R_{\ell m}^{e/o\mathcal{R}} - \frac{dH}{dx} \frac{dt_p}{ds} (\dot{r}_{\{0\}} \partial_{r_p} - im \Omega_{\{0\}}) R_{\ell m}^{e/o\mathcal{R}} \\ & + (1 - H^2) \left[ im \Omega_{\{0\}} \frac{d^2 t_p}{ds^2} + (m^2 \Omega_{\{0\}}^2 + im \partial_{r_p} \Omega_{\{0\}} \dot{r}_{\{0\}}) \left( \frac{dt_p}{ds} \right)^2 \right. \\ & \left. - \dot{r}_{\{0\}} \frac{d^2 t_p}{ds^2} \partial_{r_p} - \dot{r}_{\{0\}}^2 \left( \frac{dt_p}{ds} \right)^2 \partial_{r_p}^2 - (\ddot{r}_{\{0\}} - 2im \Omega_{\{0\}} \dot{r}_{\{0\}}) \left( \frac{dt_p}{ds} \right)^2 \partial_{r_p} \right] R_{\ell m}^{e/o\mathcal{R}} = S_{\ell m}^{e/o\text{eff}}, \end{aligned} \quad (D.4)$$

where  $S_{\ell m}^{e/o\text{eff}}(r_p, r)$  is now the effective source in  $s$  slicing. As in the main text, the effective source can be split into the ordinary point-particle source in the region  $r_p < r_{\mathcal{P}}$ ,

$$S_{\ell m}^{e/o\text{pp}}(r_p, r) := S_{\ell m}^{e/o}(r_p, r) \theta(r_{\mathcal{P}} - r_p), \quad (D.5)$$

and the following extended source in the region  $6M \geq r_p \geq r_{\mathcal{P}}$ :

$$\begin{aligned} S_{\ell m}^{e/o\text{ext}}(r_p, r) & := \theta(r_p - r_{\mathcal{P}}) [S_{\ell m}^{e/o}(r_p, r) - S_{\ell m}^{e/o}(6M, r)] \\ & - \theta(r_p - r_{\mathcal{P}}) im \left( \Omega_{\{0\}} \frac{dt_p}{ds} - \Omega_{\{0\}}|_* \left. \frac{dt_p}{ds} \right|_* \right) \left( 2H\partial_x + \frac{dH}{dx} \right) R_{\ell m}^{e/o(1)}|_* \\ & - \theta(r_p - r_{\mathcal{P}}) (1 - H^2) \left[ im \left( \Omega_{\{0\}} \frac{d^2 t_p}{ds^2} - \Omega_{\{0\}}|_* \left. \frac{d^2 t_p}{ds^2} \right|_* \right) \right. \\ & \left. + \left( \frac{dt_p}{ds} \right)^2 (m^2 \Omega_{\{0\}}^2 + im \partial_{r_p} \Omega_{\{0\}} \dot{r}_{\{0\}}) - \left( \frac{dt_p}{ds} \right)^2 m^2 \Omega_{\{0\}}^2|_* \right] R_{\ell m}^{e/o(1)}|_* \\ & + \delta(r_p - r_{\mathcal{P}}) \left[ \left. \frac{dt_p}{ds} \right|_{\mathcal{P}} \dot{r}_{\{0\}}|_{\mathcal{P}} \left( 2H\partial_x + \frac{dH}{dx} \right) R_{\ell m}^{e/o(1)}|_* \right. \\ & \left. - (1 - H^2) \left( \left. \frac{d^2 t_p}{ds^2} \right|_{\mathcal{P}} \dot{r}_{\{0\}}|_{\mathcal{P}} + \left( \left. \frac{dt_p}{ds} \right|_{\mathcal{P}} \right)^2 (\ddot{r}_{\{0\}}|_{\mathcal{P}} + 2im \Omega_{\{0\}}|_{\mathcal{P}} \dot{r}_{\{0\}}|_{\mathcal{P}}) \right) R_{\ell m}^{e/o(1)}|_* \right] \\ & + \partial_{r_p} \delta(r_p - r_{\mathcal{P}}) (1 - H^2) \left( \left. \frac{dt_p}{ds} \right|_{\mathcal{P}} \right)^2 \dot{r}_{\{0\}}^2|_{\mathcal{P}} R_{\ell m}^{e/o(1)}|_*. \end{aligned} \quad (D.6)$$

## D.2. Frequency-domain equations

We next introduce the following forward and inverse transforms:

$$\hat{g}(\omega, r) = \int_{2M}^{6M} \frac{dr_p}{\dot{r}_{\{0\}}(r_p) \left. \frac{dt_p}{ds} \right|_{r_p}} g(r_p, r) e^{i\omega s_G(r_p) - im\phi_G(r_p)}, \quad (D.7a)$$

$$g(r_p, r) = -\frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega \hat{g}(\omega, r) e^{-i\omega s_G(r_p) + im\phi_G(r_p)}, \quad (D.7b)$$

where

$$s_G(r_p) := \int^{r_p} \frac{dr'_p}{\dot{r}'_{\{0\}}(r'_p)} \frac{ds}{dt_p}(r'_p) \quad (D.8)$$

and  $dt_p/ds$  is given in equation (4.8).

The transform of equation (D.4) reads

$$(\partial_x^2 - V_\ell^{e/o}) \hat{R}_{\ell m}^{e/o \mathcal{R}} + i\omega \left( 2H\partial_x + \frac{dH}{dx} \right) \hat{R}_{\ell m}^{e/o \mathcal{R}} + (1 - H^2) \omega^2 \hat{R}_{\ell m}^{e/o \mathcal{R}} = \hat{S}_{\ell m}^{e/o \text{ eff}}, \quad (\text{D.9})$$

where, following the same reasoning as in section 5.3, no boundary terms appear. The transformed effective source  $\hat{S}_{\ell m}^{e/o \text{ eff}} = \hat{S}_{\ell m}^{e/o \text{ pp}} + \hat{S}_{\ell m}^{e/o \text{ ext}}$  is given by

$$\hat{S}_{\ell m}^{e/o \text{ pp}} := \int_{2M}^{r_{\mathcal{P}}} \frac{dr_p e^{i\omega s_G - im\phi_G}}{\dot{r}_{\{0\}} dt_p/ds} S_{\ell m}^{e/o}, \quad (\text{D.10a})$$

$$\begin{aligned} \hat{S}_{\ell m}^{e/o \text{ ext}} := & \int_{r_{\mathcal{P}}}^{6M} \frac{dr_p e^{i\omega s_G - im\phi_G}}{\dot{r}_{\{0\}} dt_p/ds} (S_{\ell m}^{e/o} - S_{\ell m}^{e/o}|_*) \\ & - \mathcal{I}_m^1(r_{\mathcal{P}}; \omega) \left( 2H\partial_x + \frac{dH}{dx} \right) R_{\ell m}^{e/o(1)} \Big|_* - (1 - H^2) \mathcal{I}_m^2(r_{\mathcal{P}}; \omega) R_{\ell m}^{e/o(1)} \Big|_* \end{aligned} \quad (\text{D.10b})$$

$$+ \left[ 2H\partial_x + \frac{dH}{dx} - i(1 - H^2) \left( \omega + m \Omega_{\{0\}} \Big|_{\mathcal{P}} \frac{dt_p}{ds} \Big|_{\mathcal{P}} \right) \right] R_{\ell m}^{e/o(1)} \Big|_* e^{i\omega s_G(r_{\mathcal{P}}) - im\phi_G(r_{\mathcal{P}})},$$

