Binary Black Hole Mergers: Conservation Laws and Gravitational Wave Memory Effects

Arwa Elhashash

Alexandria, Egypt

B.S. Physics, University of Science and Technology at Zewail City, 2017

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the degree of
Doctor of Philosophy

Department of Physics
University of Virginia
April 2025

Professor David A. Nichols
Research Advisor

Professor Kent Yagi

Professor Peter Arnold

Professor Phil Arras

Abstract

Studying binary-black-hole (BBH) systems has long been of great interest in general relativity, especially after the first observed event in 2015 discovered by LIGO, which has revolutionized our understanding of the universe and opened up new insights that were not accessible to electromagnetic astronomy. These detections offer an insightful way to test general relativity in the strong-field and high-luminosity regime. In this dissertation, we cover two ways of characterizing these systems that turn out to be connected; the symmetries of these asymptotic systems and a relativistic strong-field phenomenon associated with them.

We begin by studying BBHs as isolated systems in an empty space that approaches flatness at infinity. Changes in the spacetime can then be quantified using the changes in charges conjugate to the asymptotic symmetries of the spacetime. The symmetries of asymptotically flat spacetimes in general relativity are described by the Bondi-Metzner-Sachs group (or its proposed extensions). Associated with these symmetries are conserved charges, which include the energy-momentum, supermomentum, and relativistic angular momentum (or super-angular momentum). In flat spacetime, there is an agreed upon well defined notion of angular momentum. However, in the non-linear theory of gravity in the presence of gravitational waves several formalisms have been used to compute the spacetime angular momentum. These angular momenta do not always agree, but the different definitions have been summarized in

a two-parameter family of angular momenta. We found that a reasonable physical requirement for the angular momentum to vanish in flat spacetime restricts the two parameters to be equal, solving a part of the discrepancy that appears in the angular momentum definition. We examined the effect of this free parameter on the values of the angular momentum and super-angular momentum of nonprecessing binary-black-hole mergers. We found the definitions of angular momentum differ only when these systems are radiating gravitational waves (GWs). The definitions of super-angular momentum differ even after the GWs pass, because of a lasting effect called the GW memory effect. Using numerical-relativity surrogate waveforms, we estimate these differences to be small, but of the order of the accuracy of the angular momentum computed from these simulations.

A significant part of this thesis focuses on the relativistic non-linear GW memory effect and its detection prospects. This effect causes a permanent relative displacement between two freely falling test masses that persists after the passage of a GW signal, leaving a "memory" of the event. The memory effect has been computed first in the 1970's, but only with upcoming improvements to the LIGO, Virgo, and KAGRA detectors will the prospects of detecting the effect in a population of BBH mergers be promising. Searches for the memory effect in GW detector data require accurate waveform models, which must be evaluated many times (and, thus, need to be evaluated rapidly). Current analytical waveform models and many numerical-relativity waveforms and surrogates of BBH mergers do not include the memory effect. Instead, GW

memory is computed from waveforms without memory by using conservation laws in asymptotically flat spacetimes, which is relatively slow. We present the first time- and frequency-domain waveform models of the GW memory effect for nonspinning BBH mergers for comparable-mass systems that can be evaluated more rapidly. A part of the model involves computing a fit for the final memory offset that incorporates data from both comparable and extreme mass-ratio limits, and which could be applied in both contexts to understand the remnant properties of BBH mergers more fully. In addition to speeding up GW searches, having these analytic models give analytical insights into the time- and frequency-domain properties of the GW memory signal.

To my mom

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisor, David Nichols. I have learned an immense amount from him—both in physics and in life. His work on gravitational-wave memory effects drew me to a research topic that has continued to captivate me ever since. Our many discussions, ranging from black hole spin to why California water is cold, have enriched my academic journey in ways I will always cherish. His guidance, especially in presenting my research clearly and effectively, has been invaluable.

I am also grateful to the members of my defense committee—Prof. Peter Arnold, Prof. Kent Yagi, and Prof. Phil Arras—for their insightful questions and thoughtful suggestions. Their feedback helped sharpen my understanding and improve the quality of my work. I would also like to thank all my professors at UVA for their teaching and support—especially Prof. Diana Vaman, whose courses on quantum field theory I deeply enjoyed and which greatly influenced my appreciation for theoretical physics. I would also like to thank Niels Warburton for insightful discussions about extreme-mass-ratio inspirals and for his guidance on writing research proposals. My sincere thanks to Alex Grant for his insightful discussions and valuable input throughout my research, and to Alex Saffer for setting the bar too high with his presentations. I would like to extend my gratitude to Prof. Tamer Elkholy, a mentor whose steadfast support and guidance have been a cornerstone of my academic journey, and to Prof.

Ali Nassar, whose encouragement during my undergraduate years helped set me on this path.

A special thanks to Sid Ajith and Ben Wade for making the office an engaging and lively place through our endless discussions, and to my research group members—Nan, Shammi, Vincent, Sayantani, Siddhant, Victor, Nur, and Cuishan—for their thoughtful feedback and shared enthusiasm for research. My heartfelt appreciation goes to the support staff of the physics department for their dedication and help throughout the years—especially Peter Cline, whose kindness and reliability were a constant source of reassurance. I would also like to thank Prof. Max Bychkov for his mentorship and invaluable guidance in teaching. I am also thankful for the warm community I found in Charlottesville, especially Zahraa, Amr, and Heidi, for their friendship and support.

To my mother, whose unwavering support and encouragement sustained me when I chose to change my academic path to physics—thank you. Your strength and resilience have been a source of constant inspiration, and without you, I would not be here today. To my sisters, Alaa and Rwan—thank you. Alaa, your adventurous and ambitious spirit continues to inspire me. Rwan, your friendship and presence (even from afar) has meant the world. And to my little brother Mohamed, thank you for teaching me how to play video games and helping me unwind during stressful times. To my best friends, Mayar Shahin and Sara Khaled—thank you for being constant sources of support despite the distance. Our yearly debriefing trips were a reminder of the importance of friendship and balance.

Finally, and most importantly, I thank my best friend, then lover, and now husband, Omar. Your love, belief in me, and unwavering support from the very beginning have carried me through this journey. Thank you for helping me stay true to myself, especially during the hardest moments. I could not have done this without you.

List of Abbreviations

GR	General relativity
GW	Gravitational wave
BH	Black hole
EMRI	Extreme mass-ratio inspiral
BBH	Binary black hole
NR	Numerical Relativity
EOB	Effective-One-Body
PN	Post-Newtonian
QNM	Quasi-normal mode
ISCO	Innermost stable circular orbit
IMR	Inspiral-merger-ringdown
SNR	Signal-to-noise-ratio
BMS	Bondi-Metzner-Sachs
CM	Center-of-mass
STF	Symmetric-trace-free
LIGO	Laser Interferometer Gravitational-Wave Observatory
LISA	Laser Interferometer Space Antenna
NANOGrav	North American Nanohertz Observatory for Gravitational Waves
KAGRA	Kamioka Gravitational Wave Detector
EPTA	European Pulsar Timing Array
InPTA	International Pulsar Timing Array
LVK	LIGO-Virgo-KAGRA collaboration
FFT	Fast Fourier transform

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The study of binary-black-hole (BBH) mergers has always been of great interest, especially after the first detection of gravitational waves (GWs) made by LIGO from a BBH merger. The strong gravity associated with these systems provide the perfect environment to test Einstein's theory of general relativity (GR) in that strong-filed regime. These systems can be characterized by their masses and spins of the individual black holes, or by the mass and spin of the final forms black hole. Another way of characterizing these systems is by studying the symmetries of the spacetime in which they exist. The spacetime of these isolated systems is asymptotically flat, meaning that the spacetime become flat in some appropriate infinite limit. In this thesis, we focus on two aspects of studying these systems: the symmetries of the asymptotically flat spacetimes (and their corresponding conserved charges) and a relativistic phenomenon that arises from the strong gravitational field associated with them called the GW memory effect.

1.1 Gravitational Waves

General relativity (GR), formulated by Albert Einstein in 1915, has been successful as classical theory describing gravity, after it replaced Newtonian gravity with a geometric theory that describes gravity as the curvature of spacetime caused by stress-energy. The field equations, known as Einstein field equations, relate the curvature of spacetime (represented by the Einstein tensor $G_{\mu\nu}$, which depends on the spacetime metric $g_{\mu\nu}$) to the energy-momentum content (represented by the stress-energy tensor $T_{\mu\nu}$) through

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \,, \tag{1.1}$$

where μ is an index that runs over the four spacetime coordinates. Einstein's theory of GR has explained and predicted many physical phenomena, such as the anomalous perihelion shift in Mercury's orbit, the bending of light by gravity, black holes and event horizons, and gravitational waves (GWs). In 1918, Einstein has formulated the mathematical description of GWs in the weak-filed approximation, finding that metric perturbations could propagate as ripples in spacetime at the speed of light. His work introduced the quadrupole formula, predicting how accelerating masses could generate gravitational radiation.

There was some skepticism whether GWs were physical or just coordinate artifacts, since the metric can be transformed under coordinate changes. This skepticism, famously depicted in Arthur Eddington's comment on GWs in 1922 as "propagating at the speed of thought", was later followed by Feynman's "sticky bead argument" that states the following. Consider a rigid rod with two beads that can slide along it, and are initially at rest. When a gravitational wave perturbs spacetime, non-gravitational forces will keep the rod rigid, while the beads slide back and forth along the rod. If there is friction between the rod and the beads, then the motion of the beads will generate heat through that friction. Since heat is a form of energy, this means that gravitational waves has transferred energy to the system. Through this thought experiment, many physicists were convinced that GWs were indeed real physical phenomenon and not mere coordinate artifacts of GR.

In the weak-field approximation, when the spacetime is nearly flat, Einstein's equations can be linearized. In that approximation, GWs can be understood as perturbations on the Minkowski (flat) spacetime metric,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{1.2}$$

where $\eta_{\mu\nu}$ is the background flat spacetime metric and $h_{\mu\nu} \ll 1$ is a small perturbation (representing GWs). In the vacuum case ($T_{\mu\nu} = 0$), and using the Lorentz gauge ($\partial^\mu \bar{h}_{\mu\nu} = 0$, where $\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h_\alpha{}^\alpha$), Einstein's equations yield

$$\Box \bar{h}_{\mu\nu} = 0\,, \tag{1.3}$$

where $\Box \equiv \eta^{\mu\nu}\partial_\mu\partial_\nu$. The general solution to the wave equation takes the form

$$\bar{h}_{\mu\nu} = A_{\mu\nu} e^{ik_\alpha x^\alpha}\,, \tag{1.4}$$

where $A_{\mu\nu}$ is a complex tensor representing the wave amplitude (independent of time) and k_α is a null vector. This wave-like solution indicates that GWs propagate at the speed of light along null rays.

In 2015, the first direct detection of GWs was made by the Laser Interferometer Gravitational-Wave Observatory (LIGO) collaboration from the merger of two black holes (GW150914 [1]), confirming Einstein's predictions. Since the first detection, LIGO has detected GWs from many different sources, such as binary black hole (BBH) mergers, binary neutron star mergers, and neutron star–black hole mergers. So far, nearly 100 GWs detections of merging compact objects have been made by LIGO, Virgo, and KAGRA, during the first three observing runs (and many more detection candidates during the ongoing fourth run have been announced). The GWs detections have opened a new window into the universe, complementing traditional electromagnetic and astro-particle-based observations. The majority of these detections are of the merger of black hole binaries. These detections provide direct tests of GR in the strong-field and high-luminosity regime and offer insights into the population of BBH mergers and properties of BHs including the distributions of masses and spins.

1.2 The infrared triangle

The infrared triangle (shown in Fig. 1.1) connects three seemingly unrelated subjects in physics that were discovered to be the same subject approached from different starting points in literature (see [2] for a review). The first corner is soft theorems, first developed in QED and later extended to gravity by Weinberg, describe how scattering

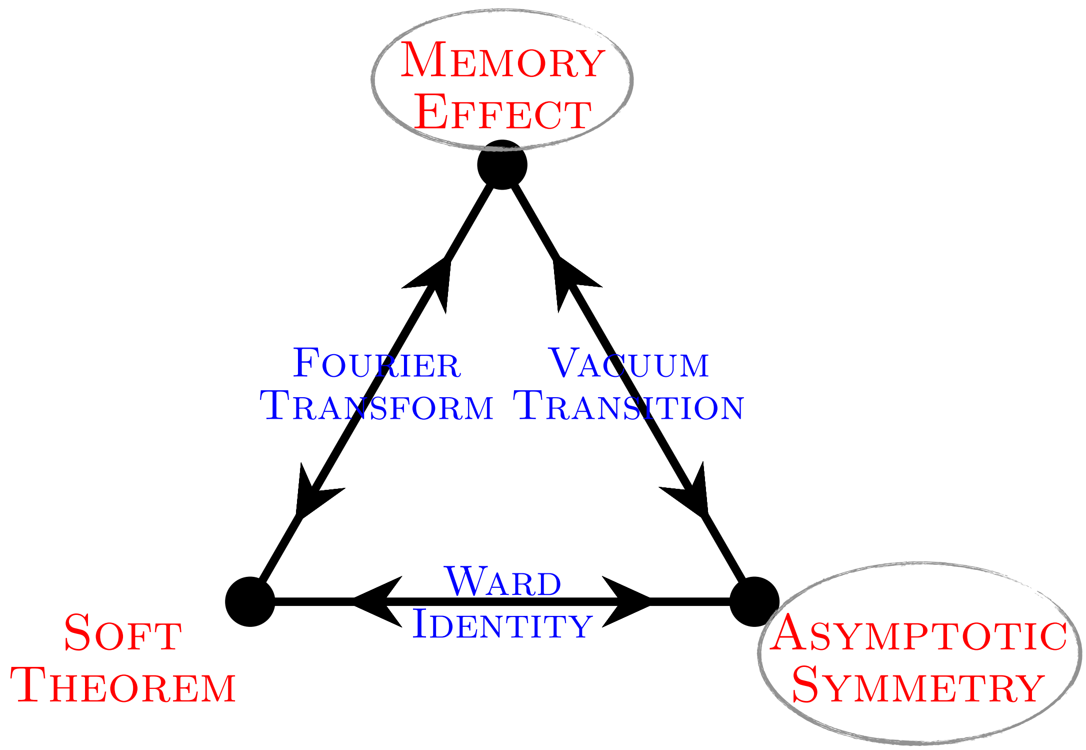

Figure 1.1: **The infrared triangle**: We focus on two corners of this triangle; the symmetries of asymptotically flat spacetimes and the memory effects. Credit: A. Strominger [2].

amplitudes behave when massless particles (e.g., photons or gravitons) become soft (their energy approach zero). This connection to gravity is important in constructing a quantum theory of gravity. The second corner is asymptotic symmetries, which studies the symmetries and the corresponding conserved chargers of any system with an asymptotic region or boundary. The third corner is the memory effect, a subtle zero-frequency (DC) effect in which the passage of gravitational waves leaves a permanent shift in the relative positions of two inertial detectors. This connection of the memory effect to classical and quantum gravitational scattering has generated interest in detecting the memory effect. There are also interesting relations between these three topics, the soft theorem is connected to memory through a Fourier transform of the momentum space poles in scattering amplitudes, which gives a step function in time (memory). That step reflects a change between two vacua, linked by asymptotic symmetries. Each symmetry has a Ward identity, which turn out to be another way of expressing the soft theorems. The two corners of the triangle we are focusing on in this thesis are the asymptotic symmetries and memory effects. We study both of these topics in the context of BBH mergers.

1.3 Asymptotically Flat Spacetimes

One way of studying black-hole binaries, and other isolated gravitational systems in GR, is through the theoretical framework of asymptotically flat spacetimes. Asymptotically flat spacetimes are characterized by the property that at large distances from a localized source, the geometry approaches that of Minkowski space. This set up

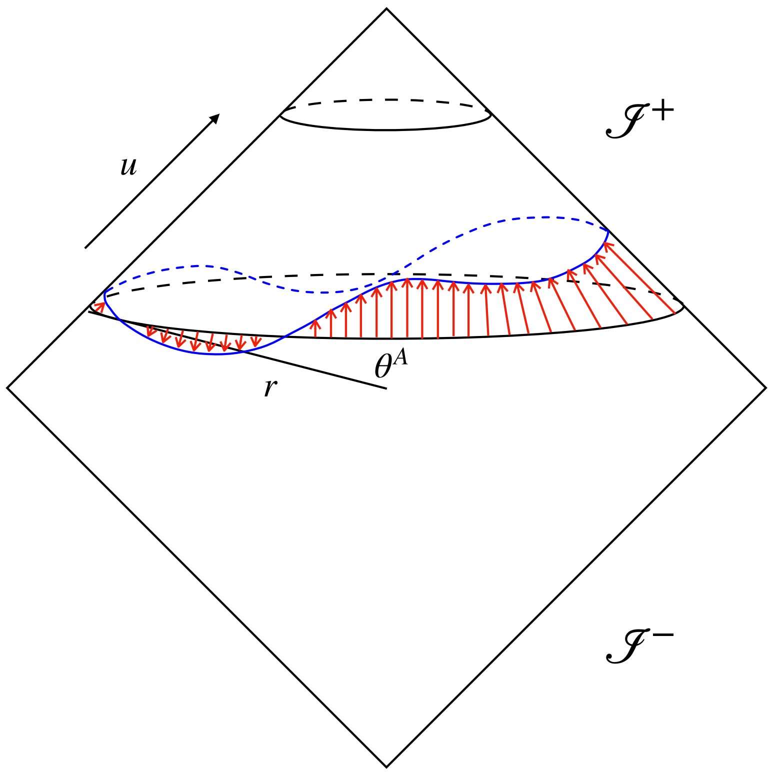

Figure 1.2: **Supertranslations**: A picture of a conformally compactified spacetime where $\mathscr{I}^+$ and $\mathscr{I}^-$ denote future and past null infinity. The lower black circle represents a 2-sphere at fixed radius r and retarded time $u = t - r$ on $\mathscr{I}^+$, parametrized by the two angles θ^A. Performing supertranslations along the direction of u translates each point on the 2-sphere independently, giving the blue 2-sphere. The supertranslations are indicated by the red arrows. In comparison, under ordinary translation (Poincaré) each point on the 2-sphere gets translated by the same amount, giving the 2-sphere shown in black at the top of the diagram.

provides a well-defined notion of gravitational radiation, asymptotic symmetries, and conserved quantities. The work of studying the asymptotic structure of spacetime in GR started in the early 1950s through independent studies by Hermann Bondi and colleagues, and Metzner and Sachs [3–5].