with

$$\mathcal{I}_m^1(r_{\mathcal{P}}; \omega) := \int_{r_{\mathcal{P}}}^{6M} \frac{dr_p e^{i\omega s_G - im\phi_G}}{\dot{r}_{\{0\}} dt_p/ds} \left[ im \left( \Omega_{\{0\}} \frac{dt_p}{ds} - \Omega_{\{0\}} \Big|_* \frac{dt_p}{ds} \Big|_* \right) \right], \quad (\text{D.11a})$$

$$\begin{aligned} \mathcal{I}_m^2(r_{\mathcal{P}}; \omega) := & \int_{r_{\mathcal{P}}}^{6M} \frac{dr_p e^{i\omega s_G - im\phi_G}}{\dot{r}_{\{0\}} dt_p/ds} \left[ im \left( \Omega_{\{0\}} \frac{d^2 t_p}{ds^2} - \Omega_{\{0\}} \Big|_* \frac{d^2 t_p}{ds^2} \Big|_* \right) \right. \\ & \left. + \left( \frac{dt_p}{ds} \right)^2 \left( m^2 \Omega_{\{0\}}^2 + im \partial_{r_p} \Omega_{\{0\}} \dot{r}_{\{0\}} \right) - \left( \frac{dt_p}{ds} \Big|_* \right)^2 m^2 \Omega_{\{0\}}^2 \Big|_* \right]. \end{aligned} \quad (\text{D.11b})$$

### D.3. Inhomogeneous solutions

The homogeneous solutions to equation (D.9) are related to the homogeneous solutions to equation (5.27) simply by an overall factor of  $e^{-i\omega\kappa(x)}$ . Recalling the  $t$ -slicing solutions in equation (5.28), we readily obtain the ‘in’ and ‘up’ solutions in  $s$  slicing as

$$\hat{R}_\ell^{e/o \text{ in}}(\omega, x) \sim \begin{cases} A_\ell^{e/o \text{ in}}(\omega) e^{-2i\omega x} + A_\ell^{e/o \text{ out}}(\omega) & \text{as } r \rightarrow +\infty (x \rightarrow +\infty), \\ 1 & \text{as } r \rightarrow 2M (x \rightarrow -\infty), \end{cases} \quad (\text{D.12a})$$

$$\hat{R}_\ell^{e/o \text{ up}}(\omega, x) \sim \begin{cases} 1 & \text{as } r \rightarrow +\infty (x \rightarrow +\infty), \\ B_\ell^{e/o \text{ in}}(\omega) + B_\ell^{e/o \text{ out}}(\omega) e^{+2i\omega x} & \text{as } r \rightarrow 2M (x \rightarrow -\infty). \end{cases} \quad (\text{D.12b})$$

Using the two independent homogeneous solutions to equation (D.9),  $\hat{R}_\ell^{e/o \text{ in}}$  and  $\hat{R}_\ell^{e/o \text{ up}}$ , we construct the following Green function:

$$\begin{aligned} \hat{G}_\ell^{e/o}(\omega, x, x') = & \frac{1}{W_\ell(\omega)} \left[ \theta(x - x') \hat{R}_\ell^{e/o \text{ in}}(\omega, x') \hat{R}_\ell^{e/o \text{ up}}(\omega, x) \right. \\ & \left. + \theta(x' - x) \hat{R}_\ell^{e/o \text{ in}}(\omega, x) \hat{R}_\ell^{e/o \text{ up}}(\omega, x') \right], \end{aligned} \quad (\text{D.13})$$

with

$$W_\ell(\omega) = e^{2i\omega\kappa(x)} \left[ \hat{R}_\ell^{e/o \text{ in}}(\omega, x) \partial_x \hat{R}_\ell^{e/o \text{ up}}(\omega, x) - \hat{R}_\ell^{e/o \text{ up}}(\omega, x) \partial_x \hat{R}_\ell^{e/o \text{ in}}(\omega, x) \right]. \quad (\text{D.14})$$

The Green function satisfies

$$\left[ (\partial_x^2 - V_\ell^{e/o}) + i\omega \left( 2H\partial_x + \frac{dH}{dx} \right) + (1 - H^2) \omega^2 \right] \hat{G}_\ell^{e/o}(\omega, x, x') = \frac{\delta(x - x')}{e^{2i\omega\kappa(x)}}. \quad (\text{D.15})$$

It is straightforward to verify that  $W_\ell$  is independent of the field point  $r$ ,  $\partial_x W_\ell = 0$ , and it can therefore be computed from the asymptotic solutions (D.12) at  $r \rightarrow +\infty$ , leading to

$$W_\ell(\omega) = 2i\omega A_\ell^{\text{in}}(\omega). \quad (\text{D.16})$$

The inhomogeneous solution can then be obtained as

$$\begin{aligned} \hat{R}_{\ell m}^{e/o \mathcal{R}}(\omega, r) &= \int_{-\infty}^{+\infty} dx' \hat{G}_{\ell}^{e/o}(\omega, x, x') e^{2i\omega\kappa(x')} \hat{S}_{\ell m}^{e/o \text{ eff}}(\omega, x') \\ &= \frac{\hat{R}_{\ell}^{e/o \text{ up}}}{2i\omega A_{\ell}^{\text{in}}(\omega)} \int_{2M}^r \frac{dr'}{f(r')} \hat{R}_{\ell}^{e/o \text{ in}}(\omega, r') e^{2i\omega\kappa(r')} \hat{S}_{\ell m}^{e/o \text{ eff}}(\omega, r') \\ &\quad + \frac{\hat{R}_{\ell}^{e/o \text{ in}}}{2i\omega A_{\ell}^{\text{in}}(\omega)} \int_r^{+\infty} \frac{dr'}{f(r')} \hat{R}_{\ell}^{e/o \text{ up}}(\omega, r') e^{2i\omega\kappa(r')} \hat{S}_{\ell m}^{e/o \text{ eff}}(\omega, r'). \end{aligned} \tag{D.17}$$

**D.4. Regularity, slicing, and punctures at the boundaries**

To assess regularity near the boundaries, we consider  $s = u$  ( $\kappa = x, H = +1$ ) at large  $r$  and  $s = v$  ( $\kappa = -x, H = -1$ ) near  $r = 2M$ . In either case, the effective source (D.10b) in a neighbourhood of each boundary reduces to

$$\hat{S}_{\ell m}^{e/o \text{ ext}} = \mp 2 \left[ \mathcal{I}_m^1(r_{\mathcal{P}}; \omega) - e^{i\omega s_G(r_{\mathcal{P}}) - im\phi_G(r_{\mathcal{P}})} \right] \partial_x R_{\ell m}^{e/o(1)} \Big|_*. \tag{D.18}$$

At large  $r$ ,  $R_{\ell m}^{e/o(1)}|_*$  behaves as a regular series in  $1/r$ , beginning at  $r^0$ , such that  $\hat{S}_{\ell m}^{e/o \text{ ext}} = \mathcal{O}(1/r^2)$ . At the horizon,  $R_{\ell m}^{e/o(1)}|_*$  behaves as a regular series in  $(r - 2M)$ , beginning at  $(r - 2M)^0$ , and  $\partial_x = f\partial_r$ , such that  $\hat{S}_{\ell m}^{e/o \text{ ext}} = \mathcal{O}(r - 2M)$ . Referring to the asymptotic behaviour (D.12), we then see that (i) the integrals in equation (D.17) converge, (ii) at the horizon, the solution behaves as the regular homogeneous solution  $\hat{R}_{\ell}^{e/o \text{ in}}$ , and (iii) at infinity, the solution behaves as the regular homogeneous solution  $\hat{R}_{\ell}^{e/o \text{ up}}$ .

If we were using  $t$  slicing rather than hyperboloidal slicing, the integrals in equation (D.17) would not converge, nor would either integral vanish in the limits  $r \rightarrow 2M$  and  $r \rightarrow \infty$ . With  $s = t$  ( $H = 0$ ), the effective source (D.10b) behaves like  $R_{\ell m}^{e/o(1)}|_*$  at the boundaries. In  $t$  slicing,  $R_{\ell m}^{e/o(1)}|_*$  behaves as  $e^{\pm i\omega x}$  toward the boundaries, and both integrals in equation (D.17) then diverge for any value of  $r$ .

Such ill behaviour is familiar from infrared divergences that arise in the inspiral multiscale expansion in  $t$  slicing [43, 70–72]. As in that case, we can obtain the physically correct solution by introducing punctures at the boundaries. The physically correct solution in this case is obtained in hyperboloidal slicing, where the field behaves as a smooth outgoing wave at future null infinity and a smooth ingoing one at the horizon. To obtain the correct boundary conditions in  $t$  slicing, we can transform the hyperboloidal-slicing solution  $\hat{R}_{\ell m}^{e/o \mathcal{R}}$  to  $t$  slicing, expand it near the horizon and infinity, and treat the results (which will be singular at the boundaries) as punctures.

To transform the hyperboloidal solution to  $t$  slicing, we first transform it back to phase space, using equation (D.7b). At a given point in spacetime, the field in  $s$  slicing is then  ${}_{[s]}R_{\ell m}^{e/o}(r_p(t_p(s)), r)e^{-im\phi_p(t_p(s))}$ , where we add a left-subscript to indicate the slicing; at the same point, the field in  $t$  slicing is  ${}_{[t]}R_{\ell m}^{e/o}(r_p(t), r)e^{-im\phi_p(t)}$ . Since the field at a given point must be independent slicing, we can infer

$${}_{[t]}R_{\ell m}^{e/o \mathcal{R}}(r_p(t), r) = {}_{[s]}R_{\ell m}^{e/o \mathcal{R}}(r_p(t_p(s)), r) e^{im[\phi_p(t) - \phi_p(t_p(s))]} \tag{D.19}$$

For this to give us the  $t$ -slicing field variable, the right-hand side must be expressed entirely in terms of the independent variables  $r_p(t)$  and  $r$ . We put it in that form by expanding for  $r_p(t)$  near the ISCO; this suffices because we only require the boundary punctures in the window  $6M > r_p > r_{\mathcal{P}}$ .

Define  $\Delta t := t_p(s) - t$ . Then

$$r_p(t_p(s)) = r_p(t) + \dot{r}_p(r_p(t)) \Delta t + \frac{1}{2} \ddot{r}_p(r_p(t)) (\Delta t)^2 + \dots \tag{D.20}$$

$$\phi_p(t_p(s)) = \phi_p(t) + \Omega_{\{0\}}(r_p(t)) \Delta t + \frac{1}{2} \dot{\Omega}_{\{0\}}(r_p(t)) (\Delta t)^2 + \dots \tag{D.21}$$

Although  $\Delta t$  need not be small, these expansions are well behaved because successive terms are smaller in the limit  $r_p \rightarrow 6M$ . Concretely, equations (3.15) and (3.16) imply  $\dot{r}_p = \mathcal{O}[(6M - r_p)^{3/2}] = \dot{\Omega}_{\{0\}}$ , and each additional derivative introduces  $d/dt = \mathcal{O}[(6M - r_p)^{1/2}]$ .

Given that  $s = t - \kappa(r)$ , we have  $t_p(s) = s + \kappa(r_p(t_p(s)))$  and so  $\Delta t = \kappa(r_p(t_p(s))) - \kappa(r)$ . Formally expanding for small time derivatives and appealing to equation (D.20) then yields

$$\Delta t = \Delta t_0(r_p, r) + \Delta t_1(r_p, r) + \dots, \tag{D.22}$$

where  $\Delta t_0 := [\kappa(r_p) - \kappa(r)]$ ,  $\Delta t_1 := \kappa'(r_p)\dot{r}_p(r_p)\Delta t_0(r_p, r)$ ,  $\kappa'(r_p) = \partial_{r_p}\kappa(r_p)$ , and here it is understood that  $r_p = r_p(t)$ . Equation (D.19) then becomes

$$\begin{aligned}
 [t]R_{\ell m}^{e/o\mathcal{R}}(r_p, r) &= e^{-im\Omega_{\{0\}}\Delta t_0} \left\{ [s]R_{\ell m}^{e/o\mathcal{R}}(r_p, r) \right. \\
 &\quad \left. + \left[ \dot{r}_p\Delta t_0\partial_{r_p} - im\left( \Omega_{\{0\}}\Delta t_1 + \frac{1}{2}\dot{\Omega}_{\{0\}}(\Delta t_0)^2 \right) \right] [s]R_{\ell m}^{e/o\mathcal{R}}(r_p, r) + \dots \right\}, \tag{D.23}
 \end{aligned}$$

where again it is understood that  $r_p = r_p(t)$ . The first subleading term, in square brackets, scales as  $(r_p - 6M)^{3/2}$ , and the ellipses denote terms of order  $(r_p - 6M)^2$  and smaller.

The transformation (D.23) can be compared to equation (161) in [43], which represents the analogous result in the inspiral. Like in the inspiral, the  $t$ -slicing field  $[t]R_{\ell m}^{e/o\mathcal{R}}$  is singular at the boundaries. This can be seen from the factors of  $\kappa(r)$  in  $\Delta t_n$ , which reduce to  $\pm x$  near the boundaries, diverging logarithmically when  $r \rightarrow 2M$  and linearly when  $r \rightarrow \infty$ .

Finally, to obtain the behaviour of the frequency-domain,  $t$ -slicing solution near the boundaries, we expand equation (D.23) in the limits  $r \rightarrow 2M$  and  $r \rightarrow \infty$  and then apply the Fourier transform (5.25a). This defines for us our punctures at the boundaries, which we move to the right-hand side of the  $t$ -slicing field equation (5.27), defining a new effective source. Such procedures are detailed extensively in [70]. Since the punctures are necessarily approximate particular solutions, they cancel the old effective source near the boundaries. If the near-horizon and large- $r$  expansions of the punctures are carried to sufficiently high order, the new effective source will behave sufficiently well toward the boundaries, the Green-function integrals over  $r'$  will converge, and the sum of the new residual field plus the new punctures will be precisely equivalent to the hyperboloidal solution, simply expressed in terms of a Fourier transform based on the singular  $t$  coordinate.

### Appendix E. Contribution from the early-time effective source

In this appendix we consider the solution to equation (5.19) with the extended source (5.22). Our aim is to show that its contribution can be neglected if we only require the solution sufficiently far to the future of the puncture window (i.e. for values of  $r_p$  not too close to  $6M$ ).