One might have thought that the symmetry group of flat spacetime (Poincaré group) should be recovered at the asymptotic regions where the spacetime is almost flat, so that spacetime has the symmetries of special relativity in that limit of large distances and weak fields. Instead, they found that the usual Poincaré symmetries of Minkowski space were not the only symmetries at null infinity, i.e., the null boundary of an asymptotically flat spacetime in the covariant conformal approach of Penrose [6, 7]. A larger set of symmetries emerged, which included angle-dependent translations known as supertranslations. This enhanced symmetry group, known as the Bondi-Metzner-Sachs (BMS) group, is the semi-direct product of the infinite-dimensional abelian group of supertranslations (of which the four spacetime translations are a subgroup) with the six-dimensional non-abelian Lorentz group. Supertranslations can be thought of as angle-dependent translations, that translate each point on the asymptotic 2-sphere independently. This is illustrated in Fig. 1.2. Beyond the BMS group, there have been two proposed extensions of the symmetry group or algebra of asymptotically flat spacetimes:

1. The first is the extended BMS algebra by Barnich and Troessaert [8–10], which includes all conformal Killing vectors of the 2-sphere, rather than the globally

defined vectors (isomorphic to the Lorentz group). These additional symmetry vectors were dubbed "super-rotations", of which the vectors that are isomorphic to Lorentz transformation are a subalgebra. Analogous to supertranslations, the super-rotations and are a kind of asymptotic angle-dependent rotations and Lorentz boosts. The supertranslations have to be correspondingly modified to include functions that are not necessarily smooth, in order to keep the algebra structure.

2. The second proposed extension is the generalized BMS algebra by Campiglia and Laddha [11, 12], which considers all diffeomorphisms of the 2-sphere rather than those equivalent to Lorentz transformations. The supertranslations are left unmodified and are the same as in the original BMS group.

According to Noether's theorem [13], associated with each symmetry there is a conserved charge. In Minkowski (flat) spacetime, the conserved charges associated with space-time translations and Lorentz transformations are the relativistic 4-momentum (which includes the radiated energy and linear momentum) and angular momentum, respectively. In asymptotically flat spacetimes, however, there are different prescriptions to define the charges and their corresponding fluxes associated with the BMS symmetries. Following the prescription proposed by Wald and Zoupas [14], in which they compute the charges and their fluxes using a generalization of Noether's theorem, the charge conjugate to supertranslations was dubbed "supermomentum" (which includes 4-momentum) and that conjugate to super-rotations was called "super angular momentum". We investigate these charges along with a subtlety that exists in the definition of the angular momentum charge in Chapter 2.

1.4 Black Hole Binaries

Binary black hole (BBH) mergers are good astrophysical systems for studying gravity in the strong-field and high-luminosity regime. The evolution of the BBH system can be divided into three major phases, each characterized by distinct physical processes and GW signatures. The first is the inspiral phase, during which the BHs orbit each other at increasing speed, gradually losing energy and angular momentum as the separation decreases. The second is the late-inspiral and merger phase, where the BHs spiral closer until they dynamically plunge, forming a single remnant BH that is highly perturbed. This is the most dynamical and non-linear phase with the strongest GW emission and highest frequency. The last phase is the ringdown, where the final perturbed BH is formed and undergoes damped oscillations (known as quasinormal modes) as it settles into a Kerr (spinning) BH, as a result of converting the angular momentum of the binary into spin angular momentum of the remnant BH.

There are different schemes of studying black hole binaries, some are analytical approximation schemes and others are numerical techniques. The different schemes could be used depending on the mass ratio $q = m_1/m_2 \geq 1$ (where m_1 is the primary BH mass and m_2 is the secondary) as well as the binary separation r_{12} (or relative velocity). The different schemes of studying black hole binaries are shown in Fig. 1.4. We review two analytical approximations below.

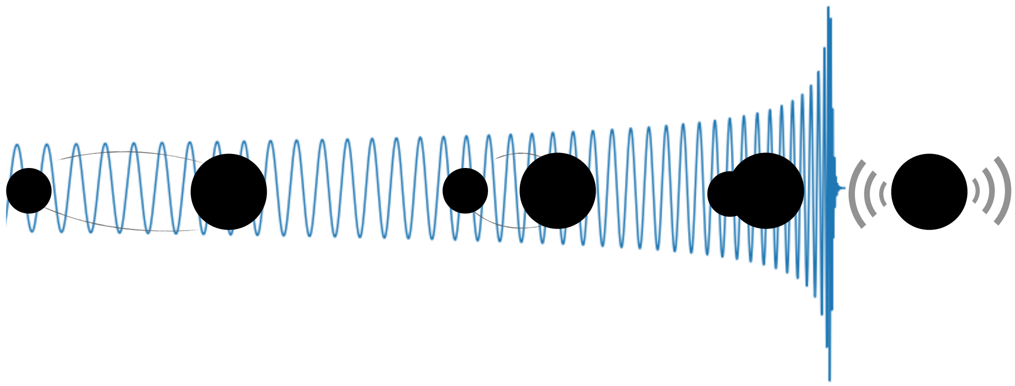

Figure 1.3: **Cartoon of the evolution of a BBH merger with the resulting GW signal**: Inspiral: starting with the two BHs at large separation, the relative velocity is small in this phase and the post-Newtonian approximation can be used ($v/c \ll 1$). Late-inspiral and merger: as the two BHs emit energy they inspiral faster until they finally plunge. This is the most dynamical phase and produces the highest amplitude of the GWs signal. Due to the lack of analytical models that could be used to describe this phase, numerical relativity simulations are needed. Ringdown: after the two BHs finally merge, they produce a perturbed remnant BH that emits gravitational waves (rings down). Black-hole perturbation theory could be used in that phase to describe the final perturbed black hole.

1.4.1 Post-Newtonian Theory

Post-Newtonian (PN) theory is an analytical approximation to GR that is particularly useful for studying BBH systems in the weak gravitational field and slow motion regime, such as the early inspiral phase. This method involves expanding Einstein's equations in powers of a small parameter related to the ratio of the typical velocity of the objects to the speed of light, called the post-Newtonian parameter and defined as

$$x \equiv \left(\frac{GM\Omega}{c^3}\right)^{2/3}, \tag{1.5}$$

where $M = m_1 + m_2$ is the total mass of the binary and Ω is the orbital frequency. The PN parameter is related to both the relative velocity of the binary v and the

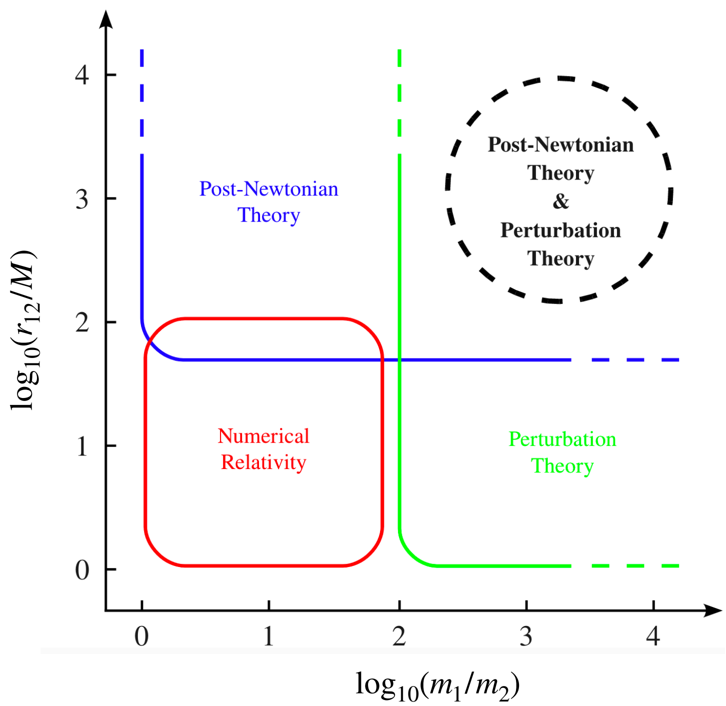

Figure 1.4: **The different schemes of studying BBH mergers**: Different analytical approximation schemes and numerical techniques used to study BBH mergers. The post-Newtonian approximation is used at large binary separation, regardless of the mass ratio. For comparable mass ratio binaries, numerical relativity simulations are required to model this highly dynamical phase. For extreme-mass-ratio binaries ($q \gg 1$), black-hole-perturbation theory is used. The PN approximation and BHPT are used at large separation for EMRIs. [Adapted from Blanchet [15]]

binary separation r_{12} (for bound systems) by

$$x \sim \frac{v^2}{c^2} \sim \frac{GM}{c^2 r_{12}}\,. \tag{1.6}$$

Henceforth in this thesis, we set $G = c = 1$. The PN approximation allows for the calculation of various aspects of BBH systems, including the equations of motion, the gravitational wave (GW) energy flux, and the orbital phase evolution. This approximation typically breaks down in the strong-field and high velocity regime, near the merger. Therefore it is only valid during the early inspiral phase of the binary, independently of the mass ratio $q = m_1/m_2$.

1.4.2 Black Hole Perturbation Theory

Another crucial analytical technique used to study BBH systems is black hole perturbation theory (BHPT), particularly in the case of extreme mass-ratio inspirals (EMRIs), which are a key source class for space-based GW detectors, such as LISA), where a stellar-mass compact-object inspirals into a supermassive black hole. In this approach, the gravitational field of the supermassive black hole is treated as a background spacetime (a Schwarzschild or Kerr black hole in GR), and the smaller object is considered as a perturbation to this background. By solving the linearized Einstein field equations on this background, one can derive highly accurate analytical waveforms in the limit where the mass ratio is very large ($q \gg 1$). The relevant solution of the linearized Einstein's equation that is relevant in this thesis is that of a Kerr background known as the Teukolsky equation [16], describing the last phase of the binary evolution, the ringdown. During the ringdown phase, the final perturbed black-hole radiates GWs at certain frequencies over characteristic time scales. The set of frequencies and damping times are known as the quasinormal modes (QNMs) which are entirely characterized by the final mass M_f and spin angular momentum S_f of the remnant BH, providing a direct test of the no-hair theorem. A Kerr's BH QNMs are the vacuum solutions of Teukolsky equation subject to the conditions that the waves cannot propagate out from the BH event horizon or in from infinity. A spectral method for solving the angular Teukolsky equation was introduced by Cook and Zalutskiy [17] as an improvement on the previously commonly used Leaver's method [18], that give the complex frequency ω and separation constant as functions of the dimensionless spin parameter $0 \leq a < 1$. The solutions to the angular Teukolsky equation are the spin-weighted spheroidal harmonics ${}_sS_{lm}(\theta, \phi; a\omega)$, which reduces to the more familiar spin-weighted spherical harmonics ${}_sY_{lm}(\theta, \phi)$ in the Schwarzschild limit ($a \to 0$). The relevant spin weight is $s = -2$ that corresponds to the Newman-Penrose scalar Ψ_4, describing outgoing radiation.

1.5 Gravitational-wave Memory Effect

One consequence of the strong gravity and high luminosities associated with BBH mergers is that the resulting GWs become strong enough that nonlinear interactions between the waves become significant. One of these nonlinear features is the nonlinear GW memory effect. The memory effect was first identified by Zel'dovich and Polnarev in 1974 [19], where they found in a linearized GR calculation that GWs radiated from

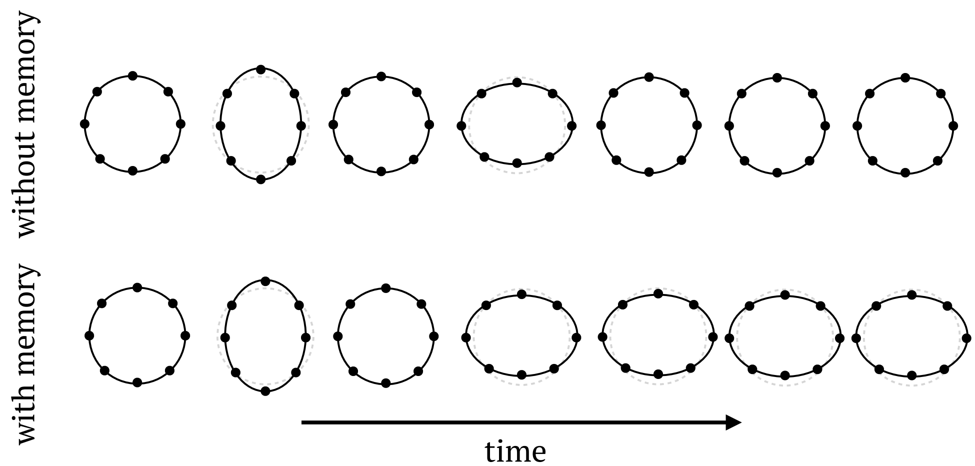

Figure 1.5: **The memory effect**: The effect of gravitational waves (propagating into the page) on a collection of free-falling observers (or test particles). As the GWs pass by the ring of particles experience the known effect of stretching and squeezing. After the wave passes, without the memory the particles would go back to their original positions as shown in the top row. However, with the memory the particles will experience a relative displacement that persists after the GWs pass.

the scattering of two compact objects will leave a permanent relative displacement between two free-falling observers that persists after the GWs pass. This type of memory was later called "linear memory" (or ordinary memory), since it was calculated using linearized GR. Related calculations of the memory effect within linearized gravity have been performed after this, such as neutrino emission from supernovae that was first noted by Turner [20] that it could be a source of memory.

While there is no memory effect that exist for bound systems in linearized gravity, it was found by Christodoulou [21] that interactions of GWs with themselves could be a source of memory effect, called "nonlinear memory". The nonlinear memory has also been called in literature "null memory" since it is a result of null radiation traveling to asymptotic null infinity ($r, t \to \infty$ at a constant Bondi time $u = t - r$), in contrast to the linear memory that is a result of astrophysical objects traveling on timelike paths to asymptotic infinity ($t \to \infty$). The nonlinear displacement memory was later calculated by Wiseman and Will [22] for compact binaries, followed by a PN expansion calculation of the effect by Blanchet and Damour [23].

There is an interesting connection between the displacement memory and the BMS group that describes the symmetries of asymptotically flat spacetimes. While the displacement memory can be understood as a permanent displacement that persists between two free-falling observers after the GWs pass as illustrated in Figure 1.5 (or permanent relative displacement between the two arms of the detector), it is also related to the supertranslation needed to reach a particular frame at later times from a related frame at early times [24]. As a result of that nice correspondence between the memory and supertranslations, the changes of the charges conjugate to supertranslations (supermomentum charges) are a source of the displacement memory. The conservation laws of flat spacetimes relate the changes in the Poincaré charges to the flux in these charges, but in asymptotically flat spacetimes, there are also changes in the BMS charges and their fluxes. The change in the supermomentum charges are closely related to the memory. The conservation laws of asymptotically flat spacetimes therefore have the form

$$\Delta \mathcal{Q} - \mathcal{F}_{\mathcal{Q}} = Memory \tag{1.7}$$

where $\mathcal{Q}$ is the BMS charge and $\mathcal{F}_{\mathcal{Q}}$ is the flux associated with that charge. The displacement memory is therefore sourced by changes in the supermomentum.

There are other subleading memory effects that have been studied; the spin memory and the center-of-mass (CM) memory, which are related to the magnetic and electric parity pieces of superrotations, respectively. While the displacement memory causes a change in the GW detector arm length (a change in the GW strain), the spin memory [25] causes a change in a portion of the time integral of the GW strain. The CM memory [26] also causes a change in the time integral of the strain, but for the other parity part of the strain.

The detection prospects of the displacement memory (the largest type of memory) have been studied before the detection of GWs from BBH mergers [27–31]. However, after the first detection of the GW150914 event by LIGO, it was noted that the GW signature from the memory effect is small compared to the dominant harmonic, despite the high SNR of that event. It was later proposed in [32] that the searches for the GW memory could be done in a population of BBH mergers with individual memory signals that are below the threshold, instead of from a single BBH merger. Forecasts of the detection of the GW memory have been performed that show that the memory could be detected during the fifth observing run [33–35], with LIGO's A+ configuration [36]. Specifically, it has been shown in [35] that the memory effect will likely be detected from a single event with the next generation ground-based detectors, Einstein Telescope and Cosmic Explorer which gives a direct test of whether the amplitude of the memory is consistent with the value predicted by GR to a few percent accuracy [37]. The detection of the memory effect from a single event is also possible with the upcoming space-base detector [37–40]; the Laser Interferometer Space Antenna (LISA) due to its sensitivity to GWs at low frequencies. Finally, searches for the memory are performed by NANOGrav and Parkes Pulsar Timing Array, where the memory accumulate on a timescale that is short compared to the shortest period they measure [41, 42].

1.6 Waveform modeling

The detection of a GW signal requires a waveform model for the signal that could be compared to the observation data. In the waveform of the GW strain from a BBH merger, the memory signal would look like a monotonic function that starts from zero (at past infinity) and slowly accumulates during the inspiral phase till it reaches the merger phase where it rapidly increases then finally saturates to a constant value during the ringdown. The waveform of the GW strain from an equal mass BBH merger is shown in Fig. 1.6 with and without the GW memory effect.