We start by obtaining the source's transform (5.25a):

$$\begin{aligned}
 \hat{S}_{\ell m}^{e/o\text{ ext}}(\omega, r) &= \int_{r_{\mathcal{P}}}^{6M} \frac{dr_p}{\dot{r}_{\{0\}}} e^{i\omega t_G - im\phi_G} [S_{\ell m}^{e/o}(r_p, r) - S_{\ell m}^{e/o}(6M, r)] \\
 &\quad - R_{\ell m}^{e/o(1)}(6M, r) \int_{r_{\mathcal{P}}}^{6M} \frac{dr_p}{\dot{r}_{\{0\}}} e^{i\omega t_G - im\phi_G} \left( m^2 \Omega_{\{0\}}^2 - m^2 \Omega_{\{0\}}^2 \Big|_* + im\partial_{r_p}\Omega_{\{0\}}\dot{r}_{\{0\}} \right) \\
 &\quad - i(\omega + m\Omega_{\{0\}} \Big|_{\mathcal{P}}) e^{i\omega t_G(r_{\mathcal{P}}) - im\phi_G(r_{\mathcal{P}})} R_{\ell m}^{e/o(1)}(6M, r). \tag{E.1}
 \end{aligned}$$

We anticipate that we are interested in the limit  $r_{\mathcal{P}} \rightarrow 6M$ . In this limit, the  $r_p$  integrals in equation (E.1) vanish. We therefore only focus on the solution at infinity sourced by the third term in equation (E.1).

Using the same Green-function technique we have adopted in section 5.4, we obtain

$$\begin{aligned}
 R_{\ell m}^{e/o\mathcal{R}} \Big|_{\text{ext}\infty}(r_p) &= -\frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega e^{-i\omega t_G(r_p) + im\phi_G(r_p)} \\
 &\quad \times \frac{e^{i\omega x}}{2i\omega A_{\ell}^{\text{in}}(\omega)} \int_{2M}^{\infty} \frac{dr'}{f(r')} \hat{R}_{\ell}^{e/o\text{ in}}(\omega, r') \hat{S}_{\ell m}^{e/o\text{ ext}}(\omega, r'). \tag{E.2}
 \end{aligned}$$

As discussed in appendix D.4, the radial integral here does not converge if we use the effective source described in the body of the paper, due to the poor behaviour of the source in  $t$  slicing. We adopt the prescription in appendix D.4 to avoid that divergence. We can then write the equation above as

$$R_{\ell m}^{e/o\mathcal{R}} \Big|_{\text{ext}\infty}(r_p) = -\frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega F_{\ell m}^{e/o}(\omega; r_p, r_{\mathcal{P}}) e^{i\omega[t_G(r_{\mathcal{P}}) - t_G(r_p) + x]}, \tag{E.3}$$

where we have defined

$$\begin{aligned}
 F_{\ell m}^{e/o}(\omega; r_p, r_{\mathcal{P}}) &:= -\frac{e^{im[\phi_G(r_p) - \phi_G(r_{\mathcal{P}})]}}{2A_{\ell}^{\text{in}}(\omega)} \left( 1 + \frac{m\Omega_{\{0\}} \Big|_{\mathcal{P}}}{\omega} \right) \\
 &\quad \times \int_{2M}^{\infty} \frac{dr'}{f(r')} \hat{R}_{\ell}^{e/o\text{ in}}(\omega, r') R_{\ell m}^{e/o(1)}(6M, r'). \tag{E.4}
 \end{aligned}$$

The integral (E.3) contains a rapidly oscillating integrand when  $r_{\mathcal{P}}$  is near  $6M$  (and  $r_p$  is not too close to  $6M$ ). Integrating by parts  $N + 1$  times, we find

$$R_{\ell m}^{e/o} \mathcal{R} |_{\text{ext } \infty} (r_p) = \frac{1}{2\pi} \sum_{n=0}^N \left[ i^{n+1} \frac{d^n F_{\ell m}^{e/o}(\omega; r_p, r_{\mathcal{P}})}{d\omega^n} \frac{e^{i\omega[t_G(r_{\mathcal{P}}) - t_G(r_p) + x]}}{[t_G(r_{\mathcal{P}}) - t_G(r_p) + x]^{n+1}} \right]_{-\infty}^{+\infty} + \mathcal{O}(N + 1), \tag{E.5}$$

where  $\mathcal{O}(N + 1)$  denotes the remaining integral, which is proportional to  $1/[t_G(r_{\mathcal{P}}) - t_G(r_p) + x]^{N+1}$ . Assuming the solution (E.3) exists,  $F_{\ell m}^{e/o}$  must decay to zero as  $\omega \rightarrow \pm\infty$ , implying its derivatives do likewise. Hence, the sum in equation (E.5) vanishes, leaving the  $\mathcal{O}(N + 1)$  remainder. If  $r_{\mathcal{P}}$  is taken toward  $6M$ , then  $t_G(r_{\mathcal{P}}) \rightarrow -\infty$ , causing the  $\mathcal{O}(N + 1)$  remainder to go to zero. Since  $N$  is arbitrary, it follows that  $R_{\ell m}^{e/o} \mathcal{R} |_{\text{ext } \infty} (r_p)$  goes to zero faster than any power of  $1/t_G(r_{\mathcal{P}})$  as  $r_{\mathcal{P}} \rightarrow 6M$  (at fixed values of  $r_p$  and  $x$ ).

### Appendix F. Regularization procedure for the excitation coefficients

In this appendix we justify the regularization procedure introduced in section 5.6. We start by considering equation (5.40),

$$R_{\ell m}^{e/o} \mathcal{R} |_{\text{pp } \infty} (r_p) e^{-im\phi_G(r_p)} = -\frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega e^{-i\omega[t_G(r_p) - x]} \frac{C_{\ell m}^{e/o}(\omega)}{2i\omega A_{\ell}^{\text{in}}(\omega)}, \tag{E.1}$$

where the integration along the real  $\omega$  axis gives a finite result. The numerator of the integrand,  $C_{\ell m}^{e/o}(\omega)$ , is given by the radial integral (5.41), which we rewrite here for convenience:

$$C_{\ell m}^{e/o}(\omega) := \int_{2M}^{r_{\mathcal{P}}} \frac{dr'}{\dot{r}_{\{0\}}(r')} e^{i\omega t_G(r') - im\phi_G(r')} K_{\ell m}^{e/o}(\omega, r'). \tag{E.2}$$

Now, analogously to the procedure described at the end of section 5.6, we can rewrite  $C_{\ell m}^{e/o}(\omega)$  by subtracting the near-horizon behaviour  $q_{\ell m}^{e/o}$ , given in equation (5.54), from the integrand and adding back its antiderivative  $Q_{\ell m}^{e/o}$ :

$$C_{\ell m}^{e/o}(\omega) = C_{\ell m}^{e/o} |_{\text{reg}}(\omega) - Q_{\ell m}^{e/o}(\omega, 2M). \tag{E.3}$$

Integrating equation (5.54) gives  $Q_{\ell m}^{e/o}(\omega, r) \sim (r - 2M)^{1-4im\omega}$ , which vanishes at the horizon. Therefore, equation (5.40) is equivalent to

$$R_{\ell m}^{e/o} \mathcal{R} |_{\text{pp } \infty} (r_p) e^{-im\phi_G(r_p)} = -\frac{1}{2\pi} \int_{-\infty}^{+\infty} d\omega e^{-i\omega[t_G(r_p) - x]} \frac{C_{\ell m}^{e/o} |_{\text{reg}}(\omega)}{2i\omega A_{\ell}^{\text{in}}(\omega)}. \tag{E.4}$$

After closing the contour in the complex plane and using the residue theorem, this then justifies computing the excitation factors appearing in equation (5.49) by using  $C_{\ell m}^{e/o} |_{\text{reg}}$  rather than simply  $C_{\ell m}^{e/o}$  in equation (5.50).