Numerical relativity (NR) simulations produce the most accurate waveforms that describe the coalescence of black holes. Because of the nonlinearity of Einstein's equations in GR, solving these equations for a systems of a BBH merger is very computationally expensive and requires supercomputers. Solving Einstein's equations is typically done in NR simulations by constructing initial data for the spacetime then evolving the initial data using Einstein's equations. Since the first stable simulation of BBH mergers was done [43], several NR codes for solving Einstein's equations were

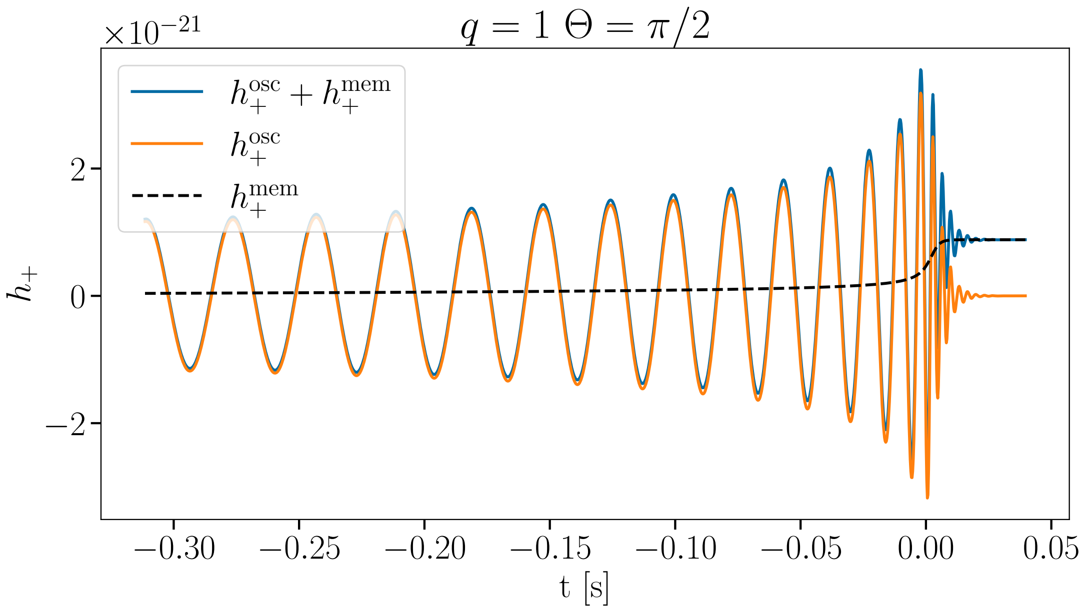

Figure 1.6: **Waveform of a GW with and without the memory**: Gravitational waves (h_+ polarization) from an equal mass BBH system, with inclination angle of $\Theta = \pi/2$ (the binary is being observed edge-on, which produces the largest memory signal). The GW waveform without the memory effect is shown in the solid blue curve, while the waveform *with* the memory is shown in the solid orange curve. The accumulating memory signal is shown in the dashed black curve.

developed. One such code is the Spectral Einstein Code, created by the Simulating eXtreme Spacetimes Collaboration. Because these numerical relativity simulations can be computationally expensive to run, there have been a need to develop models that use the output of the NR simulations to speed up the waveform evaluations. These models could be surrogate models that are modeling the NR waveforms directly [44–47] or in the form of phenomenological models that aim to simply describe the different features of the NR waveforms.

Because surrogate models are directly trained on NR waveforms, they are the most accurate to reproduce the NR waveforms. The process of building these models involves the following steps. A set of waveforms is computed using NR simulations for selected points in parameter space (e.g., different mass rations, spins). From the computed full set of waveforms, the algorithm identifies a smaller bases set that captures most of the variations in the waveform space. Then a small set of time nodes (sampling points) is chosen such that they can represent the whole waveform effectively. These points are then used to construct the full waveform using only the values at the the selected times. At each interpolation point, a fit across the parameter space (of mass ratios or spins) is constructed to model how the waveform changes with these parameters. This step builds a smooth, interpolated function that can be evaluated at arbitrary parameter values (that might not be included in the bases set). Finally, to evaluate the waveform at a random parameter point, the fit is evaluated for this point which gives the basis coefficients that are then used to construct the full waveform. The result is a fast and accurate model that reproduces the original waveforms to high precision but at a fraction of the computational cost. We use different surrogate models throughout this thesis to produce the waveforms for the GW multipole moments, specifically the NRHybSur3dq8 [47] that covers a certain range on the parameter space that we are considering.

NR waveforms (and therefore the surrogate models trained on them) computed using extrapolation techniques fail to capture the memory. However, simulations done using Cauchy-characteristic extraction (CCE) [48, 49] are able to resolve the memory signal [31, 50]. The reason behind this difference comes down to how each approach evolve Einstein's equations to null infinity (where the memory is properly defined). The extrapolation method (often used in NR simulations) computes the GW strain at finite radii, then extrapolates to $r \to \infty$ assuming a particular falloff in $1/r$ for the strain. This does not truly solve Einstein's equations along the outgoing null rays that link the simulated region to future null infinity, and can miss some of the nonlinear physics sourced during the propagation of the GW signal to future null infinity (including, the memory effect). CCE however uses the results of the usual Cauchy simulations as initial data that gets evolved on null hypersurfaces that connect the finite volume of Cauchy simulation to future null infinity. This method therefore evolves Einstein's equations in the whole spacetime.

Matched-filtering-based searches for the memory require a signal model that is both accurate and rapid-to-evaluate, since these searches require many evaluations for the model. However, generating these numerical-relativity waveforms is computationally intensive, especially since they need to be generated multiple times to produce sufficient templates for GW searches and parameter estimation. The displacement memory can be computed from the balance laws for the flux of supermomentum, using waveform models for the oscillatory modes of the GW strain. Computing the memory from the GW strain multipole moments require many numerical calculations: differentiating these modes, integrating different products of the modes, then summing the different products to compute the memory signal. These numerical calculations can be expensive and slow searches for the memory signal in the detector's data.

1.7 Summary and outline of the thesis

We now provide a summary and outline of this thesis. In Chapter 2 we study the symmetries of asymptotically flat spacetimes in general relativity, the BMS group and its proposed extensions. We discuss the charges associated with these symmetries, particularly the supermomentum and the relativistic angular momentum. The extensions of the BMS group includes a generalization for the angular momentum called super angular momentum. We investigate a subtlety that exists in the definition of angular momentum that exists from using different formalisms to define the angular momentum, leading to nonequivalent definitions for the charge. The different definitions could be summarized in a two-parameter family of angular momentum that satisfy reasonable physical conditions. We found that requiring that the angular momentum charge should vanish in flat spacetime forces the two parameters to be equal, reducing this two-parameter family to a one-parameter family of angular momentum. We then investigated the effect of this one parameter on the angular momentum in the context of nonspinning BBH mergers. We found that the angular momentum definitions only differ while the system is radiating gravitational waves, but agree at early and late times after the GWs pass. We also proposed a similar one parameter family of the super angular momentum and found that the different definitions do *not* agree even at late times after the passage of GWs, which is a result of the memory effect. We compute the difference in both the angular momentum and super angular momentum using numerical relativity surrogate waveforms of BBH mergers. We found that the difference is small but could be resolved given the current accuracy of numerical relativity simulations.

Chapters 3 and 4 cover the topic of the gravitational-wave memory effect and developing accurate waveform models that are both accurate and fast to evaluate. We focus on the $l = 2$, $m = 0$ spin-weighted spherical-harmonic mode of the memory signal from nonspinning BBH mergers in quasicircular orbits, which is the dominant memory mode for these systems. In Chapter 3, we take the first step of developing this waveform model by developing a model for the final memory offset (the net change in the GW strain between early and late times). We discuss two main aspects of constructing this model:

(i) Computing the memory signal for extreme-mass-ratio inspirals using a high post-Newtonian calculation (22PN-order), as a function of the PN parameter (or time).

(ii) Two polynomial fits for the final memory offset. One fit exclusively used numerical-relativity data, and the other fit used NR data in combination with the result of the extreme-mass-ratio calculation (where we used the EMRI result to fix the coefficient in the fit that is linear in the mass ratio).

We also provide a fit for the memory accumulated during the early inspiral phase that could be used to determine the initial data in CCE simulations. The fit for the final memory offset could also be used for interpreting the results of a potential pulsar-timing-array detection of a signal with GW memory, where the memory signal accumulates on a time-scale that is shorter that the shortest period they can measure.

In Chapter 4, we construct the time-domain model for the GW memory effect. We construct this model over three main phases of the binary evolution:

(i) Inspiral memory, where we use post-Newtonian approximation to model the memory in that phase.

(ii) Ringdown memory, where we use black-hole-perturbation theory to model the memory in terms of a superposition of products of quasi-normal modes.

(iii) Intermediate memory, where we used a phenomenological approach to bridge between the inspiral model and the ringdown model.

We enforce continuity of the memory signal as well as its first and second time derivatives to ensure smoothness of the time-domain model. We use the fit for the memory offset developed in Chapter 3 to require that the memory from the time-domain model saturates to the value computed from this fit. We also compute the analytic Fourier transform of the time-domain model to obtain a frequency-domain model for the GW memory effect. Finally, we compute the mismatch between the memory model and the true NR memory computed from surrogate models to evaluate the accuracy of our model. We use the advanced LIGO sensitivity curve from the fourth observing run to compute the mismatch for different BBH systems.

Chapter 2

Definitions of (super) angular momentum in asymptotically flat spacetimes

A. Elhashash and D. Nichols, Phys. Rev. D **104**, 024020 (2021),

2.1 Abstract

The symmetries of asymptotically flat spacetimes in general relativity are given by the Bondi-Metzner-Sachs (BMS) group, though there are proposed generalizations of its symmetry algebra. Associated with each symmetry is a charge and a flux, and the values of these charges and their changes can characterize a spacetime. The charges of the BMS group are relativistic angular momentum and supermomentum (which includes 4-momentum); the extensions of the BMS algebra also include generalizations of angular momentum called "super angular momentum." Several different formalisms have been used to define angular momentum, and they produce nonequivalent expressions for the charge. It was shown recently that these definitions can be summarized in a two-parameter family of angular momenta, which we investigate in this paper.

We find that requiring that the angular momentum vanishes in flat spacetime restricts the two parameters to be equal. If we do not require that the angular momentum agrees with a common Hamiltonian definition, then we are left with a one-parameter family of angular momenta that includes the definitions from the several different formalisms. We then also propose a similar two-parameter family of super angular momentum. We examine the effect of the free parameters on the values of the angular momentum and super angular momentum from nonprecessing binary-black-hole mergers. The definitions of angular momentum differ at a high post-Newtonian order for these systems, but only when the system is radiating gravitational waves (not before and after). The different super-angular-momentum definitions occur at lower orders, and there is a difference in the change of super angular momentum even after the gravitational waves pass, which arises because of the gravitational-wave memory effect. We estimate the size of these effects using numerical-relativity surrogate waveforms and find they are small but resolvable.

2.2 Introduction

The LIGO, Virgo, and KAGRA collaborations have now announced the detection of almost fifty binary-black-hole (BBH) mergers during the first three observing runs of the advanced-detector era beginning in 2015 [51, 52]. There are a few ways in which these BBH mergers are characterized: for example, by the masses and spins of the individual black holes (BHs) plus the orbital elements of the binary at a given reference frequency or by the final mass and spin of the BH formed after the merger and ringdown (e.g., [51,52]). An alternate way to characterize asymptotically flat systems is in terms of the "conserved" quantities conjugate to the symmetries of asymptotically flat spacetimes and the net fluxes of these conserved quantities. The symmetries of asymptotically flat spacetimes form the Bondi-Metzner-Sachs (BMS) group, which consists of transformations isomorphic to the Lorentz group and supertranslations (of which the four spacetime translations are a subgroup) [3–5]. The radiated energy and linear momentum (often expressed as a recoil velocity) being the quantities conjugate to the translation symmetries are often quoted when describing BBH mergers (see, e.g., [53] and references therein).

The flux of angular momentum (the quantity related to Lorentz symmetries) is somewhat more subtle. Angular momentum must be computed about an origin in flat spacetime; in terms of the symmetries that form the Poincaré group, this implies that a translation must be specified to identify the particular Lorentz transformation under consideration. There is thus a four-parameter family of Lorentz transformations spanned by a basis of the spacetime translations in the Poincaré group. In asymptotically flat spacetimes, this four-parameter family is enlarged to a countably infinite family of Lorentz transformations, each of which is associated with some basis element of the infinite-dimensional supertranslation subgroup in the BMS group. In stationary spacetimes, there is a natural way to choose a "preferred" set of supertranslations that reduces the dependence of the angular momentum to a choice of origin as in flat spacetime (see [54, 55] or more recently [24]); however, in nonstationary solutions, there is no such natural choice, though there are several different proposals to "fix" the supertranslation freedom (see, e.g., [56] for a review). The absence of this preferred Poincaré group is referred to as the "supertranslation ambigutity" of angular momentum in asymptotically flat spacetimes, which is, in essence, a statement that angular momentum in asymptotically flat spacetimes is different from its counterpart in flat spacetimes.

This additional complexity in describing the value of angular momentum for an asymptotically flat spacetime may have contributed to it and its flux being less frequently quoted in the output of numerical-relativity (NR) simulations of merging black holes. The six degrees of freedom in the relativistic angular momentum are often split into the three spin parts (corresponding to rotations) and three center-of-mass (CM) parts (corresponding to Lorentz boosts). Of these six components, the most commonly given from NR simulations of BBHs are the magnitude of the final BH's spin (though this spin is most often computed from quasilocal constructions on the BH's apparent horizon rather than in terms of quantities measured at or near future null infinity [57–59]); additional components of the angular momentum were computed in [60], for example.

In addition to the supertranslation ambiguities, a number of different definitions of the angular momentum of an asymptotically flat spacetime were (and continue to be) used. A nonexhaustive list of some of these definitions include one based on the Landau-Lifshitz pseudotensor for the intrinsic part of the angular momentum (in the

CM frame of the source) [61], a definition based on constructions called "linkages" [62], ones inspired from twistor theory [63,64], and those related to Hamiltonians conjugate to conserved quantities [14,65]. When considered in their respective domains of validity, the different definitions of the angular momentum described above agree [14, 25]. More recently, however, new definitions of angular momenta arose from revisiting the Landau-Lifshitz formalism when not restricted to the CM frame [66] and from considerations about soft theorems [67] (particularly a subleading correction to Weinberg's soft theorem [68]; see [2] for a review).

It was pointed out in [69] that these new definitions of angular momentum differ from the Hamiltonian definition of Wald and Zoupas [14].[1] Moreover, it was shown that the discrepancies in these definitions can be written in terms of two functions that are quadratic in the shear related to the outgoing GWs in asymptotically flat spacetimes. The different definitions were parametrized in terms of two real coefficients multiplying these two quadratic functions, respectively, and when the coefficients equal one, the Hamiltonian definition of [14] is recovered. All members of this two-parameter family of angular momenta satisfy flux balance laws, are covariant with respect to quantities defined on 2-sphere cross sections of null infinity, and lead to the same correspondence with the subleading soft theorem [69]. This led Compère

[1] Note that what we call the six-parameter (Lorentz-covariant) angular momentum, Compère *et al.* in [69] call the "Lorentz charge." We also have different usages for how we describe the parts that correspond to the rotations and the Lorentz boosts. We both call the part corresponding to Lorentz boosts "center-of-mass angular momentum," but Compère *et al.* call the parts corresponding to rotations simply "angular momentum," whereas we refer to it as "intrinsic" or "spin" angular momentum, because it reduces to those quantities in the rest-frame of the source.

et al. in [69] to conclude that there was not a compelling physical reason to prefer one definition over another and to suggest that there could be a two-parameter family of self-consistent definitions of angular momentum of asymptotically flat spacetimes. Compère *et al.* later described in [70] the sense in which these different definitions can all be considered to be Hamiltonian definitions [which is why we take care to describe which (or whose) Hamiltonian definition of the charge is being used].

In this chapter, we investigate this new two-parameter family of angular momenta in greater detail. Ashtekar and Winicour [71] had a larger set of criteria that a charge at null infinity should satisfy than the conditions discussed in [69].[2] Among these conditions was requiring that the charges and fluxes vanish in flat spacetime. We find that if we require the angular momentum to vanish in flat spacetime, then two of the parameters must be equal, thereby reducing the two parameters to one. This calculation further implies that the one-parameter family of angular momenta will agree in any region of spacetime in which there is only electric-parity shear (which includes stationary solutions and some radiative solutions). If we do not require that the angular momentum agree with the Wald-Zoupas definition, then we are left with a one-parameter definition that encompasses several other definitions used in the literature.

Ashtekar and Winicour further require that a charge agree with the Komar formula whenever there is an exact (as opposed to asymptotic) symmetry. The same calculation showing that the charge vanishes in flat spacetime also implies that the angular

[2] We thank Laurent Friedel for pointing out this reference to us.

momentum will agree with the Komar formula [72] in regions of vanishing electric-partiy shear (which include stationary regions); however, in regions with shear of generic parity, it is only the Wald-Zoupas charge that agrees with the Komar formula (by construction).[3] While this is arguably a compelling reason to consider only the Wald-Zoupas charge, we do not aim to settle the issue of whether there is a preferred definition of angular momentum among this one-parameter family here; rather, we explore whether the different commonly used definitions of angular momentum have significant differences for strongly gravitating and dynamical systems, such as the binary black holes, which have been measured observationally by LIGO and Virgo. In this sense, our investigation is similar in spirit to that of [73], in which the effect of the supertranslation ambiguities on the angular momentum radiated from compact-binary coalescences was studied as a way to assess how large the effect could be for this class of sources.

With this approach in mind, for this residual one-parameter family of angular momenta, we expand the difference of the angular momentum from the Wald-Zoupas definition in terms of spin-weighted spherical-harmonic moments of the GW strain. These difference terms involve only products of electric- and magnetic-type spherical-harmonic coefficients (unlike the flux of the Wald-Zoupas angular momentum), which is consistent with the results of [69]. This implies that the difference will vanish in stationary regions of spacetimes and nonradiative regions of spacetime with vanishing magnetic shear, though more generally, it will not vanish. We compute the time-

[3] We thank Kartik Prabhu for making us aware of this property of the angular momentum.

dependent difference terms for nonspinning BBH mergers, and we find that they are small compared to the total radiated angular momentum.

In addition to the BMS group, there are two different proposals for larger symmetry groups or algebras of asymptotically flat spacetimes. The first, due to Barnich and Troessaert [8–10], considers all the conformal Killing vectors of the 2-sphere, rather than the globally defined vectors, which are isomorphic to the Lorentz group. These vectors were dubbed "super-rotations," and, analogously to the supertranslations, they are a kind of asymptotic angle-dependent rotations and Lorentz boosts. To maintain the algebra structure of these asymptotic symmetries, the supertranslations must be correspondingly modified. A second extended symmetry group, due to Campiglia and Laddha [11, 12], considers all the diffeomorphisms of the 2-sphere rather than those equal to the Lorentz transformations, but the supertranslations are the same as in the BMS group. The 2-sphere diffeomorphisms are often referred to as super Lorentz transformations [74].

Both the super-rotations and super Lorentz transformations have corresponding conserved charges. The charges for both algebras have been called "super angular momentum," but they have also been called simply super-rotation charges or super Lorentz charges, for the respective algebras. We shall primarily focus on the generalized BMS algebra, and we shall refer to the charges associated with this algebra as the super angular momentum (and will call those associated with the super-rotations the "super-rotation charges."). Note that we will call the split of the charges into their electric- and magnetic-parity parts by super CM and superspin, respectively, in analogy with the convention used initially in [24] for the super-rotation charges, and subsequently for the super angular momentum in [25, 26].[4]

The super-rotation charges have a similar form to the angular momenta, but a super-rotation vector field enters into the expression for the charge rather than a Lorentz vector field (see, e.g., [10,24]). The super Lorentz charges constructed defined in [74] also have a similar form to the angular momentum with the Lorentz vector field is replaced by a super Lorentz transformation, but they have an additional term linear in the shear tensor needed to satisfy a flux balance law [74]. Given that there is a one-parameter family of angular momentum that satisfies a number of reasonable physical conditions, it is also natural to ask whether there is such a parametrization for the super angular momentum. We investigate this issue as well by allowing for a two-parameter family of super angular momentum that generalizes the Hamiltonian definition of [74] in a way completely analogous to the two-parameter extension of the Wald-Zoupas angular momentum given in [69]. In this case, setting the parameters to be equal (thereby reducing it to a one-parameter family) does not seem to make the super Lorentz charges vanish. This is consistent with a calculation performed by Compère and Long [75] for the Hamiltonian charges. There is a choice of parameters that makes the super angular momentum vanish, but this choice does not correspond to the Hamiltonian definition of [74]. Rather, this choice is the same as the one used

[4] This is a second discrepancy with the nomenclature used in [69]. There, what we call superspin is called super angular momentum, and what we call super angular momentum is called a super Lorentz charge. Our usages of super center-of-mass are equivalent, however.

in [70] to determine a representation of the extended BMS algebra in nonradiative regions of spacetime for the super Lorentz charges in terms of the standard Poisson bracket. This also leads to the possibility that properties of the generalized BMS algebra and charges could provide a criteria to prefer a certain definition of the angular momentum (though we will not discuss this possibility further in this paper; see instead [76]).

We then compute the multipolar expansion of the difference of the two-parameter family of super angular momentum from the Hamiltonian super angular momentum of [74]. This allows us to see that unlike the angular momentum, the change in the difference in the super angular momentum will be nonvanishing even in stationary regions. As a concrete example, we estimate the value of the change in the difference of the super angular momentum for nonspinning, quasicircular BBH mergers. The relative size of the net change in Hamiltonian value of the super angular momentum and the net change in the difference term is small for these BBH mergers (a roughly one-percent effect). Although it is small, it can be resolved given the current accuracy of numerical relativity (NR) simulations.

Overview The outline of the rest of this chapter is as follows. Section 2.3 is mostly a review in which we introduce Bondi coordinates, the metric in these coordinates, the evolution equations for the Bondi mass and angular-momentum aspects, the (extended) BMS symmetries of asymptotically flat spacetimes, and the expressions for the various definitions of angular momentum in Bondi coordinates. We end this section, however, by introducing the proposed two-parameter definition of the super angular momentum. In Sec. 2.4, we compute the (super) angular momentum in flat spacetime (where we show two of the parameters must be equal for the angular momentum to vanish). In the next section, Sec. 2.5, we perform a multipolar expansion of the (super) angular momentum that is valid for general asymptotically flat spacetimes. In Sec. 2.6, we estimate the effect that the remaining free parameter in the angular momentum and super angular momentum has on BBH mergers of different mass ratios. We compute results in the post-Newtonian approximation and using NR surrogate waveforms. We conclude in Sec. 2.7. In Appendix A, we compare our multipolar expansion of the angular momentum with a related expansion performed in [69]. In this paper, we use geometric units $G = c = 1$, and the conventions on the metric and curvature tensors in [77].

2.3 Bondi-Sachs framework, symmetries, and charges

In this section, we review aspects of the Bondi-Sachs framework including the metric, some components of Einstein's equations, the asymptotic symmetries, and the corresponding charges. We then discuss different definitions of angular momentum and super angular momentum.

2.3.1 Metric and Einstein's equations

We will perform our calculations in Bondi coordinates [3,4] (u, r, θ^A), where $A = 1, 2$, and we review the properties of these coordinates and the solutions of Einstein's equations below. We will use the notation and conventions given in [24]. The metric in these coordinates is written in the form

$$ds^2 = -Ue^{2\beta}du^2 - 2e^{2\beta}dudr + r^2\gamma_{AB}(d\theta^A - U^A du)(d\theta^B - U^B du) \tag{2.1}$$

where the functions and tensors U, β, γ^{AB}, and U^A depend on all four Bondi coordinates (u, r, θ^A). The metric by construction satisfies the Bondi gauge conditions $g_{rr} = 0$ and $g_{rA} = 0$; Bondi coordinates also are defined such that $\det(\gamma_{AB}) = \gamma(\theta^A)$ is independent of u and r. Some important properties of these coordinates are that u is a retarded time variable (i.e., u =const. are null hypersurfaces), r is an areal radius, and θ^A (with $A = 1, 2$) are coordinates on 2-spheres of constant r and u.

Near future null infinity (i.e., where r is large), the metric functions U, β, γ_{AB}, and U^A can be expanded as series in $1/r$. Asymptotically flat solutions postulate a given form of the expansion of these Bondi metric functions. For the tensor γ_{AB} the conditions of asymptotic flatness generally impose

$$\gamma_{AB} = h_{AB} + \frac{1}{r}C_{AB} + O(r^{-2})\,, \tag{2.2}$$

where $h_{AB}(\theta^C)$ is the metric on the unit 2-sphere, C_{AB} is a function of (u, θ^A), and the determinant condition of Bondi gauge implies that $C_{AB}h^{AB} = 0$. The remaining functions U, β, and U^A are assumed to have the following limits as r approaches infinity[5]

$$\lim_{r\to\infty}\beta = \lim_{r\to\infty} U^A = 0\,, \qquad \lim_{r\to\infty} U = 1\,. \tag{2.3}$$

We will now specify to vacuum spacetimes to discuss Einstein's equations, for simplicity. The ru, rA, and trace of the AB components of Einstein's equations take the form of hypersurface equations that can be solved on surfaces of constant u by integrating radially outward. The form of these equations is summarized in the review [78], for example. The results of substituting Eq. (2.2) into these hypersurface equations, radially integrating, and applying the boundary conditions in Eq. (2.3) gives the following solutions for the remaining functions U, β, and U^A:

$$\beta = -\frac{1}{32r^2}C_{AB}C^{AB} + O(r^{-3})\,, \tag{2.4a}$$

$$U = 1 - \frac{2m}{r} + O(r^{-2})\,, \tag{2.4b}$$

$$\begin{aligned} U^A = &-\frac{1}{2r^2}D_B C^{AB} + \frac{1}{r^3}\Big[-\frac{2}{3}N^A + \frac{1}{16}D^A(C_{BC}C^{BC}) \\ &+ \frac{1}{2}C^{AB}D^C C_{BC}\Big] + O(r^{-4})\,. \end{aligned} \tag{2.4c}$$

We have introduced a number of new pieces of notation in the above equation, which we will now explain: First, the function $m(u,\theta^A)$ is the Bondi mass aspect and $N^A(u,\theta^B)$ is the angular momentum aspect. They are related to "functions of in-

[5] Although we consider generalized BMS charges in this paper, we still impose the standard boundary conditions of asymptotic flatness and assume h_{AB} is the round 2-sphere metric with constant Ricci scalar curvature and U approaches unity as r approaches infinity. We restrict to these conditions, because we consider binary-black-hole mergers in this paper. These are asymptotically flat solutions that remain in a fixed super Lorentz frame, and we then restrict to the trivial super Lorentz rest frame of the system. Even with this restriction on the set of super Lorentz frames, the super angular momentum is nontrivial for these spacetimes. If one considers a space of solutions that are super Lorentz transformed from the boundary conditions given here, then one would need to consider the more general set of boundary conditions given, e.g., in [74]

tegration" that arise from integrating the hypersurface equations radially. Second, in the above equation, we have raised and lowered indices of tensors and vectors on the 2-sphere using the metric h^{AB} (respectively h_{AB}). Third, we have defined the derivative operator D_A as the torsion-free, metric-compatible derivative associated with the metric h_{AB}.

The evolution equation for γ_{AB}, when expanded to leading order in $1/r$, shows that the u derivative of C_{AB} is unconstrained by Einstein's equations and is defined to be the Bondi news tensor $N_{AB} = \partial_u C_{AB}$. The leading-order parts of the uu and uA components of Einstein equations are the conservation equations, which look like evolution equations for the Bondi mass aspect m and the angular momentum aspect N_A at fixed radii:

$$\dot{m} = -\frac{1}{8}N_{AB}N^{AB} + \frac{1}{4}D_A D_B N^{AB} \tag{2.5a}$$

$$\dot{N}_A = D_A m + \frac{1}{4}D_B D_A D_C C^{BC} - \frac{1}{4}D_B D^B D^C C_{CA} + \frac{1}{4}D_B(N^{BC}C_{CA}) + \frac{1}{2}D_B N^{BC} C_{CA} \tag{2.5b}$$

These equations are important for establishing flux balance laws for the charges conjugate to the asymptotic symmetries that form the BMS group and its extensions; we turn to the subject of these symmetries in the next subsection.

2.3.2 Asymptotic symmetries

The Bondi-Metzner-Sachs (BMS) group [3, 5] can be obtained from set of transformations that preserve the Bondi gauge conditions of the metric (2.1) and the asymptotic form of the functions that appear in the metric [Eqs. (2.2) and (2.4)]. The BMS group is the semidirect product of the infinite-dimensional abelian group of supertranslations with a six-dimensional group of conformal transformations of the 2-sphere (which is isomorphic to the proper, isochronous Lorentz group). The four spacetime translations are a subgroup of the supertranslation group. More recent generalizations of the BMS algebra take two forms. (i) The first is the extended BMS algebra proposed by Barnich and Troessaert [8–10] (see also [79]). In this proposal, all conformal Killing vectors of the 2-sphere are added to the algebra, including those with complex-analytic singularities on the 2-sphere. These additional symmetry vector fields were dubbed super-rotations, and the vectors that are isomorphic to the Lorentz transformations are a subalgebra of the super-rotations. The supertranslations also are extended to include functions that are not necessarily smooth. (ii) The second proposal has been called the generalized BMS algebra, and is due to Campiglia and Laddha [11, 12]. Here all smooth diffeomorphisms of the 2-sphere are considered instead of those equivalent to the Lorentz transformations, but the supertranslations are the same as in the original BMS group (though it is no longer possible to identify a preferred spacetime translation subgroup [80]).

The BMS symmetries and their generalizations are described by infinitesimal vector fields $\vec{\xi}$ that formally are defined at future null infinity, the null boundary of an asymptotically flat spacetime in the covariant conformal approach of Penrose [6, 7]. The form of the vector fields at future null infinity can be written in Bondi coordinates by restricting the vector fields that preserve the Bondi gauge conditions and the fall off rates of the metric to the tangent space of surfaces of constant r, and then taking the limit as r goes to infinity. In this limit, the vector fields for the BMS group and its extensions all take the same form; they are parameterized by a scalar function $T(\theta^A)$ and a vector on the 2-sphere $Y^A(\theta^B)$:

$$\vec{\xi} = \left[T(\theta^A) + \frac{1}{2}uD_A Y^A(\theta^B)\right]\vec{\partial}_u + Y^A(\theta^B)\vec{\partial}_A \tag{2.6}$$

The function $T(\theta^A)$ parametrizes the supertranslations in the BMS algebra and its generalizations (for the standard and generalized BMS algebras, it is assumed to be a smooth function, whereas for the extended BMS algebra, it can have complex analytic singular points). The vector field $Y^A(\theta^B)$ is a conformal Killing vector on the 2-sphere for the standard and extended BMS algebras (it is spanned by a six-parameter basis for the standard BMS algebra, or an infinite dimensional basis for the extended BMS algebra), or a smooth vector field for the generalized BMS group.

The symmetries at future null infinity can also be extended into the interior of the spacetime at large, but finite r by requiring that the diffeomorphisms generated by these vector fields preserve the Bondi gauge conditions and the asymptotic fall-off conditions imposed on the metric. Under these transformations, the functions C_{AB}, N_{AB}, m, and N_A transform in a nontrivial way. For the discussion that follows, we will only need the transformation law for C_{AB}, and we denote this transformation by $C_{AB} \to C_{AB} + \delta_\xi C_{AB}$, which was derived, e.g., in [9]. It is convenient to first define a quantity

$$f = T + \frac{u}{2} D_A Y^A \,, \tag{2.7}$$

which appears in $\delta_\xi C_{AB}$ as follows:

$$\delta_\xi C_{AB} = f N_{AB} - (2D_A D_B - h_{AB} D^2) f + \mathcal{L}_Y C_{AB} - \frac{1}{2} D_C Y^C C_{AB} \,. \tag{2.8}$$

This transformation of C_{AB} is useful for defining fluxes of conserved quantities associated with the BMS symmetries, which we will discuss in the next subsection. Before we do so, it is useful to introduce a decomposition of the tensor C_{AB} into its electric and magnetic (parity) parts as follows:

$$C_{AB} = \left(D_A D_B - \frac{1}{2} h_{AB} D^2 \right) \Phi + \epsilon_{C(A} D_{B)} D^C \Psi \,. \tag{2.9}$$

The scalars Φ and Ψ are both smooth functions of the coordinates (u, θ^A). From the transformation of C_{AB} in Eq. (2.8), it follows that a supertranslation affects the electric part of C_{AB}, but leaves the magnetic part invariant. This property of the shear has been understood for quite some time (see, e.g., [55]).

2.3.3 Fluxes and charges

There are a few different prescriptions used to define the charges and the fluxes of charges that are associated with BMS symmetries. We will describe here the procedure of Wald and Zoupas [14], in which the charges and fluxes are computed using a generalization of Noether's theorem that allows for the charges to change from emitted fluxes of gravitational waves and other matter fields. We denote the charges by $Q_\xi[\mathcal{C}]$, where the charges depend linearly upon a BMS vector field $\vec{\xi}$ and are defined on a cross section of null infinity $\mathcal{C}$ (in Bondi coordinates, a surface of constant u at fixed r in the limit of $r \to \infty$). We call the flux $\mathcal{F}_\xi[\Delta\mathscr{I}]$. Like the charge, it has a linear dependence on a BMS vector field $\vec{\xi}$, but the flux depends on a region of null infinity $\Delta\mathscr{I}$ between two cuts (in Bondi coordinates, the region between two surfaces of constant u at fixed r in the limit of $r \to \infty$). The flux balance law for the charges requires that

$$Q_\xi[\mathcal{C}_2] - Q_\xi[\mathcal{C}_1] = \mathcal{F}_\xi[\Delta\mathscr{I}]\,. \tag{2.10}$$

The explicit expression for the flux has a simple form in Bondi coordinates in vacuum (see, e.g., [24])

$$\mathcal{F}_\xi[\Delta\mathscr{I}] = -\frac{1}{32\pi}\int_{\Delta\mathscr{I}} du\, d^2\Omega\, N^{AB}\delta_\xi C_{AB}\,, \tag{2.11}$$

where $\delta_\xi C_{AB}$ is given in Eq. (2.8) and $d^2\Omega$ is the area element on the 2-sphere cuts of constant u. Using Eq. (2.8) and the conservation equations for the Bondi mass and angular momentum aspects in Eq. (2.5), it is possible to show that the charge is given by

$$Q_\xi = \frac{1}{8\pi}\int_{\mathcal{C}} d^2\Omega \left\{ 2Tm + Y^A \left[N_A - uD_A m - \frac{1}{16}D_A(C_{BC}C^{BC}) - \frac{1}{4}C_{AB}D_C C^{BC} \right] \right\} \tag{2.12}$$

(again, see, e.g., [24]). We dropped the dependence of the charge on the cut $\mathcal{C}$ to simplify the notation, and because it is made explicit in the domain of the integral on the right-hand side of the equation.