### ORCID iDs

Lorenzo Küchler  0000-0002-6609-1684  
 Geoffrey Compère  0000-0002-1977-3295  
 Adam Pound  0000-0001-9446-0638

### References

- [1] Abbott B P et al (LIGO Scientific and Virgo Collaboration) 2016 Observation of gravitational waves from a binary black hole merger *Phys. Rev. Lett.* **116** 061102
- [2] Abbott B P et al (LIGO Scientific and Virgo Collaboration) 2019 GWTC-1: a gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs *Phys. Rev. X* **9** 031040
- [3] Abbott B P et al (LIGO Scientific and Virgo Collaboration) 2021 GWTC-2: compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run *Phys. Rev. X* **11** 021053
- [4] Abbott B P et al (KAGRA, VIRGO and LIGO Scientific Collaboration) 2023 GWTC-3: compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run *Phys. Rev. X* **13** 041039
- [5] Maggiore M et al (ET Collaboration) 2020 Science case for the Einstein telescope *J. Cosmol. Astropart. Phys.* **JCAP03(2020)050**
- [6] Evans M et al 2021 A horizon study for cosmic explorer: science, observatories, and community (arXiv:2109.09882)

- [7] Baibhav V, Berti E, Gerosa D, Mapelli M, Giacobbo N, Bouffanais Y and Di Carlo U N 2019 Gravitational-wave detection rates for compact binaries formed in isolation: LIGO/Virgo O3 and beyond *Phys. Rev. D* **100** 064060
- [8] Kalogera V et al 2021 The next generation global gravitational wave observatory: the science book (arXiv:2111.06990)
- [9] Amaro-Seoane P et al 2017 Laser interferometer space antenna (arXiv:1702.00786)
- [10] Babak S, Gair J, Sesana A, Barausse E, Sopuerta C F, Berry C P L, Berti E, Amaro-Seoane P, Petiteau A and Klein A 2017 Science with the space-based interferometer LISA. V: Extreme mass-ratio inspirals *Phys. Rev. D* **95** 103012
- [11] Seoane P A et al (LISA Collaboration) 2023 Astrophysics with the laser interferometer space antenna *Living Rev. Relativ.* **26** 2
- [12] Barausse E, Dvorkin I, Tremmel M, Volonteri M and Bonetti M 2020 Massive black hole merger rates: the effect of kiloparsec separation wandering and supernova feedback *Astrophys. J.* **904** 16
- [13] Mangiagli A, Caprini C, Volonteri M, Marsat S, Vergani S, Tamanini N and Inchauspé H 2022 Massive black hole binaries in LISA: multimessenger prospects and electromagnetic counterparts *Phys. Rev. D* **106** 103017
- [14] Bellovary J, Luo Y, Quinn T, Munshi F, Tremmel M and Wadsley J 2024 Intermediate mass ratio inspirals in milky way galaxies (arXiv:2411.12117)
- [15] Jani K, Shoemaker D and Cutler C 2019 Detectability of intermediate-mass black holes in multiband gravitational wave astronomy *Nat. Astron.* **4** 260–5
- [16] Afshordi N et al (LISA Consortium Waveform Working Group Collaboration) 2023 Waveform modelling for the laser interferometer space antenna (arXiv:2311.01300)
- [17] Damour T 2010 Gravitational self force in a schwarzschild background and the effective one body formalism *Phys. Rev. D* **81** 024017
- [18] Nagar A and Albanesi S 2022 Toward a gravitational self-force-informed effective-one-body waveform model for nonprecessing, eccentric, large-mass-ratio inspirals *Phys. Rev. D* **106** 064049
- [19] van de Meent M, Buonanno A, Mihaylov D P, Ossokine S, Pompili L, Warburton N, Pound A, Wardell B, Durkan L and Miller J 2023 Enhancing the SEOBNRv5 effective-one-body waveform model with second-order gravitational self-force fluxes *Phys. Rev. D* **108** 124038
- [20] Leather B, Buonanno A and van de Meent M 2025 Inspiral-merger-ringdown waveforms with gravitational self-force results within the effective-one-body formalism (arXiv:2505.11242)
- [21] Wardell B, Pound A, Warburton N, Miller J, Durkan L and Le Tiec A 2023 Gravitational waveforms for compact binaries from second-order self-force theory *Phys. Rev. Lett.* **130** 241402
- [22] Albertini A, Nagar A, Pound A, Warburton N, Wardell B, Durkan L and Miller J 2022 Comparing second-order gravitational self-force, numerical relativity and effective one body waveforms from inspiralling, quasicircular and nonspinning black hole binaries *Phys. Rev. D* **106** 084061
- [23] Katz M L, Chua A J K, Speri L, Warburton N and Hughes S A 2021 Fast extreme-mass-ratio-inspiral waveforms: new tools for millihertz gravitational-wave data analysis *Phys. Rev. D* **104** 064047
- [24] Amaro-Seoane P 2018 Relativistic dynamics and extreme mass ratio inspirals *Living Rev. Relativ.* **21** 4
- [25] Abac A et al 2025 The science of the Einstein telescope (arXiv:2503.12263)
- [26] Blanchet L 2014 Post-Newtonian theory for gravitational waves *Living Rev. Relativ.* **17** 2
- [27] Barack L and Pound A 2019 Self-force and radiation reaction in general relativity *Rept. Prog. Phys.* **82** 016904
- [28] Pound A and Wardell B 2022 Black hole perturbation theory and gravitational self-force *Handbook of Gravitational Wave Astronomy* p 38
- [29] Duez M D and Zlochower Y 2019 Numerical relativity of compact binaries in the 21st century *Rep. Prog. Phys.* **82** 016902
- [30] London L, Shoemaker D and Healy J 2014 Modeling ringdown: beyond the fundamental quasinormal modes *Phys. Rev. D* **90** 124032
- London L, Shoemaker D and Healy J 2016 *Phys. Rev. D* **94** 069902 (Erratum)
- [31] Berti E et al 2025 Black hole spectroscopy: from theory to experiment (arXiv:2505.23895)
- [32] Buonanno A and Damour T 1999 Effective one-body approach to general relativistic two-body dynamics *Phys. Rev. D* **59** 084006
- [33] Buonanno A and Damour T 2000 Transition from inspiral to plunge in binary black hole coalescences *Phys. Rev. D* **62** 064015
- [34] Damour T, Iyer B R, Jaranowski P and Sathyaprakash B S 2003 Gravitational waves from black hole binary inspiral and merger: The Span of third postNewtonian effective one-body templates *Phys. Rev. D* **67** 064028
- [35] Damour T and Nagar A 2007 Faithful effective-one-body waveforms of small-mass-ratio coalescing black-hole binaries *Phys. Rev. D* **76** 064028
- [36] Buonanno A, Pan Y, Pfeiffer H P, Scheel M A, Buchman L T and Kidder L E 2009 Effective-one-body waveforms calibrated to numerical relativity simulations: coalescence of non-spinning, equal-mass black holes *Phys. Rev. D* **79** 124028
- [37] Taracchini A, Buonanno A, Khanna G and Hughes S A 2014 Small mass plunging into a Kerr black hole: Anatomy of the inspiral-merger-ringdown waveforms *Phys. Rev. D* **90** 084025
- [38] Riemenschneider G, Retegno P, Breschi M, Albertini A, Gamba R, Bernuzzi S and Nagar A 2021 Assessment of consistent next-to-quasicircular corrections and postadiabatic approximation in effective-one-body multipolar waveforms for binary black hole coalescences *Phys. Rev. D* **104** 104045
- [39] Ajith P et al 2007 Phenomenological template family for black-hole coalescence waveforms *Class. Quantum Grav.* **24** S689–700
- [40] Ajith P et al 2008 A Template bank for gravitational waveforms from coalescing binary black holes. I. Non-spinning binaries *Phys. Rev. D* **77** 104017
- Ajith P et al 2009 *Phys. Rev. D* **79** 129901 (Erratum)
- [41] Ajith P et al 2011 Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins *Phys. Rev. Lett.* **106** 241101
- [42] Santamaria L et al 2010 Matching post-Newtonian and numerical relativity waveforms: systematic errors and a new phenomenological model for non-precessing black hole binaries *Phys. Rev. D* **82** 064016
- [43] Miller J and Pound A 2021 Two-timescale evolution of extreme-mass-ratio inspirals: waveform generation scheme for quasicircular orbits in Schwarzschild spacetime *Phys. Rev. D* **103** 064048
- [44] Mathews J and Pound A 2025 Post-adiabatic waveform-generation framework for asymmetric precessing binaries (arXiv:2501.01413)
- [45] Hinderer T and Flanagan E E 2008 Two timescale analysis of extreme mass ratio inspirals in Kerr. I. Orbital Motion *Phys. Rev. D* **78** 064028
- [46] Stein L C and Warburton N 2020 Location of the last stable orbit in Kerr spacetime *Phys. Rev. D* **101** 064007