When the vector field $\vec{\xi}$ has $Y^A = 0$ and $T \neq 0$, then it is a supertranslation, and the corresponding charge is the supermomentum. The other case, a vector field with $Y^A \neq 0$ and $T = 0$, has as its corresponding charge the angular momentum, when Y^A is equivalent to a Lorentz transformation for the standard BMS group. The angular momentum is often split into its intrinsic (or spin) and center-of-mass (CM) parts, which correspond to the rotation and boost symmetries in the Lorentz group, respectively. It was observed in [24] that the charge in Eq. (2.12) does not satisfy the flux balance law (2.10) for the extended or generalized BMS vector fields. A charge that does satisfy a flux balance for the super Lorentz charges was determined in [74]. It is the same as that in Eq. (2.12), up to the addition of two new terms linear in the tensor C_{AB}, and it is given below:[6]

$$\begin{aligned} Q_\xi = \frac{1}{8\pi}\int_{\mathcal{C}} d^2\Omega \Bigg\{ 2Tm + Y^A \Bigg[N_A - uD_A m - \frac{1}{16} D_A(C_{BC}C^{BC}) - \frac{1}{4} C_{AB} D_C C^{BC} \\ + \frac{u}{8}(D^2 D^B C_{AB} - D_B D_A D_C C^{BC}) \Bigg] \Bigg\} . \end{aligned} \tag{2.14}$$

Note that the integral of the two additional terms in the final line Eq. (2.14) can be shown to vanish for the Y^A corresponding to Lorentz vector fields; note also that these two terms in the integrand are proportional to a differential operator acting on the magnetic part Ψ of the shear in Eq. (2.9) (see, e.g., [24]). The super angular

[6] The flux for which this charge satisfies the flux balance law differs from Eq. (2.11). It is necessary to add a term of the form

$$\frac{1}{64\pi}\int_{\Delta\mathscr{I}} du\, d^2\Omega\, u(D^2 D^B N_{AB} - D_B D_A D_C N^{BC}) \tag{2.13}$$

to the right-hand side of Eq. (2.11) to restore the balance law with the definition of the charge in Eq. (2.14) (see [69] for further details).

momentum in (2.14) can be divided into a magnetic-parity part called superspin and an electric-parity part called super center-of-mass, in analogy to the standard angular momentum. In the next subsection, we focus on the angular momentum and discuss a subtlety in its definition.

2.3.4 Definitions of angular momentum and their properties

As discussed in the introduction, the angular momentum computed by Wald and Zoupas is not the only notion of the angular momentum of an isolated system that is commonly used. While a number of the different angular momenta are equivalent, not all the definitions agree. First, for convenience, let us specialize the general BMS charges in Eq. (2.14) to a vector field $\vec{\xi}$ with $T = 0$ and Y^A being a generator of Lorentz transformations:

$$Q_Y = \frac{1}{8\pi}\int_{\mathcal{C}} d^2\Omega\, Y^A \left[N_A - uD_A m - \frac{1}{16}D_A(C_{BC}C^{BC}) - \frac{1}{4}C_{AB}D_C C^{BC}\right] . \qquad (2.15)$$

We used the notation Q_Y rather than Q_ξ to emphasize that it depends only on Y^A. It has been shown in [14] that the flux of this angular momentum agrees with that of Ashtekar and Streubel [65] and the charge defined by Dray and Streubel [64] (which came from twistorial definitions of the angular momentum [63]). The Landau-Lifshitz definition of angular momentum in [61] (which is restricted to the center-of-mass frame of the source and averaged over a few wavelengths of the emitted gravitational waves) also agrees with the flux of the angular momentum charge in Eq. (2.15), when the expression is restricted to this context [25].

There are a few notable examples of definitions of angular momentum that differ from the one in Eq. (2.15), a fact that was recently pointed out in a paper by Compère *et al.* in [69]. First, in the context of conservation laws of gravitational scattering, a definition of an angular momentum involving just the mass and angular momentum aspects and the vector field on the 2-sphere, Y^A, was used in [67, 81] to define the (super) angular momentum: i.e.,

$$Q_Y^{(0)} = \frac{1}{8\pi} \int_{\mathcal{C}} d^2\Omega Y^A (N_A - u D_A m) \,. \tag{2.16}$$

Also recently, a more general definition of the Landau-Lifshitz angular momentum was proposed by by Bonga and Poisson [66], who no longer required that the result be defined in the CM frame or by averaging over a few wavelengths of the gravitational waves. They specialized to the intrinsic (as opposed to CM) angular momentum, which they defined by using a collection of vector fields on the 2-sphere, $Y_i^A = \epsilon^{AD}\partial_D n_i$. Here n_i is a unit vector normal to the 2-sphere in quasi-Cartesian coordinates constructed from the spatial Bondi coordinates (r, θ^A), and ϵ^{AD} is the Levi-Civita tensor on the unit 2-sphere. After converting their definition of the intrinsic angular momentum into our notation, their result can be written as

$$J_i = \frac{1}{8\pi}\int_{\mathcal{C}} d^2\Omega \epsilon^{AD}\partial_D n_i \left[N_A - u D_A m - \frac{3}{4} C_{AB} D_C C^{BC} \right] . \tag{2.17}$$

There is a definition of the CM part of the angular momentum in the Landau-Lifshitz formalism from Blanchet and Faye [82], but it was shown in [69] that it cannot easily be written in terms of the 2-sphere-covariant Bondi-metric functions. As we discuss further below, the three definitions of the angular momentum in Eqs. (2.15)–(2.17) all vanish in flat spacetime, give the same angular momentum of a Kerr black hole and satisfy flux balance laws; they thus appear to be equally viable definitions of the angular momentum of an isolated source.

Given that the angular momenta in Eqs. (2.15)–(2.17) differ in the factors in front of the two terms quadratic in C_{AB} in Eq. (2.15), Compère *et al.* [69] observed that a two-parameter family of charges could be defined by allowing the coefficients in front of these terms to be arbitrary real numbers. When the coefficients are restricted to specific values, the two-parameter family of charges reduces to one of the specific definitions in Eqs. (2.15)–(2.17). Thus, the two-parameter family of angular momentum of Compère *et al.* [69] is given by

$$Q_Y^{(\alpha,\beta)} = \frac{1}{8\pi}\int_{\mathcal{C}} d^2\Omega Y^A \left[N_A - uD_A m - \frac{\alpha}{4}C_{AB}D_C C^{BC} - \frac{\beta}{16}D_A(C_{BC}C^{BC})\right], \quad (2.18)$$

where α and β are real constants.[7] The Wald-Zoupas angular-momentum corresponds to the case $\alpha = \beta = 1$; the angular momentum in Eq. (2.16) corresponds to $\alpha = \beta = 0$; and the intrinsic angular momentum in Eq. (2.17) corresponds to $\alpha = 3$ (and β can take on any real value, because it does not contribute to the intrinsic part). For all values of α and β, the angular momentum in Eq. (2.18) satisfies flux balance laws, but it is not immediately apparent that they will vanish in flat spacetime. In the next section, we will derive the conditions under which the angular momentum in

[7] The terms $D_A(C_{BC}C^{BC})$ and $C_{AB}D_C C^{BC}$ form a kind of basis of vectors constructed from contractions of C_{AB} and $D_A C_{BC}$, in the sense that other possible contractions can be rewritten in terms of these two quantities [69].

Eq. (2.18) vanishes in flat spacetime.

2.3.5 Definitions of super angular momentum

The charge in Eq. (2.18) was defined specifically for the angular momentum. There are also differing definitions of the super angular momentum, however, because several of the definitions of the super angular momentum were defined through promoting the vector field Y^A that enters into the charge from a Lorentz vector field to a super Lorentz vector. The definition in Eq. (2.16) was also used for a super-rotation charge (where Y^A is a super-rotation vector field, for example), and this definition differs from that in Eq. (2.15). The main difference between the two charges is are the terms quadratic in the shear tensor. It thus seems reasonable to define a two-parameter family of charges that satisfy a flux balance law by generalizing Eq. (2.14) (when $T = 0$) to include real coefficients α and β in front of the terms quadratic in C_{AB}. Thus, we will also consider a two-parameter family of super angular momentum defined by

$$Q_Y^{(\alpha,\beta)} = \frac{1}{8\pi}\int_{\mathcal{C}} d^2\Omega\, Y^A \bigg[N_A - uD_A m + \frac{u}{8}(D^2 D^B C_{AB} - D_B D_A D_C C^{BC}) \\ - \frac{\alpha}{4} C_{AB} D_C C^{BC} - \frac{\beta}{16} D_A (C_{BC} C^{BC}) \bigg] . \tag{2.19}$$

We will investigate the properties of this charge in flat spacetime next.

2.4 (Super) angular momentum in flat spacetime

While the focus in this section will be determining the values of the coefficients α and β for which the angular momentum vanishes in flat spacetime, much of the calculation holds for any smooth vector field on the 2-sphere Y^A, and thus applies to the super angular momentum of the generalized BMS algebra.[8] In the derivation that follows, it is structured so that the first part applies to smooth generalized BMS vectors Y^A, and the next part is specified to Y^A that generate Lorentz transformations. Note that a similar calculation was performed by Compère and Long in [83] for the Wald-Zoupas charges (i.e., $\alpha = \beta = 1$).

In flat spacetime, there is no radiation, and the news tensor vanishes [84]. In this case, the Bondi mass aspect and the Bondi angular momentum are also proportional to components of the vacuum Riemann tensor (see, e.g., [24]) and thus they must also vanish. From Eq. (2.5b), one can then also show that Ψ, the scalar that parametrizes the magnetic part of C_{AB} must also vanish. Because C_{AB} is electric type, then by performing a supertranslation it follows from Eq. (2.8) that it is possible to choose a frame in which the tensor C_{AB} vanishes (note that from the transformation properties of m and N_A given in, e.g., [24], the mass and angular momentum aspects will remain zero under this transformation). We will not work in the frame where C_{AB} vanishes, but rather we will choose a frame where it has a nonzero electric part. Thus, the values of the relevant functions needed to compute the super angular momentum in

[8] Note however that if Y^A is a super-rotation vector field of the extended BMS algebra, then the singular points of the vector fields make integration by parts on the 2-sphere more challenging. Although the 2-sphere is a compact manifold without boundary, when integrating by parts one must carefully analyze the contributions that come from boundary-like terms at the singular points of the super-rotation vectors, which can contribute to the integral (see, e.g., [83] for further details).

Eq. (2.19) are given by

$$m = 0\,, \tag{2.20a}$$

$$N_A = 0\,, \tag{2.20b}$$

$$C_{AB} = \left(D_A D_B - \frac{1}{2}h_{AB}D^2\right)\Phi\,. \tag{2.20c}$$

In flat spacetime, therefore, the additional terms in the second line of Eq. (2.19) do not contribute, and the super angular momentum is given by

$$Q_Y^{(\alpha,\beta)} = -\frac{1}{128\pi}\int_{\mathcal{C}} d^2\Omega\left[4\alpha Y^A C_{AB} D_C C^{BC} + \beta Y^A D_A (C_{BC}C^{BC})\right]\,. \tag{2.21}$$

We will now substitute in the expression in Eq. (2.20c) for C_{AB} in Eq. (2.21) in several places, and begin simplifying the expression. Because we are assuming Y^A is a smooth vector on the 2-sphere and Φ is a smooth function, we can integrate the first term by parts and drop the terms involving divergences of vector fields on the 2-sphere. For the second term, we use the fact that the covariant derivative acting on the shear tensor in Eq. (2.20c) is given by

$$D^B C_{AB} = D^B D_A D_B \Phi - \frac{1}{2}D_A D^2\Phi\,. \tag{2.22}$$

We can then use the definition of the Riemann tensor (associated with the derivative operator D_A) to commute the first two covariant derivatives in the first term. We find that it can be written as

$$D^B C_{AB} = D_A D^2\Phi + R_{AB}D^B\Phi - \frac{1}{2}D_A D^2\Phi\,, \tag{2.23}$$

where R_{AB} is the Ricci tensor on the 2-sphere. Assuming that the metric is that of a round 2-sphere, then the scalar curvature of the sphere is given by $R = 2$, the Ricci tensor is $R_{AB} = h_{AB}$, and the Riemann tensor can be written as

$$R_{ABCD} = h_{AC}h_{BD} - h_{AD}h_{BC}\,. \tag{2.24}$$

This implies that $D^B C_{AB}$ simplifies to

$$D^B C_{AB} = \frac{1}{2} D_A (D^2 + 2)\Phi\,. \tag{2.25}$$

Next, substituting Eqs. (2.25) and (2.20c) into Eq. (2.21), we can write the charge in terms of Y^A, Φ, and derivative operators D_A (though we leave one term involving C_{AB}). If we integrate by parts once more for both the terms proportional to α and β, we find the super angular momentum is given by

$$\begin{aligned} Q_Y^{(\alpha,\beta)} =& \frac{1}{128\pi}\int_{\mathcal{C}} d^2\Omega \left\{ \beta D_A Y^A [D_B D_C \Phi D^B D^C \Phi - \frac{1}{2}(D^2\Phi)^2] \right. \\ & \left. + 2\alpha \left[D^B Y^A C_{AB} + \frac{1}{2} Y^A D_A (D^2+2)\Phi \right] (D^2+2)\Phi \right\}. \end{aligned} \tag{2.26}$$

While for each Φ and Y^A there should exist a choice of α and β that makes Q_Y vanish, a choice of α and β that makes the super angular momentum vanish for all Φ and Y^A in flat spacetime is $\alpha = \beta = 0$. However, it is not necessarily clear that one should require that the super angular momentum should vanish, as Compère and collaborators have argued that the super angular momentum can be used to distinguish vacuum states that differ by a supertranslation [74, 83]. We thus only identify $\alpha = \beta = 0$ as a choice that makes the super angular momentum vanish in flat spacetime, but do not require the charge to satisfy this property.

Angular momentum We do require that the charge Q_Y vanish for vectors Y^A that generate Lorentz transformations. We now continue our simplification of Eq. (2.26) by using the fact that Y^A is a conformal Killing vector on the 2-sphere; i.e., it satisfies the conformal Killing equation

$$2D_{(A}Y_{B)} - D_C Y^C h_{AB} = 0\,. \tag{2.27}$$

Because C_{AB} is symmetric and trace free, then $C_{AB}D^B Y^A$ involves only the symmetric-trace-free part of $D^B Y^A$. By the conformal Killing equation (2.27), however, $D^B Y^A$ is proportional to h_{AB}, so $C_{AB}D^B Y^A$ vanishes. After performing a large number of integration by parts (so as to write the expression mostly in terms of squares of Φ and its derivatives) and using the following identity

$$D^2 D^C \Phi = D^C D^2 \Phi + D^C \Phi\,, \tag{2.28}$$

we find that the angular momentum can be written as

$$\begin{aligned} Q_Y^{(\alpha,\beta)} = \frac{1}{256\pi}\int_{\mathcal{C}} d^2\Omega\, \big\{ &(D_A Y^A)\big[(\beta-\alpha)(D^2\Phi)^2 - 4\alpha\Phi^2 + 2(2\alpha-\beta)D_C\Phi D^C\Phi\big] \\ &-2D^2(D_A Y^A)\big[\alpha\Phi^2 - \beta D_C\Phi D^C\Phi\big] - 2\beta D_B D_C D_A Y^A D^B\Phi D^C\Phi\big\}\,. \end{aligned} \tag{2.29}$$

Conformal Killing vectors also satisfy the property that

$$(D^2+2)(D_A Y^A) = 0\,, \tag{2.30}$$

which leads to the cancellation of some terms proportional to α in Eq. (2.29). The globally defined conformal Killing vectors (the vector fields Y^A that can be written as a superposition of the six $l = 1$ vector spherical harmonics on the 2-sphere) satisfy the additional property

$$D_B D_C D_A Y^A = -h_{BC} D_A Y^A \,. \tag{2.31}$$

After using the property in Eq. (2.31) in Eq (2.29), we see that the angular momentum in flat spacetime can be written as

$$Q_Y^{(\alpha,\beta)} = \frac{1}{256\pi}(\beta - \alpha) \int_{\mathcal{C}} d^2\Omega D_A Y^A [(D^2\Phi)^2 - 4 D_C \Phi D^C \Phi] \,. \tag{2.32}$$

The intrinsic angular momentum (i.e., the charge $Q_Y^{(\alpha,\beta)}$ for vectors Y^A with $D_A Y^A = 0$) vanishes for all values of α and β. For the center-of-mass angular momentum (i.e., the charge with Y^A that has nonvanishing $D_A Y^A$), the charge will typically be nonvanishing unless $\alpha = \beta$. Having the physical requirement that the angular momentum should vanish in flat spacetime thus reduces the two-parameter family of charges to a one-parameter family given by α. We will typically work with this reduced one-parameter family in the rest of the paper, unless we note otherwise.

We conclude this section with an important note. Because our expressions for the mass and angular momentum aspects vanish in flat spacetime, our calculations in this section apply to the α- and β-dependent terms in any region of spacetime, where the tensor C_{AB} can be written in terms of the electric part as in Eq. (2.20c). While this section is then nominally about flat spacetime, the results in this part directly imply that the different definitions of angular momentum that vanish in flat spacetime will all agree in any region of spacetime with electric shear (stationary or radiative). In particular, the result that the angular momentum vanishes when $\alpha = \beta$ in flat spacetime means that in stationary regions, the angular momenta for any real value of α will be equivalent. Requiring the angular momentum takes on a particular value in a particular stationary solution cannot be used to restrict this remaining parameter α. For the angular momentum, we will then focus on the differences that arise in radiative regions with magnetic-parity shear. For the super angular momentum, which only manifestly vanishes when $\alpha = \beta = 0$, there can be differences in its value for distinct α and β values for the same spacetime, as we also illustrate in more detail below.

2.5 Multipolar expansion of the (super) angular momentum

We will first summarize our conventions for the spherical harmonics that we use in our multipolar expansion. We will then perform multipolar expansions of the super angular momentum, which we will subsequently specialize to the standard angular momentum.