- [47] Ori A and Thorne K S 2000 The Transition from inspiral to plunge for a compact body in a circular equatorial orbit around a massive, spinning black hole *Phys. Rev. D* **62** 124022
- [48] Compère G and Küchler L 2021 Self-consistent adiabatic inspiral and transition motion *Phys. Rev. Lett.* **126** 241106
- [49] Compère G and Küchler L 2022 Erratum: self-consistent adiabatic inspiral and transition motion [phys. rev. lett. 126, 241106 (2021)] *Phys. Rev. Lett.* **128** 029901
- [50] Compère G and Küchler L 2022 Asymptotically matched quasi-circular inspiral and transition-to-plunge in the small mass ratio expansion *SciPost Phys.* **13** 043
- [51] Küchler L, Compère G, Durkan L and Pound A 2024 Self-force framework for transition-to-plunge waveforms *SciPost Phys.* **17** 056
- [52] Sundararajan P A, Khanna G and Hughes S A 2007 Towards adiabatic waveforms for inspiral into Kerr black holes. I. A New model of the source for the time domain perturbation equation *Phys. Rev. D* **76** 104005
- [53] Sundararajan P A, Khanna G, Hughes S A and Drasco S 2008 Towards adiabatic waveforms for inspiral into Kerr black holes: II. Dynamical sources and generic orbits *Phys. Rev. D* **78** 024022
- [54] Sundararajan P A, Khanna G and Hughes S A 2010 Binary black hole merger gravitational waves and recoil in the large mass ratio limit *Phys. Rev. D* **81** 104009
- [55] Rifat N E M, Field S E, Khanna G and Varma V 2020 Surrogate model for gravitational wave signals from comparable and large-mass-ratio black hole binaries *Phys. Rev. D* **101** 081502
- [56] Islam T, Field S E, Hughes S A, Khanna G, Varma V, Giesler M, Scheel M A, Kidder L E and Pfeiffer H P 2022 Surrogate model for gravitational wave signals from nonspinning, comparable-to large-mass-ratio black hole binaries built on black hole perturbation theory waveforms calibrated to numerical relativity *Phys. Rev. D* **106** 104025
- [57] Hadar S and Kol B 2011 Post-ISCO ringdown amplitudes in extreme mass ratio inspiral *Phys. Rev. D* **84** 044019
- [58] Folacci A and Ould El Hadj M 2018 Multipolar gravitational waveforms and ringdowns generated during the plunge from the innermost stable circular orbit into a Schwarzschild black hole *Phys. Rev. D* **98** 084008
- [59] Cheung M H-Y, Berti E, Baibhav V and Cotesta R 2024 Extracting linear and nonlinear quasinormal modes from black hole merger simulations *Phys. Rev. D* **109** 044069
- Cheung M H-Y, Berti E, Baibhav V and Cotesta R 2024 *Phys. Rev. D* **110** 049902 (Erratum)
- [60] Lim H, Hughes S A and Khanna G 2022 Measuring quasinormal mode amplitudes with misaligned binary black hole ringdowns *Phys. Rev. D* **105** 124030
- [61] Carullo G 2024 Ringdown amplitudes of nonspinning eccentric binaries *J. Cosmol. Astropart. Phys.* **JCAP10(2024)061**
- [62] Mitman K et al 2025 Probing the ringdown perturbation in binary black hole coalescences with an improved quasi-normal mode extraction algorithm (arXiv:2503.09678)
- [63] Pound A 2012 Second-order gravitational self-force *Phys. Rev. Lett.* **109** 051101
- [64] Pound A 2012 Nonlinear gravitational self-force. I. Field outside a small body *Phys. Rev. D* **86** 084019
- [65] Pound A 2017 Nonlinear gravitational self-force: second-order equation of motion *Phys. Rev. D* **95** 104056
- [66] Upton S D and Pound A 2021 Second-order gravitational self-force in a highly regular gauge *Phys. Rev. D* **103** 124016
- [67] Detweiler S 2012 Gravitational radiation reaction and second order perturbation theory *Phys. Rev. D* **85** 044048
- [68] Pound A and Miller J 2014 Practical, covariant puncture for second-order self-force calculations *Phys. Rev. D* **89** 104020
- [69] Lewis J, Kakehi T, Pound A and Tanaka T 2025 Post-adiabatic dynamics and waveform generation in self-force theory: an invariant pseudo-Hamiltonian framework (arXiv:2507.08081)
- [70] Miller J, Leather B, Pound A and Warburton N 2024 Worldtube puncture scheme for first- and second-order self-force calculations in the Fourier domain *Phys. Rev. D* **109** 104010
- [71] Pound A 2015 Second-order perturbation theory: problems on large scales *Phys. Rev. D* **92** 104047
- [72] Cunningham K, Kavanagh C, Pound A, Trestini D, Warburton N and Neef J 2024 Gravitational memory: new results from post-Newtonian and self-force theory (arXiv:2410.23950)
- [73] Misner C W, Thorne K S and Wheeler J A 1973 *Gravitation* (W.H. Freeman & Co Ltd)
- [74] Pound A 2010 Self-consistent gravitational self-force *Phys. Rev. D* **81** 024023
- [75] Küchler L 2023 Inspiral, transition and plunge: a framework for complete waveforms in the small-mass-ratio expansion *PhD Thesis* U. Brussels
- [76] Gralla S E and Wald R M 2008 A rigorous derivation of gravitational self-force *Class. Quantum Grav.* **25** 205009
- Gralla S E and Wal R M 2011 *Class. Quantum Grav.* **28** 159501 (Erratum)
- [77] Pound A 2015 Gauge and motion in perturbation theory *Phys. Rev. D* **92** 044021
- [78] Warburton N, Barack L and Sago N 2013 Isosynchronous pairing of geodesic orbits in Kerr geometry *Phys. Rev. D* **87** 084012
- [79] Folacci A and Ould El Hadj M 2018 Electromagnetic radiation generated by a charged particle plunging into a Schwarzschild black hole: Multipolar waveforms and ringdowns *Phys. Rev. D* **98** 024021
- [80] Rom B and Sari R 2022 Extreme mass-ratio binary black hole merger: characteristics of the test-particle limit *Phys. Rev. D* **106** 104040
- [81] Strusberg I, Rom B and Sari R 2025 Universal waveforms for extreme mass-ratio inspiral (arXiv:2505.07941)
- [82] Chandrasekhar S 1983 *The Mathematical Theory of Black Holes (The International Series of Monographs on Physics)* (Oxford University Press)
- [83] Martel K and Poisson E 2005 Gravitational perturbations of the Schwarzschild spacetime: a practical covariant and gauge-invariant formalism *Phys. Rev. D* **71** 104003
- [84] Spiers A, Pound A and Wardell B 2024 Second-order perturbations of the Schwarzschild spacetime: practical, covariant and gauge-invariant formalisms *Phys. Rev. D* **110** 064030
- [85] Regge T and Wheeler J A 1957 Stability of a schwarzschild singularity *Phys. Rev.* **108** 1063–9
- [86] Zerilli F J 1970 Gravitational field of a particle falling in a schwarzschild geometry analyzed in tensor harmonics *Phys. Rev. D* **2** 2141–60
- [87] Hopper S and Evans C R 2010 Gravitational perturbations and metric reconstruction: Method of extended homogeneous solutions applied to eccentric orbits on a Schwarzschild black hole *Phys. Rev. D* **82** 084010
- [88] Cunningham C T, Price R H and Moncrief V 1978 Radiation from collapsing relativistic stars. I - Linearized odd-parity radiation *Astrophys. J.* **224** 643
- [89] Moncrief V 1974 Gravitational perturbations of spherically symmetric systems. I. The exterior problem *Ann. Phys.* **88** 323–42
- [90] Cunningham C T, Price R H and Moncrief V 1979 Radiation from collapsing relativistic stars. II. Linearized even-parity radiation *Astrophys. J.* **230** 870–92