Because the multipolar expansion of Hamiltonian charges and fluxes had been computed previously (see, e.g., [25,26,69]), we will focus on the difference of the two-parameter family of charges from the charge defined in [74]. Thus, for a vector field

Y^A we will write

$$Q_Y^{(\alpha,\beta)} = Q_Y^{(\alpha=1,\beta=1)} + (\alpha - 1)\delta Q_Y^{(\alpha=1)} + (\beta - 1)\delta Q_Y^{(\beta=1)}\,, \tag{2.33a}$$

where $Q_Y^{(\alpha=1,\beta=1)}$ is the charge with $\alpha = \beta = 1$ and $\delta Q_Y^{(\alpha=1)}$ and $\delta Q_Y^{(\beta=1)}$ are defined by

$$\delta Q_Y^{(\alpha=1)} = -\frac{1}{32\pi}\int_{\mathcal{C}} d^2\Omega\, Y^A C_{AB} D_C C^{BC}\,, \tag{2.33b}$$

$$\delta Q_Y^{(\beta=1)} = -\frac{1}{128\pi}\int_{\mathcal{C}} d^2\Omega\, Y^A D_A(C_{BC} C^{BC})\,. \tag{2.33c}$$

In the special case of angular momentum, we will also use the notation $\delta J_Y^{(\alpha=1)}$ and $\delta k_Y^{(\alpha=1)}$ (and similarly for the β term) for the difference in the intrinsic and CM angular momentum, respectively, associated with a vector Y^A (which is a rotation or Lorentz boost, respectively). A similar calculation was performed in [69]; however, here we also compute the α-dependent term in the CM angular momentum, and we write the result in terms of the multipole moments U_{lm} and V_{lm} (defined below) rather than the rank-l symmetric-trace-free (STF) tensors U_L and V_L (discussed in Appendix A). The moments U_{lm} and V_{lm} are somewhat easier to relate to the moments of the gravitational-wave strain h_{lm} that can be obtained from numerical-relativity simulations or surrogate models fit to simulations (the latter of which we will use later in Sec. 2.6).

In the cases where we restrict to $\alpha = \beta$ (so that the angular momentum vanishes in flat spacetime), then we will use the notation

$$Q_Y^{(\alpha=\beta)} = Q_Y^{(\alpha=\beta=1)} + (\alpha-1)\delta Q_Y^{(\alpha=\beta=1)}\,, \tag{2.34a}$$

where $Q_Y^{(\alpha=\beta=1)} = Q_Y^{(\alpha=1,\beta=1)}$ is the charge with $\alpha=\beta=1$ and $\delta Q_Y^{(\alpha=\beta=1)}$ is defined by

$$\delta Q_Y^{(\alpha=\beta=1)} = -\frac{1}{128\pi}\int_{\mathcal{C}} d^2\Omega\, Y^A[4C_{AB}D_C C^{BC} + D_A(C_{BC}C^{BC})]\,. \tag{2.34b}$$

We will similarly use the notation $\delta J_Y^{(\alpha=\beta=1)}$ and $\delta k_Y^{(\alpha=\beta=1)}$ for the intrinsic and CM angular momentum, respectively, when Y^A is a rotation or Lorentz boost (also respectively).

2.5.1 Spherical harmonics and multipolar expansion of the gravitational-wave data

In addition to the scalar spherical harmonics (with the usual Condon-Shortly phase convention), $Y_{lm}(\theta,\phi)$, we will use vector and tensor harmonics on the unit 2-sphere, which we define as in [25]. The vector harmonics are given by

$$T^A_{(e),lm} = \frac{1}{\sqrt{l(l+1)}} D^A Y_{lm}\,, \tag{2.35a}$$

$$T^A_{(b),lm} = \frac{1}{\sqrt{l(l+1)}} \epsilon^{AB} D_B Y_{lm}\,, \tag{2.35b}$$

which are nonzero for $l \geq 1$ and the tensor harmonics

$$T^{(e),lm}_{AB} = \frac{1}{2}\sqrt{\frac{2(l-2)!}{(l+2)!}}\left(2D_A D_B - h_{AB}D^2\right) Y_{lm}\,, \tag{2.36a}$$

$$T_{AB}^{(b),lm} = \sqrt{\frac{2(l-2)!}{(l+2)!}}\epsilon_{C(A}D_{B)}D^C Y_{lm}\,, \tag{2.36b}$$

which are nonzero for $l \geq 2$.

We use these harmonics to expand the shear tensor as

$$C^{AB} = \sum_{l,m}(U_{lm}T^{AB}_{(e),lm} + V_{lm}T^{AB}_{(b),lm})\,. \tag{2.37}$$

Because the shear is real, the coefficients in this expansion obey the properties

$$U_{l,-m} = (-1)^m \bar{U}_{lm}\,, \quad V_{l,-m} = (-1)^m \bar{V}_{lm}\,, \tag{2.38}$$

where the overline means to take the complex conjugate. By using Eqs. (2.35a)–(2.36b) and (2.23), we can write the covariant derivative of the shear tensor in terms of vector harmonics as follows:

$$D_C C^{BC} = \sum_{l,m}\sqrt{\frac{(l-1)(l+2)}{2}}(U_{lm}T^B_{(e),lm} - V_{lm}T^B_{(b),lm}). \tag{2.39}$$

The vector and tensor harmonics are related to spin-weighted spherical harmonics ${}_sY_{lm}$ of spin weight $s = \pm 1$ and $s = \pm 2$, respectively, and a complex null dual vector on the 2-sphere

$$m^A\partial_A = \frac{1}{\sqrt{2}}(\partial_\theta + i\csc\theta\partial_\phi)\,. \tag{2.40}$$

and its complex conjugate $\bar{m}^A$. The relationships for the vector harmonics are

$$T_A^{(e),lm} = \frac{1}{\sqrt{2}}\left({}_{-1}Y_{lm}m_A - {}_1Y_{lm}\bar{m}_A\right)\,, \tag{2.41a}$$

$$T_A^{(b),lm} = \frac{i}{\sqrt{2}}\left({}_{-1}Y_{lm}m_A + {}_1Y_{lm}\bar{m}_A\right)\,, \tag{2.41b}$$

and for the tensor harmonics are

$$T_{AB}^{(e),lm} = \frac{1}{\sqrt{2}} \left({}_{-2}Y_{lm} m_A m_B + {}_{2}Y_{lm} \bar{m}_A \bar{m}_B \right) , \tag{2.42a}$$

$$T_{AB}^{(b),lm} = -\frac{i}{\sqrt{2}} \left({}_{-2}Y_{lm} m_A m_B - {}_{2}Y_{lm} \bar{m}_A \bar{m}_B \right) . \tag{2.42b}$$

The spin-weighted spherical harmonics satisfy the well-known complex-conjugate property ${}_s\bar{Y}_{lm} = (-1)^{s+m} {}_{-s}Y_{l-m}$.

The charges are quadratic in C_{AB} and involve a vector field Y^A, and we will expand all three quantities in terms of spin-weighted spherical harmonics using Eqs. (2.35a)–(2.42b). When evaluating the charges, we will frequently encounter integrals of three spin-weighted spherical harmonics over S^2. We use the notation of [25] to describe these integrals, which we denote by

$$C_l(s', l', m'; s'', l'', m'') \equiv \int d^2\Omega \, ({}_{s'+s''}\bar{Y}_{lm'+m''})({}_{s'}Y_{l'm'})({}_{s''}Y_{l''m''}) . \tag{2.43}$$

The complex-conjugated spherical harmonic ${}_{s'+s''}\bar{Y}_{lm'+m''}$ has spin-weight $s = s' + s''$ and azimuthal number $m = m' + m''$, because for all other values of s and m, the integral vanishes. It can be shown that the coefficients $C_l(s', l', m'; s'', l'', m'')$ can be written in terms of Clebsch-Gordon coefficients $\langle l', m'; l'', m'' | l, m' + m'' \rangle$ as follows:

$$\begin{aligned} C_l(s', l', m'; s'', l'', m'') = {} & (-1)^{l+l'+l''} \sqrt{\frac{(2l'+1)(2l''+1)}{4\pi(2l+1)}} \\ & \times \langle l', s'; l'', s'' | l, s' + s'' \rangle \langle l', m'; l'', m'' | l, m' + m'' \rangle \, . \end{aligned} \tag{2.44}$$

The coefficients are also nonvanishing only when the l index is in the range $\{\max(|l' -$

$l''|, |m'+m''|, |s'+s''|), ..., l'+l''-1, l'+l''\}$. There are two additional useful identities under sign flips of the spin weight and azimuthal numbers that we will need in the discussion below

$$C_l(s', l', m'; s'', l'', m'') = (-1)^{l+l'+l''} C_l(-s', l', m'; -s'', l'', m''), \tag{2.45a}$$

$$C_l(s', l', m'; s'', l'', m'') = (-1)^{l+l'+l''} C_l(s', l', -m'; s'', l'', -m''). \tag{2.45b}$$

We can now turn to the evaluation of the terms $\delta Q_Y^{(\alpha=1)}$ and $\delta Q_Y^{(\beta=1)}$ in a few specific cases of interest next.

2.5.2 Multipolar expansion of the super angular momentum

In this part, we will compute the multipolar expansion of the α and β "difference terms" in Eqs. (2.33b) and (2.33c) from the super angular momentum of [74]. We will consider two types of vector fields Y^A to compute the charges: namely, the electric- and magnetic-parity vectors harmonics defined in Eqs. (2.35a) and (2.35b). We will thus denote these terms by $\delta Q_{(e),lm}^{(\alpha=1)}$ and $\delta Q_{(b),lm}^{(\alpha=1)}$, respectively, for Eq. (2.33b) and $\delta Q_{(e),lm}^{(\beta=1)}$ and $\delta Q_{(b),lm}^{(\beta=1)}$, respectively, for Eq. (2.33c). The results here hold for both the standard BMS charges (CM and intrinsic angular momentum) and the generalized BMS charges (super angular momentum). There are a number of additional simplifications that occur for the intrinsic and CM angular momentum, and we will therefore treat these simpler cases separately afterwards.

In this calculation, we will not require initially that the two parameters α and β be equal, because this choice was made to require that the standard (rather than the super) angular momentum vanishes in flat spacetimes. For the super angular momentum, the choice of $\alpha = \beta$ does not guarantee that these charges vanish in flat spacetimes, and it is not agreed upon universally that these charges should vanish in flat spacetime (see, e.g., [83]).

Before we begin the calculations, note that because $D^A T_A^{(b),lm} = 0$, then by performing an integration by parts of Eq. (2.33c), one can show that

$$\delta Q^{(\beta=1)}_{(b),lm} = 0\,; \tag{2.46}$$

we will thus focus on the three quantities $\delta Q^{(\alpha=1)}_{(e),lm}$, $\delta Q^{(\alpha=1)}_{(b),lm}$, and $\delta Q^{(\beta=1)}_{(e),lm}$. The calculation of these three quantities is quite similar, so we will describe in detail the procedure for just $\delta Q^{(\alpha=1)}_{(e),lm}$ (and the other two quantities can be determined through a nearly identical calculation).

Starting from Eq. (2.33b), we then substitute in the multipolar expansion of C_{AB} and $D_A C^{AB}$ given in Eqs. (2.37) and (2.39) and the vector spherical harmonic in Eq. (2.35a). We then use the relationships between the vector and tensor spherical harmonics and the spin-weighted spherical harmonics in Eqs. (2.41a)–(2.42b) to write $\delta Q^{(\alpha=1)}_{(e),lm}$ in terms of the multipole moments U_{lm} and V_{lm} as well as the integrals of three spin-weighted spherical harmonics in Eq. (2.43). We then make use of the identities for the coefficients $C_l(s', l', m'; s'', l'', m'')$ in Eq. (2.45) and the complex conjugate properties of U_{lm} and V_{lm} in Eq. (2.38) to simplify the expression. It is useful to make the definitions (similar to those in [26])

$$s^{l,(\pm)}_{l';l''} = 1 \pm (-1)^{l+l'+l''} \,, \tag{2.47a}$$

$$f^{l}_{l',m';l'',m''} = \sqrt{(l'+2)(l'-1)}C_l(-1,l',m';2,l'',m'') \,, \tag{2.47b}$$

$$g^{l}_{l',m';l'',m''} = \sqrt{l(l+1)}C_l(-2,l',m';2,l'',m'') \,. \tag{2.47c}$$

The result can then be written as is

$$\delta Q^{(\alpha=1)}_{(e),lm} = -\frac{1}{128\pi} \sum_{l',m';l'',m''} f^{l}_{l',m';l'',m''} [s^{l,(+)}_{l';l''}(U_{l'm'}U_{l''m''} + V_{l'm'}V_{l''m''}) + i s^{l,(-)}_{l';l''}(U_{l'm'}V_{l''m''} - V_{l'm'}U_{l''m''})] \,, \tag{2.48a}$$

where the indices on the charges should be integers in the ranges $l \geq 1$ and $-l \leq m \leq l$, and where the sums run over integers in the ranges $l' \geq 2$, $-l' \leq m' \leq l'$, $l'' \geq 2$, and $-l'' \leq m'' \leq l''$ This gives the α-dependent difference from the super-CM charge of [74]. A similar calculation shows that the α-dependent correction to the superspin can be written as

$$\delta Q^{(\alpha=1)}_{(b),lm} = \frac{i}{128\pi} \sum_{l',m';l'',m''} f^{l}_{l',m';l'',m''} [s^{l,(-)}_{l';l''}(U_{l'm'}U_{l''m''} + V_{l'm'}V_{l''m''}) + i s^{l,(+)}_{l';l''}(U_{l'm'}V_{l''m''} - V_{l'm'}U_{l''m''})] \,. \tag{2.48b}$$

Finally, the β-dependent correction to the super-CM charge is given by

$$\delta Q^{(\beta=1)}_{(e),lm} = -\frac{1}{256\pi} \sum_{l',m';l'',m''} g^{l}_{l',m';l'',m''} [s^{l,(+)}_{l';l''}(U_{l'm'}U_{l''m''} + V_{l'm'}V_{l''m''}) + i s^{l,(-)}_{l';l''}(U_{l'm'}V_{l''m''} - V_{l'm'}U_{l''m''})] \,. \tag{2.48c}$$

The values of l, l', l'', m, m', and m'' in Eqs. (2.48b) and (2.48c) are the same as in Eq. (2.48a). From these difference terms and the super-CM and superspin charges with $\alpha = 1$ and $\beta = 1$ (i.e., $Q^{(\alpha=1,\beta=1)}_{(e),lm}$ and $Q^{(\alpha=1,\beta=1)}_{(b),lm}$) one can then construct the full α and β dependent super CM and superspin (i.e., $Q^{(\alpha,\beta)}_{(e),lm}$ and $Q^{(\alpha,\beta)}_{(b),lm}$).

Although we do not require that the superspin and super CM vanish in flat spacetime, it is still useful to write down the expressions for the α- and β-dependent difference terms in this case: namely, the quantities $\delta Q^{(\alpha=\beta=1)}_{(e),lm}$ and $\delta Q^{(\alpha=\beta=1)}_{(b),lm}$. It is then straightforward to specialize our previous results to find that

$$\delta Q^{(\alpha=\beta=1)}_{(e),lm} = -\frac{1}{256\pi} \sum_{l',m';l'',m''} (2 f^l_{l',m';l'',m''} + g^l_{l',m';l'',m''}) \\ \times [s^{l,(+)}_{l';l''} (U_{l'm'} U_{l''m''} + V_{l'm'} V_{l''m''}) + i s^{l,(-)}_{l';l''} (U_{l'm'} V_{l''m''} - V_{l'm'} U_{l''m''})] \,. \tag{2.49a}$$

The superspin is the same, because the term $\delta Q^{(\beta=1)}_{(b),lm}$ vanishes: i.e.,

$$\delta Q^{(\alpha=\beta=1)}_{(b),lm} = \delta Q^{(\alpha=1)}_{(b),lm} \,. \tag{2.49b}$$

In the next subsections, we will further specialize Eqs. (2.49a) and (2.49b) to $l = 1$ spherical harmonics to compute the CM and intrinsic angular momentum.

2.5.3 Multipolar expansion of the intrinsic angular momentum

We begin by simplifying the expression in Eq. (2.48b) in the case where $l = 1$ (which corresponds to the correction to the intrinsic angular momentum). When $l = 1$, the coefficients $f^1_{l',m';l'',m''}$ are nonvanishing for $l'' = l'$ or $l'' = l' \pm 1$. Thus, the coefficient $s^{1,(-)}_{l';l''}$ is nonvanishing only when $l'' = l'$ and the coefficient $s^{1,(+)}_{l';l''}$ is nonvanishing for $l'' = l' \pm 1$. Because the index m satisfies $m = 0$ or $m = \pm 1$, then for the first set of terms in Eq. (2.48b) proportional to $s^{1,(-)}_{l';l''}$ the nonzero terms in the double sum will be one of the terms of the form $f^1_{l',m';l',-m'}$ or $f^1_{l',m';l',-m'\pm 1}$. Given the complex-conjugate relationships for the U_{lm} and V_{lm} moments in Eq. (2.38) and the symmetries of the coefficients $C_1(-1, l', m'; 2, -l', m'')$ under the change of sign of m' in Eq. (2.45), then one can show that the terms proportional to $s^{1,(-)}_{l';l''}$ vanish. The difference term from the Wald-Zoupas angular momentum is then given by

$$\delta J^{(\alpha=1)}_{1,m} \equiv \delta Q^{(\alpha=1)}_{(b),1,m} = \frac{1}{128\pi} \sum_{l',m',l'',m''} s^{1,(+)}_{l';l''} f^1_{l',m';l',m''} (U_{l'm'} V_{l'',m''} - V_{l'm'} U_{l'',m''}) \,. \tag{2.50}$$

Note that although we left the expression as a double sum over l' and l'', the l'' sum is restricted to $l'' = l' - 1$ or $l'' = l' + 1$; similarly, the m'' sum is restricted to the values $m'' = m - m'$, where $m = 0$ or $m = \pm 1$. If we evaluate the coefficients $f^1_{l',m';l'\pm 1,-m'}$, $f^1_{l',m';l'\pm 1,-m'-1}$, and $f^1_{l',m';l'\pm 1,-m'+1}$ in the sum using the expression in Eq. (2.44), then the expressions can be simplified to square roots of rational functions in these cases. We follow [26] and define coefficients a_l, $b^{(\pm)}_{lm}$, c_{lm} and $d^{(\pm)}_{lm}$ by

$$a_l = \sqrt{\frac{(l-1)(l+3)}{(2l+1)(2l+3)}} \,, \tag{2.51a}$$

$$b^{(\pm)}_{lm} = \sqrt{(l \pm m + 1)(l \pm m + 2)} \,, \tag{2.51b}$$

$$c_{lm} = \sqrt{(l - m + 1)(l + m + 1)} \,, \tag{2.51c}$$

$$d_{lm}^{(\pm)} = \sqrt{(l \pm m + 1)(l \mp m)} \tag{2.51d}$$

(though we do not use $d_{lm}^{(\pm)}$ until the next subsection). In terms of these quantities, and after relabelling l' with l and m' with m in the sum, we can write the difference term from the Wald-Zoupas angular momentum as

$$\delta J_{1,0}^{(\alpha=1)} = \frac{1}{16}\sqrt{\frac{3}{2\pi}} \sum_{l\geq 2,m} \frac{a_l c_{lm}}{l+1} (\bar{U}_{lm} V_{l+1,m} - \bar{V}_{lm} U_{l+1,m}) \,, \tag{2.52a}$$

$$\delta J_{1,\pm 1}^{(\alpha=1)} = \frac{1}{32}\sqrt{\frac{3}{\pi}} \sum_{l\geq 2,m} \frac{a_l b_{lm}^{(\pm)}}{l+1} (\bar{U}_{lm} V_{l+1,m\pm 1} - \bar{V}_{lm} U_{l+1,m\pm 1}) \,. \tag{2.52b}$$

The calculation to arrive at these simplified expressions requires some relabelling of indices in the sum so that only terms with $l+1$ appear rather than $l-1$.