- [91] Chandrasekhar S and Detweiler S 1975 The quasi-normal modes of the schwarzschild black hole *Proc. R. Soc. A* **344** 441–52
- [92] Berti E, Cardoso V and Starinets A O 2009 Quasinormal modes of black holes and black branes *Class. Quantum Grav.* **26** 163001
- [93] Berti E and Cardoso V 2006 Quasinormal ringing of Kerr black holes. I. The Excitation factors *Phys. Rev. D* **74** 104020
- [94] Zhang Z, Berti E and Cardoso V 2013 Quasinormal ringing of Kerr black holes. II. Excitation by particles falling radially with arbitrary energy *Phys. Rev. D* **88** 044018
- [95] Oshita N 2021 Ease of excitation of black hole ringing: Quantifying the importance of overtones by the excitation factors *Phys. Rev. D* **104** 124032
- [96] Leaver E W 1986 Spectral decomposition of the perturbation response of the Schwarzschild geometry *Phys. Rev. D* **34** 384–408
- [97] Küchler L, Compere G and Pound A 2025 Self-force framework for merger-ringdown waveforms (ancillary files) (available at: <https://github.com/gcompere/Self-force-framework-for-merger-ringdown-waveforms>)
- [98] Black hole perturbation toolkit (available at: [bhptoolkit.org](https://github.com/bhptoolkit))
- [99] Vega I, Diener P, Tichy W and Detweiler S L 2009 Self-force with (3+1) codes: a primer for numerical relativists *Phys. Rev. D* **80** 084021
- [100] Berti E Ringdown (available at: <https://pages.jh.edu/eberti2/ringdown/>)
- [101] Berti E, Cardoso V and Will C M 2006 On gravitational-wave spectroscopy of massive black holes with the space interferometer LISA *Phys. Rev. D* **73** 064030
- [102] Leaver E W 1985 An Analytic representation for the quasi normal modes of Kerr black holes *Proc. R. Soc. A* **402** 285–98
- [103] SXS gravitational waveform database (available at: <https://data.black-holes.org/waveforms/index.html>)
- [104] Boyle M et al 2019 The SXS Collaboration catalog of binary black hole simulations *Class. Quantum Grav.* **36** 195006
- [105] Boyle M and Scheel M The sxs package (available at: <https://doi.org/10.5281/zenodo.4034006>)
- [106] Le Tiec A, Mroue A H, Barack L, Buonanno A, Pfeiffer H P, Sago N and Taracchini A 2011 Periastron advance in black hole binaries *Phys. Rev. Lett.* **107** 141101
- [107] Le Tiec A, Barausse E and Buonanno A 2012 Gravitational self-force correction to the binding energy of compact binary systems *Phys. Rev. Lett.* **108** 131103
- [108] Le Tiec A et al 2013 Periastron advance in spinning black hole binaries: gravitational self-force from numerical relativity *Phys. Rev. D* **88** 124027
- [109] Nagar A 2013 Gravitational recoil in nonspinning black hole binaries: the span of test-mass results *Phys. Rev. D* **88** 121501
- [110] Le Tiec A and Grandclément P 2018 Horizon surface gravity in corotating black hole binaries *Class. Quantum Grav.* **35** 144002
- [111] van de Meent M and Pfeiffer H P 2020 Intermediate mass-ratio black hole binaries: applicability of small mass-ratio perturbation theory *Phys. Rev. Lett.* **125** 181101
- [112] Blackman J (SXS Collaboration) 2024 Binary black-hole simulation SXS:BBH:1220 [Data set] (Zenodo) <https://doi.org/10.5281/zenodo.13159569>
- [113] SXS Collaboration 2024 Binary black-hole simulation SXS:BBH:2477 [Data set] (Zenodo) <https://doi.org/10.5281/zenodo.13147606>
- [114] SXS Collaboration 2019 Binary black-hole simulation SXS:BBH:1931 [Data set] (Zenodo) <https://doi.org/10.5281/zenodo.3305100>
- [115] SXS Collaboration 2024 Binary black-hole simulation SXS:BBH:2471 [Data set] (Zenodo) <https://doi.org/10.5281/zenodo.13147573>
- [116] Giesler M, Isi M, Scheel M A and Teukolsky S 2019 Black hole ringdown: the importance of overtones *Phys. Rev. X* **9** 041060
- [117] Bhagwat S, Forteza X J, Pani P and Ferrari V 2020 Ringdown overtones, black hole spectroscopy and no-hair theorem tests *Phys. Rev. D* **101** 044033
- [118] Okounkova M 2020 Revisiting non-linearity in binary black hole mergers (arXiv:2004.00671)
- [119] Cook G B 2020 Aspects of multimode Kerr ringdown fitting *Phys. Rev. D* **102** 024027
- [120] Dhani A 2021 Importance of mirror modes in binary black hole ringdown waveform *Phys. Rev. D* **103** 104048
- [121] Mourier P, Jiménez Forteza X, Pook-Kolb D, Krishnan B and Schnetter E 2021 Quasinormal modes and their overtones at the common horizon in a binary black hole merger *Phys. Rev. D* **103** 044054
- [122] Finch E and Moore C J 2021 Modeling the ringdown from precessing black hole binaries *Phys. Rev. D* **103** 084048
- [123] Forteza X J and Mourier P 2021 High-overtone fits to numerical relativity ringdowns: Beyond the dismissed  $n = 8$  special tone *Phys. Rev. D* **104** 124072
- [124] Baibhav V, Cheung M H-Y, Berti E, Cardoso V, Carullo G, Cotesta R, Pozzo W D and Duque F 2023 Agnostic black hole spectroscopy: quasinormal mode content of numerical relativity waveforms and limits of validity of linear perturbation theory *Phys. Rev. D* **108** 104020
- [125] Nee P J, Völkel S H and Pfeiffer H P 2023 Role of black hole quasinormal mode overtones for ringdown analysis *Phys. Rev. D* **108** 044032
- [126] Zhu H, Ripley J L, Cárdenas-Avendaño A and Pretorius F 2024 Challenges in quasinormal mode extraction: perspectives from numerical solutions to the Teukolsky equation *Phys. Rev. D* **109** 044010
- [127] Qiu Y, Forteza X J and Mourier P 2024 Linear versus nonlinear modeling of black hole ringdowns *Phys. Rev. D* **109** 064075
- [128] Giesler M et al 2024 Overtones and nonlinearities in binary black hole ringdowns (arXiv:2411.11269)
- [129] Buonanno A, Cook G B and Pretorius F 2007 Inspiral, merger and ring-down of equal-mass black-hole binaries *Phys. Rev. D* **75** 124018
- [130] Oshita N and Cardoso V 2025 Reconstruction of ringdown with excitation factors *Phys. Rev. D* **111** 104043
- [131] De Amicis M, Albanesi S and Carullo G 2024 Inspiral-inherited ringdown tails *Phys. Rev. D* **110** 104005
- [132] Fernando M, Neilsen D, Lim H, Hirschmann E and Sundar H 2019 Massively parallel simulations of binary black hole intermediate-mass-ratio inspirals *SIAM J. Sci. Comput.* **41** C97–C138
- [133] Lousto C O and Healy J 2020 Exploring the small mass ratio binary black hole merger via zeno’s dichotomy approach *Phys. Rev. Lett.* **125** 191102
- [134] Dhisi M, Rüter H R, Pound A, Barack L and Pfeiffer H P 2021 Worldtube excision method for intermediate-mass-ratio inspirals: scalar-field toy model *Phys. Rev. D* **104** 124002
- [135] Albertini A, Nagar A, Pound A, Warburton N, Wardell B, Durkan L and Miller J 2022 Comparing second-order gravitational self-force and effective one body waveforms from inspiralling, quasicircular and nonspinning black hole binaries. II. The large-mass-ratio case *Phys. Rev. D* **106** 084062
- [136] Wittek N A, Barack L, Pfeiffer H P, Pound A, Deppe N, Kidder L E, Macedo A, Nelli K C, Throwe W and Vu N L 2025 Relieving scale disparity in binary black hole simulations *Phys. Rev. Lett.* **134** 251402