A similar calculation was performed in [69] using STF l-index tensors rather than expanding C_{AB} in the harmonics in Eq. (2.37). The two formalisms can be related, and we compared the result of the difference term in [69] for the intrinsic angular momentum to our expressions in Eqs. (2.52a) and (2.52b). We found that our result differs from Eq. (4.16) of [69] by an additional factor of $1/(l+1)$, and we could not identify from where this discrepancy was arising. We give a detailed calculation of this comparison in Appendix A. Given our results in the next subsection, we believe our result to be correct, so we suspect that the error lies in the conversion between the two formalisms.

2.5.4 Multipolar expansion of the center-of-mass angular momentum

We now derive a similar expression for the difference terms from the Wald-Zoupas center-of-mass angular momentum when expanded in terms of the mass and current multipole moments of C_{AB} in Eq. (2.37). We first give a result for general real coefficients α and β, and we then specify to the $\alpha = \beta$ choice. The calculation is quite similar to that in the previous subsection for the intrinsic angular momentum. When the expression in Eq. (2.48a) is restricted to $l = 1$, then there is again a similar cancellation of the terms proportional to $s^{1,(-)}_{l';l''}$ leaving just the terms proportional to $s^{1,(+)}_{l';l''}$. Again, because the allowed values of l'' are given by $l'' = l' \pm 1$, the coefficients $f^{1}_{l',m';l'',m''}$ simplify to square roots of rational functions. The α-dependent difference terms are then given by

$$\delta Q^{(\alpha=1)}_{(e),1,0} \equiv \delta k^{(\alpha=1)}_{1,0} = \frac{1}{16}\sqrt{\frac{3}{2\pi}} \sum_{l\geq 2,m} \frac{a_l c_{lm}}{l+1} (\bar{U}_{lm} U_{l+1,m} + \bar{V}_{lm} V_{l+1,m}) \,, \tag{2.53a}$$

$$\delta Q^{(\alpha=1)}_{(e),1,\pm 1} \equiv \delta k^{(\alpha=1)}_{1,\pm 1} = \frac{1}{32}\sqrt{\frac{3}{\pi}} \sum_{l\geq 2,m} \frac{a_l b^{(\pm)}_{lm}}{l+1} (\bar{U}_{lm} U_{l+1,m\pm 1} + \bar{V}_{lm} V_{l+1,m\pm 1}) \,, \tag{2.53b}$$

for the $m = 0$ and $m = \pm 1$ modes, respectively.

For the β-dependent difference term in Eq. (2.48c), it is no longer the case that the $s^{1,(-)}_{l';l''}$ terms vanish. However, because the coefficients $g^{1}_{l',m';l'',m''}$ also have the property that they vanish except when $l'' = l'$ or $l'' = l' \pm 1$ and when $m'' = m - m'$ for $m = 0$ or $m = \pm 1$, then the coefficients can similarly be evaluated in terms of rational functions and their square roots. The result of this calculation is as follows:

$$\delta Q^{(\beta=1)}_{(e),1,0} \equiv \delta k^{(\beta=1)}_{1,0} = -\frac{1}{16}\sqrt{\frac{3}{2\pi}}\sum_{l\geq 2,m}\frac{1}{l+1} \times \left[a_l c_{lm}(\bar{U}_{lm}U_{l+1,m} + \bar{V}_{lm}V_{l+1,m}) - \frac{2im}{l}\bar{U}_{lm}V_{lm}\right], \tag{2.54a}$$

$$\delta Q^{(\beta=1)}_{(e),1,\pm 1} \equiv \delta k^{(\beta=1)}_{1,\pm 1} = -\frac{1}{32}\sqrt{\frac{3}{\pi}}\sum_{l\geq 2,m}\frac{1}{l+1} \times \left[a_l b^{(\pm)}_{lm}(\bar{U}_{lm}U_{l+1,m\pm 1} + \bar{V}_{lm}V_{l+1,m\pm 1}) \pm \frac{2i}{l}d^{(\pm)}_{lm}\bar{U}_{lm}V_{l,m\pm 1}\right]. \tag{2.54b}$$

The coefficients $d^{(\pm)}_{lm}$ are defined in Eq. (2.51).

A significant simplification occurs when the two parameters are equal; only the terms involving products of U_{lm} and V_{lm} moments remain. We find that the result is given by

$$\delta Q^{(\alpha=\beta=1)}_{(e),1,0} \equiv \delta k^{(\alpha=\beta=1)}_{1,0} = \frac{i}{8}\sqrt{\frac{3}{2\pi}}\sum_{l\geq 2,m}\frac{m}{l(l+1)}\bar{U}_{lm}V_{lm}\,, \tag{2.55a}$$

$$\delta Q^{(\alpha=\beta=1)}_{(e),1,\pm 1} \equiv \delta k^{(\alpha=\beta=1)}_{1,\pm 1} = \mp\frac{i}{16}\sqrt{\frac{3}{\pi}}\sum_{l\geq 2,m}\frac{d^{(\pm)}_{lm}}{l(l+1)}\bar{U}_{lm}V_{l,m\pm 1}\,. \tag{2.55b}$$

This result is consistent with our calculation in flat spacetime in Sec. 2.4. In that section, we showed that when $\alpha = \beta$, the angular momentum should vanish in flat spacetime. Because the tensor C_{AB} can be decomposed using just electric-type tensor harmonics (i.e., the U_{lm} modes can be nonvanishing but all V_{lm} modes must vanish), then the multipolar expansion should not involve products of U_{lm} moments with other U_{lm} moments, because these terms would be nonvanishing in flat spacetime.

Our result for the β-dependent term in Eqs. (2.54) agrees with Eq. (4.17) of [69] after performing the same conversion between their STF l-index tensors and our mass and current multipoles U_{lm} and V_{lm}. This comparison is given in detail in Appendix A. The α-dependent terms in Eq. (2.53) was not computed in [69]. Note, however, that the coefficients in $\delta k_{1m}^{(\alpha=1)}$ in Eq. (2.53) that multiply the products of U_{lm} and V_{lm} moments are precisely the same ones that appear in Eq. (2.52) for $\delta J_{1m}^{(\alpha=1)}$. Since the coefficients are the same in Eqs. (2.52) and (2.53), and since these coefficients are needed to have the angular momentum vanish in flat spacetime, then this provides a consistency check on the result in Eq. (2.52).

Now that we have the multipolar expressions for the difference terms from the Wald-Zoupas definition of the angular momentum, it is possible to assess how large these terms are for different systems of interest. We will focus on nonspinning compact binaries in the next section.

2.6 Standard and super angular momentum for nonprecessing BBH mergers

In this part, we compute the effect of the remaining free parameter α on the standard and super angular momentum from nonprecessing binary-black-hole mergers. As discussed in the introduction, the value of the (super) angular momentum depends on a choice of Bondi frame. For the explicit calculations using PN theory and NR surrogate models in this section, we will work in the canonical frame (e.g., [24]) associated with the binary as $u \to -\infty$. This frame is a type of asymptotic rest frame in which $C_{AB} = 0$ and the system has vanishing mass dipole moment (i.e., a CM frame).

For the difference of the angular momentum from the Wald-Zoupas values [i.e., Eqs. (2.52) and (2.55)], this difference depends on products of both the U_{lm} and the V_{lm} modes. As we discuss in the first subsection in this part, the U_{lm} modes can be nonvanishing after the passage of GWs for these BBH mergers, because of the GW memory effect. The V_{lm} modes vanish after the radiation passes for these BBH systems (see, e.g., [50]); thus, the difference terms in Eqs. (2.52) and (2.55) will vanish after the passage of the GWs. This implies that the net change in the angular momentum between two nonradiative regions for these binaries will be the same. Nevertheless, while the binary is emitting GWs, the instantaneous value of the angular momentum will differ from the Wald-Zoupas value. We compute the size of this effect in the post-Newtonian (PN) approximation and using surrogate models fit to numerical-relativity (NR) simulations in the following subsections.

We then perform similar calculations involving the difference terms from the super angular momentum of [74]. Because the super angular momentum terms in Eq. (2.48a) involve products of U_{lm} moments, then the super angular momentum can differ from the $\alpha = \beta = 1$ values when there is the GW memory effect. We thus estimate the magnitude of this difference in the PN approximation and from the dominant waveform modes from NR simulations. As we will discuss further below, the effect is small compared to the change in the super angular momentum, but is within the numerical accuracy of the simulations.

Because we are interested in investigating the order-of-magnitudes of the effects rather than their precise values, we will generally work with the leading-order approximations to the results in this section, as we will describe in more detail in the relevant parts below.

2.6.1 Computing the leading GW memory effect and spin memory effect

In post-Newtonian theory, the GW memory effect and the spin memory effect have been computed, and the relevant results can be obtained from, e.g., [22] or [25], respectively. For NR simulations, GW memory effects are not captured in most Cauchy simulations (see, e.g., [85]) and the additional post-processing step of Cauchy-characteristic extraction [48] needs to be performed [31,50] to get the memory effects directly from simulations. However, by enforcing the flux balance laws in Eq. (2.10), one can determine constraints on the GW memory effects from waveforms that do not contain the memory (e.g., [25,86]). This approximate procedure is quite accurate [50]. We summarize our procedure for computing GW memory effects below.

2.6.1.1 (Displacement) GW memory effect

The GW memory effect can be computed by integrating the conservation equation for the Bondi mass aspect in Eq. (2.5a) with respect to u [this equation contains equivalent information to the flux balance law (2.10) for a basis of supertranslation vectors]. Integrating the term $D_A D_B N^{AB}$ in Eq. (2.5a) with respect to u gives rise to a change in the shear, which we will denote by $D_A D_B \Delta C^{AB}$. This quantity $D_A D_B \Delta C^{AB}$ is constrained by changes in the mass aspect Δm and the integrated flux of energy per solid angle (a term proportional to $\int du N_{AB} N^{AB}$; see, e.g., [24] and references therein for further discussion). This equation constrains only the electric part of ΔC_{AB}, and for this reason it is convenient to write the memory using a single scalar function $\Delta\Phi$ as

$$\Delta C_{AB} = \left(D_A D_B - \frac{1}{2} h_{AB} D^2 \right) \Delta\Phi \,. \tag{2.56}$$

It is then useful to expand $\Delta\Phi$ in scalar spherical harmonics Y_{lm}. Once this is done, when the the operator $(2D_A D_B - h_{AB} D^2)$ acts on these scalar harmonics, Eq. (2.56) can be written in terms of the electric-parity tensor harmonics in Eq. (2.36a) as

$$\Delta C_{AB} = \sum_{l,m} \sqrt{\frac{(l+2)!}{2(l-2)!}} T^{(e),lm}_{AB} \Delta\Phi_{lm} \,. \tag{2.57}$$

By comparing Eq. (2.57) with Eq. (2.37), it is straightforward to see that the change in the U_{lm} moments can be related to the $\Delta\Phi_{lm}$ modes via the relationship

$$\Delta U_{lm} = \sqrt{\frac{(l+2)!}{2(l-2)!}} \Delta\Phi_{lm} \,. \tag{2.58}$$

Although both changes in the Bondi mass aspect and the flux of energy per solid angle produce GW memory effects, for nonprecessing BBH mergers, the flux term produces the much larger memory effect (i.e., the nonlinear memory is much larger than the linear memory; this is true in both the post-Newtonian approximation [87] and in NR simulations [50]). For this reason, just the contributions from the nonlinear memory to $\Delta\Phi$ were computed in [25], and the result is given in terms of the mass and current multipole moments by

$$\Delta\Phi_{lm} = \frac{1}{2}\frac{(l-2)!}{(l+2)!}\sum_{l',l'',m',m''} C_l(-2,l',m';2,l'',m'') \\ \times \int_{-\infty}^{\infty} du \{2is^{l,(-)}_{l';l''}\dot{U}_{l'm'}\dot{V}_{l''m''} + s^{l,(+)}_{l';l''}(\dot{U}_{l'm'}\dot{U}_{l'',m''} + \dot{V}_{l'm'}\dot{V}_{l'',m''})\}\,. \tag{2.59}$$

Both in the PN approximation and in NR simulations, the largest contribution to the GW memory effect from nonprecessing BBH mergers comes from terms involving products of U_{22} and $U_{2-2} = \bar{U}_{22}$ modes in Eq. (2.59). The dominant memory effect produced by the U_{22} mode appears in the $\Delta\Phi_{20}$ and $\Delta\Phi_{40}$ modes. Evaluating the appropriate coefficients in Eq. (2.59) and using Eq. (2.58), we find that the leading GW memory effect in the mode U_{20} is given by

$$\Delta U_{20} = \frac{1}{42}\sqrt{\frac{15}{\pi}}\int_{-\infty}^{\infty} du |\dot{U}_{22}|^2\,. \tag{2.60a}$$

The expression for the U_{40} mode is given by

$$\Delta U_{40} = \frac{1}{504\sqrt{5\pi}}\int_{-\infty}^{\infty} du |\dot{U}_{22}|^2 = \frac{1}{60\sqrt{3}}\Delta U_{20}\,. \tag{2.60b}$$

We will also consider quantities U_{20} and U_{40} which are obtained by integrating Eq. (2.60) from $-\infty$ up to a finite retarded time u rather than taking the limit $u \to \infty$.

2.6.1.2 GW modes that produce the spin-memory effect

The other type of GW memory that we will need to consider in this paper is the GW spin memory effect. Like the GW memory effect in the previous subsection, the spin memory effect can also be determined from the flux balance law in Eq. (2.10). Unlike the displacement memory, the spin memory is constrained by changes in the super angular momentum (rather than the supermomentum) and the flux of angular momentum per solid angle (rather than the flux of energy per solid angle). In addition, the spin memory effect appears in the magnetic-parity part of the retarded-time integral of the shear tensor, rather than the electric part of the change in the shear. We will not need the spin memory itself, but we do need the GW modes that produce the spin memory effect. Nevertheless, it is easiest to describe the calculation of these modes by summarizing the calculation of the spin memory. We thus begin by writing the shear tensor C_{AB} as a sum of two terms of electric- and magnetic-parity parts

$$C_{AB} = \frac{1}{2}\left(2D_A D_B - h_{AB}D^2\right)\Phi + \epsilon_{C(A}D_{B)}D^C\Psi\,, \tag{2.61}$$

where Φ and Ψ are smooth functions of the coordinates (u, θ^A). The spin memory is related to the retarded time integral of the function Ψ [25]

$$\Delta\Sigma \equiv \int_{-\infty}^{+\infty} du\,\Psi\,. \tag{2.62}$$

The full multipolar expansion of the spin memory is a somewhat lengthy expression, so we do not reproduce it here (although it is given in [25]). Analogously to the displacement GW memory effect, there are two contributions to the spin memory effect from the linear and nonlinear terms. However, the linear terms are smaller than the nonlinear terms for nonprecessing compact binaries (see, e.g., [50]), so we focus on just the nonlinear terms. We will also give just the largest terms that are computed from the mode U_{22} (which is the dominant term in the PN approximation, and also the most significant term in NR simulations). The U_{22} mode produces a spin memory effect that appears in the u integral of the $l = 3$, $m = 0$ mode of the waveform; it was computed in [25] to be

$$\Delta\Sigma = \frac{1}{80\sqrt{7\pi}} Y_{30} \int du \Im(\bar{U}_{22}\dot{U}_{22}) \,. \tag{2.63}$$

Acting on $\Delta\Sigma$ with the operator $\epsilon_{C(A}D_{B)}D^C$ gives the retarded-time integral of the magnetic-parity part of the shear tensor C_{AB}:

$$\epsilon_{C(A}D_{B)}D^C\Delta\Sigma = \frac{1}{40}\sqrt{\frac{15}{7\pi}} T^{(b),30}_{AB} \int du \Im(\bar{U}_{22}\dot{U}_{22}) \,. \tag{2.64}$$

By differentiating Eq. (2.64) with respect to u, we can obtain the magnetic part of the shear that produces the spin memory effect. Because Eq. (2.64) is already expanded in magnetic-parity tensor harmonics, we can immediately determine that the relevant spin-memory mode is V_{30}, which is given by

$$V_{30} = \frac{1}{40}\sqrt{\frac{15}{7\pi}} \Im(\bar{U}_{22}\dot{U}_{22}) \,. \tag{2.65}$$

We will use Eqs. (2.60) and (2.65) to add in the contributions of the memory and spin memory effects that are not included in the NR surrogate waveform model that we use to compute the difference terms from the respective Hamiltonian definitions of [14] and of [74] for the angular momentum and super angular momentum in the next subsections.