- [137] Küchler L 2024 Progress towards merger-ringdown waveforms from self-force theory *Talk Given at 27th Capra Meeting on Radiation Reaction in General Relativity* (National University of Singapore) (available at: [www.caprameeting.org/capra-meetings/capra-27/abstracts](http://www.caprameeting.org/capra-meetings/capra-27/abstracts))
- [138] Küchler L 2024 Self-force framework for transition-to-plunge waveforms *Talk Given at 15th Int. LISA Symp.* (University College Dublin) (available at: [www.lisasymposium2024ie/programme/](http://www.lisasymposium2024ie/programme/))
- [139] Küchler L 2024 Self-force framework for transition-to-plunge waveforms *Talk Given at Fundamental Physics Meets Waveforms With LISA Workshop* (Max Planck Institute for Gravitational Physics) (available at: <https://workshops.aei.mpg.de/fpmeetswavelisa/program/>)
- [140] Mummery A and Balbus S 2022 Inspirals from the innermost stable circular orbit of kerr black holes: exact solutions and universal radial flow *Phys. Rev. Lett.* **129** 161101
- [141] Dyson C and van de Meent M 2023 Kerr-fully diving into the abyss: analytic solutions to plunging geodesics in Kerr *Class. Quantum Grav.* **40** 195026
- [142] Teukolsky S A 1972 Rotating black holes - separable wave equations for gravitational and electromagnetic perturbations *Phys. Rev. Lett.* **29** 1114–8
- [143] Honet L, Küchler L, Pound A and Compère G 2025 Transition-to-plunge self-force waveforms with a spinning primary (arXiv:2510.13958)
- [144] Lhost G and Compère G 2025 Approach to the separatrix with eccentric orbits *SciPost Phys. Core* **8** 059
- [145] Faggioli G, van de Meent M, Buonanno A and Khanna G 2025 Characterizing the merger of equatorial-eccentric-geodesic plunges in rotating black holes *Phys. Rev. D* **112** 084009
- [146] Albanesi S, Nagar A and Bernuzzi S 2021 Effective one-body model for extreme-mass-ratio spinning binaries on eccentric equatorial orbits: testing radiation reaction and waveform *Phys. Rev. D* **104** 024067
- [147] Albanesi S, Bernuzzi S, Damour T, Nagar A and Placidi A 2023 Faithful effective-one-body waveform of small-mass-ratio coalescing black hole binaries: the eccentric, nonspinning case *Phys. Rev. D* **108** 084037
- [148] Faggioli G, van de Meent M, Buonanno A, Gamboa A, Khalil M and Khanna G 2025 Testing eccentric corrections to the radiation-reaction force in the test-mass limit of effective-one-body models *Phys. Rev. D* **111** 044036
- [149] Becker D R and Hughes S A 2025 Transition from adiabatic inspiral to plunge for eccentric binaries *Phys. Rev. D* **111** 064003
- [150] Spiers A, Maselli A and Sotiriou T P 2024 Measuring scalar charge with compact binaries: high accuracy modeling with self-force *Phys. Rev. D* **109** 064022
- [151] Speri L, Barsanti S, Maselli A, Sotiriou T P, Warburton N, van de Meent M, Chua A J K, Burke O and Gair J 2024 Probing fundamental physics with Extreme Mass Ratio Inspirals: a full Bayesian inference for scalar charge (arXiv:2406.07607)
- [152] Dyson C, Spiexsma T F M, Brito R, van de Meent M and Dolan S 2025 Environmental effects in extreme mass ratio inspirals: perturbations to the environment in Kerr (arXiv:2501.09806)
- [153] Roy A, Küchler L, Pound A and Panosso Macedo R 2025 Black hole mergers beyond general relativity: a self-force approach (arXiv:2510.11793)
- [154] Compère G, Chakraborty S and Machet L Plunging in black hole surrounded by a dark matter halo (in preparation)