2.6.2 Standard angular momentum

We noted above that the different definitions of the angular momentum for non-precessing BBH mergers will agree after the gravitational waves pass, but they will differ while these systems are radiating gravitational waves. We will calculate the size of this difference first in the post-Newtonian (PN) approximation and second in full general relativity using numerical-relativity waveforms from BBH mergers. The NR waveforms are usually given in terms of the multipole moments of the strain h, which is related to the tensor C_{AB} by the relation

$$h \equiv h_+ - ih_\times = \frac{1}{r} C_{AB} \bar{m}^A \bar{m}^B \,. \tag{2.66}$$

This expression defines the two polarizations h_+ and $h_\times$ and $\bar{m}^A$ is the complex conjugate of the dyad defined in Eq. (2.40). The strain h can be expanded in terms of spin-weighted spherical harmonics ${}_{-2}Y_{lm}$ as

$$h = \sum_{lm} h_{lm} \left({}_{-2}Y_{lm} \right) \,. \tag{2.67}$$

It then follows that the moments h_{lm} are related to U_{lm} and V_{lm} by

$$h_{lm} = \frac{1}{r\sqrt{2}} \left(U_{lm} - iV_{lm} \right) \tag{2.68}$$

(see, e.g., [25] and references therein).

Because of the symmetries of nonprecessing binaries, the relationship between the h_{lm} mode and the U_{lm} and V_{lm} modes simplifies. Specifically, the mass multipole moments U_{lm} are nonzero only when $l+m$ is even, and the current multipole moments V_{lm} are nonzero only when $l+m$ is odd (see, e.g., [15]). Therefore, the mass and current multipole moments can be written in terms of the strain modes for these systems as

$$U_{lm} = r\sqrt{2}h_{lm}\,, \quad \text{for } l+m \text{ even}\,, \tag{2.69a}$$

$$V_{lm} = ir\sqrt{2}h_{lm}\,, \quad \text{for } l+m \text{ odd}\,. \tag{2.69b}$$

Note that our definition of the polarizations h_+ and $h_\times$ (and hence h_{lm}) have a relative minus sign to those in [15], though the U_{lm} and V_{lm} moments agree in sign. Combining these properties of the U_{lm} and V_{lm} moments with the expressions for the difference terms in Eqs. (2.52) and (2.55), we find that multipole moments $\delta J_{1\pm1}^{(\alpha=\beta=1)}$ and $\delta k_{10}^{(\alpha=\beta=1)}$ vanish. Thus, we focus on the $\delta J_{10}^{(\alpha=\beta=1)}$ and $\delta k_{1\pm1}^{(\alpha=\beta=1)}$ modes below.

The waveforms from PN calculations and surrogate models from NR simulations contain a finite number of (l, m) modes [in the PN context, the waveform has only been computed up to a finite PN order, whereas for surrogate models, the NR simulations extract only a subset of all (l, m) modes, and the surrogate models only fit to a further subset of the extracted modes]. The number of modes that we use in the calculations of the quantities $\delta J_{10}^{(\alpha=\beta=1)}$ and $\delta k_{1\pm1}^{(\alpha=\beta=1)}$ will differ, but it is chosen such that we capture the leading nonvanishing effect in the PN approximation. We will then use the same set of modes for the calculations with the NR surrogate waveform (absent any modes that the surrogate model does not contain). As we will discuss in more detail below, we will use waveform modes that go up to 2.5PN orders above the leading part of the U_{22} mode to compute $\delta J_{10}^{(\alpha=\beta=1)}$, whereas for $\delta k_{1\pm1}^{(\alpha=\beta=1)}$, we can capture the leading effect using just the leading U_{22} mode and the V_{21} mode. Thus, to compute $\delta J_{10}^{(\alpha=\beta=1)}$ we use the expression

$$\begin{aligned}\delta J_{10}^{(\alpha=\beta=1)} = \frac{1}{8}\sqrt{\frac{3}{2\pi}}\Re\Bigg[&\frac{a_2 c_{22}}{3}\bar{U}_{22}V_{32} + \frac{a_3 c_{33}}{4}\bar{U}_{33}V_{43} + \frac{a_3 c_{31}}{4}\bar{U}_{31}V_{41} + \frac{a_2 c_{20}}{6}\bar{U}_{20}V_{30} \\ &- \frac{a_2 c_{21}}{3}\bar{V}_{21}U_{31} - \frac{a_3 c_{32}}{4}\bar{V}_{32}U_{42} - \frac{a_4 c_{43}}{5}\bar{V}_{43}U_{53} - \frac{a_3 c_{30}}{8}\bar{V}_{30}U_{40}\Bigg],\end{aligned} \tag{2.70}$$

Note that the real part of the quantity in parentheses is being taken, which arises from using the complex-conjugate properties of the modes U_{lm} and V_{lm} in Eq. (2.38). For $\delta k_{1\pm1}^{(\alpha=\beta=1)}$, we use the expressions

$$\delta k_{11}^{(\alpha=\beta=1)} = \frac{i}{96}\sqrt{\frac{3}{\pi}}\left(d_{2-2}^{(+)}U_{22}\bar{V}_{21} - d_{20}^{(+)}\bar{U}_{20}V_{21}\right), \tag{2.71a}$$

$$\delta k_{1-1}^{(\alpha=\beta=1)} = \frac{i}{96}\sqrt{\frac{3}{\pi}}\left(d_{22}^{(-)}\bar{U}_{22}V_{21} - d_{20}^{(-)}U_{20}\bar{V}_{21}\right). \tag{2.71b}$$

Here note that $\delta k_{11}^{(\alpha=\beta=1)} = -\delta\bar{k}_{1-1}^{(\alpha=\beta=1)}$, since $\delta k_{11}^{(\alpha=\beta=1)}$ and $\delta k_{1-1}^{(\alpha=\beta=1)}$ can both be related to the real difference terms from the x and y components of the Wald-Zoupas CM angular momentum (see Appendix A).

2.6.2.1 Post-Newtonian results

For nonprecessing binaries, the mass and current multipole moments U_{lm} and V_{lm} are expressed conveniently in terms of several different mass parameters and mass ratios. Here we denote the individual masses by m_1 and m_2 with $m_1 > m_2$. We then denote the total mass by $M = m_1 + m_2$, the relative mass difference by $m_{12} = (m_1 - m_2)/M$, the mass ratio by $q = m_1/m_2 \geq 1$, and the symmetric mass ratio $\nu = m_1 m_2/M^2$. We also use the notation Ω for the orbital frequency, ψ for the orbital phase, and $x = (M\Omega)^{2/3}$ for the PN parameter, as in [15]. It is shown in [15] that all the waveform modes h_{lm} can be written in the form

$$h_{lm} = -\frac{8M\nu x}{r}\sqrt{\frac{\pi}{5}}\mathcal{H}_{lm}e^{-im\psi}\,, \tag{2.72}$$

where the terms $\mathcal{H}_{lm}$ are given in Eqs. (328)–(329) of [15] and can be written as polynomials in the square root of the PN parameter (i.e., $\sqrt{x}$). We do not use the full expressions for $\mathcal{H}_{lm}$ in Eqs. (328)–(329) of [15]; rather we only go up to 2.5PN order (i.e., $x^{5/2}$) in these equations. After substituting these expressions into Eq. (2.70), we find that the result for $\delta J_{1,0}^{\alpha=\beta=1}$ is given by

$$\delta J_{10}^{(\alpha=\beta=1)} = \frac{8}{5}\sqrt{\frac{3\pi}{2}}M^2\nu^2\left(-\frac{10}{21} - \frac{m_{12}^2}{210} + \frac{9329}{4410}\nu\right)x^{9/2} + O(x^5)\,. \tag{2.73}$$

The angular momentum in the Newtonian limit goes as $x^{-1/2}$, so the correction term in Eq. (2.73) appears at 5PN order with respect to the leading-order effect. During the inspiral when the PN parameter x is small, $\delta J_{10}^{\alpha=\beta=1}$ is not expected to be very large.

Given the fact that the product $\bar{U}_{22}V_{32}$ scales with the PN parameter as x^3, it might initially seem unusual that the net effect $\delta J_{10}^{\alpha=\beta=1}$ goes like $x^{9/2}$. Because there is a real part in Eq. (2.70), there are a number of cancellations that occur between different modes. These cancellations in the U_{lm} and V_{lm} moments occur in the conservative part of the dynamics, but not the dissipative part from GW radiation reaction. These dissipative dynamics appear as a relative 1.5PN correction to V_{32}, which explains why the leading order part of $\delta J_{10}^{\alpha=\beta=1}$ goes like $x^{9/2}$. Analogous arguments can be made for the other terms in Eq. (2.70).

There is another feature of Eq. (2.73) worth describing that relates to the dependence of $\delta J_{10}^{\alpha=\beta=1}$ on the mass ratio q (and which is a feature that also appears in the NR simulations, which we discuss later). Specifically, the sign of $\delta J_{10}^{\alpha=\beta=1}$ changes, and there is a specific mass ratio at which the leading PN expression vanishes. The value of the mass ratio can be computed from Eq. (2.73) to be $q \approx 1.9$. The physical reason for this value was less clear to us, though it arises from the change in amplitudes of the multipole moments U_{lm} and V_{lm} as a function of mass ratio q.

The leading-order contribution to $\delta k_{1\pm1}^{(\alpha=\beta=1)}$ turns out to require fewer terms to compute, as indicated in Eq. (2.71), and it only requires the leading-order parts of the moments U_{22} and V_{21}. It is reasonably straightforward to show that $\delta k_{1\pm1}^{(\alpha=\beta=1)}$ is given by

$$\delta k_{1,\pm1}^{(\alpha=\beta=1)} = -i\frac{22}{35}\sqrt{\frac{\pi}{3}}M^2\nu^2 m_{12}x^{5/2}e^{\mp i\psi} + O(x^3)\,. \tag{2.74}$$

The difference term from the Wald-Zoupas definition of the CM angular momentum scales as $x^{5/2}$, which is two PN orders lower than the correction term to the intrinsic angular momentum. However, this effect also goes as $e^{\mp i\psi}$, so the average over an orbital period vanishes. As was discussed in [26], while the change in the Wald-Zoupas definition of the CM angular momentum scales with the PN parameter as $x^0 = O(1)$, there is a choice of reference time u_0 that can set the change in the CM angular momentum to zero through 2PN order (i.e., through x^2). At 2.5PN order ($x^{5/2}$), there is no longer just a choice of reference time that allows the effect to be set to zero, which also preserves the fact that the binary was initially chosen to be in the CM frame and rest-frame of the source with the supertranslations chosen such that $C_{AB} = 0$ initially. Thus, the terms $\delta k_{1\pm1}^{(\alpha=\beta=1)}$ in Eq. (2.74) are of the same PN order as the nontrivial (in the sense discussed here) Wald-Zoupas CM angular momentum. The impact of the different definitions of angular momentum is thus largest for the CM angular momentum (although the impact of the CM angular momentum on the evolution of compact binaries has not been discussed as extensively as that of the other charges associated with the Poincaré group).

Finally, we also point out that from Eq. (2.74) it can be shown that the maximum effect happens approximately at $q = 2.6$. This is comparable to the value of the mass ratio that results in the maximum kick velocity for nonspinning binaries ($q = 2.8 \pm 0.23$) [88].

2.6.2.2 Results from NR surrogate models

While the PN approximation gives useful intuition about the effect of the remaining free parameter α on the intrinsic and CM angular momentum during the inspiral phase of a compact binary, it is not expected to be accurate during the merger and ringdown phases. Instead, it is preferable to use the results of NR simulations during these late stages of a BBH merger. In particular, we will use the hybrid NR surrogate model NRHyb3dq8 [47] to generate the waveform modes that enter into Eqs. (2.70) and (2.71). The surrogate produces the waveform modes rh_{lm}/M, which we convert to the U_{lm} and V_{lm} moments using Eq. (2.69). Because the surrogate does not model the modes h_{40}, h_{41} and h_{53}, we cannot include the surrogate model's contribution to these modes in Eq. (2.70). Also, because the surrogate does not have the memory or spin memory contributions to the modes h_{20}, h_{30}, and h_{40}, we add these contributions to those of the surrogate model. The procedure we use to compute these memory modes is reviewed in Sec. 2.6.1.

For presenting our results from the surrogate waveforms, we opt to show the Cartesian components of the intrinsic or CM angular momentum instead of the multipole moments that were described in the previous parts. The conversion between these two descriptions is reasonably straightforward and is described in further detail in Appendix A. We thus quote the results here. First, the z component for the intrinsic angular momentum $\delta J_z^{(\alpha=\beta=1)}$ can be related to $\delta J_{10}^{(\alpha=\beta=1)}$ by

$$\delta J_z^{(\alpha=\beta=1)} = -2\sqrt{\frac{2\pi}{3}}\delta J_{10}^{(\alpha=\beta=1)} \,. \tag{2.75}$$

Similarly, $\delta k_x^{(\alpha=\beta=1)}$ and $\delta k_y^{(\alpha=\beta=1)}$ can be related to $\delta k_{1\pm1}^{(\alpha=\beta=1)}$ by

$$\delta k_x^{(\alpha=\beta=1)} = -4\sqrt{\frac{\pi}{3}}\Re\left[\delta k_{11}^{(\alpha=\beta=1)}\right] \,, \tag{2.76a}$$

$$\delta k_y^{(\alpha=\beta=1)} = 4\sqrt{\frac{\pi}{3}}\Im\left[\delta k_{11}^{(\alpha=\beta=1)}\right] \tag{2.76b}$$

(see also [26]). Because $\delta k_z^{(\alpha=\beta=1)}$ is proportional to $\delta k_{10}^{(\alpha=\beta=1)} = 0$ for nonspinning BBHs, then the magnitude of the difference of the CM angular momentum is given by

$$|\delta \mathbf{k}^{(\alpha=\beta=1)}| = \sqrt{\left(\delta k_x^{(\alpha=\beta=1)}\right)^2 + \left(\delta k_y^{(\alpha=\beta=1)}\right)^2} \,. \tag{2.77}$$

We first show the difference of the intrinsic angular momentum from the Wald-Zoupas value, $\delta J_z^{(\alpha=\beta=1)}$, for BBHs with different mass ratios. The top panel of Fig. 2.1 displays $\delta J_z^{(\alpha=\beta=1)}$ as a function of retarded time for three different mass ratios, $q = 1$, 2, and 4 as solid blue, orange dashed, and green dotted curves, respectively. The extreme values of the time series for $\delta J_z^{(\alpha=\beta=1)}$ approach the largest positive, the closest to zero, and the most negative value for these three mass ratios, respectively. The dependence of the extreme value of $\delta J_z^{(\alpha=\beta=1)}$ as a function of mass ratio is illustrated in more detail in the bottom panel of Fig. 2.1. As was noted in the discussion of $\delta J_z^{(\alpha=\beta=1)}$ in the PN approximation, the extreme value of this quantity changes sign as a function of mass ratio. The value at which it undergoes this sign change for the surrogate model is $q \approx 2.2$, which is close to the value predicted by the leading PN result of $q \approx 1.9$. There is a sharp feature in the curve near the mass ratio where $\delta J_z^{(\alpha=\beta=1)}$ goes to zero, because (what is for most mass ratios) the primary peak (which changes smoothly with mass ratio) becomes smaller than (what is for most mass ratios) the secondary peak (which also varies smoothly with mass ratio, but at a different rate from the primary peak). When the roles of primary and secondary peak reverse for a small range of mass ratios, the slope changes abruptly, and this leads to this slight sharp feature.

We also mention a few implications of the results presented in Fig. 2.1. During the inspiral, the Newtonian value of the orbital angular momentum is given by $M^2\nu x^{-1/2}$. For an equal mass binary separated by a distance of $100M$, the angular momentum will initially be of order $\sim 2.5M^2$. The final black hole is a Kerr black hole with spin of order $\sim 0.67M_\mathrm{f}^2$, where M_f is the final mass of the black hole (which is typically at least ninety percent of the total mass M). Thus, the fact that $\delta J_z^{(\alpha=\beta=1)}$ is of order a few times $10^{-4}M^2$ at its largest implies that the discrepancies in the definitions of angular momentum will be small for definitions where α is of order unity. However, the final spin parameter of the black hole formed from a BBH merger is often quoted to an accuracy which is smaller than the values of $\delta J_z^{(\alpha=\beta=1)}$ described here (see, e.g., [85]). Thus, for completeness, NR simulations should specify which definition of angular momentum is being used.

We now turn to the difference of the CM angular momentum from the Wald-

Zoupas value. We use the same surrogate model to compute $\delta k_x^{(\alpha=\beta=1)}$ and $|\delta\mathbf{k}^{(\alpha=\beta=1)}|$ as functions of retarded time. We plot these quantities in the top panel of Fig. 2.2 for $q = 3$. The bottom panel of Fig. 2.2 shows the peak value of the time series $|\delta\mathbf{k}^{(\alpha=\beta=1)}|$ as a function of the binary's mass ratio, q. For an equal mass black-hole binary, $q = 1$, the change in the CM angular momentum vanishes. This occurs because there is no linear momentum radiated from such a system, so the initial and final rest frames are the same (and we have chosen the initial rest frame to be the CM frame). The peak value of $|\delta\mathbf{k}^{(\alpha=\beta=1)}|$ is reached at a mass ratio of roughly $q \approx 2.5$. This is similar to the PN prediction of $q \approx 2.6$ computed earlier. It is also near the peak value of the gravitational recoil computed in [88] of $q \approx 2.8$. The decrease in the magnitude of $|\delta\mathbf{k}^{(\alpha=\beta=1)}|$ at mass ratios greater than $q \sim 2.5$ is likely related to the fact that the gravitational recoil also decreases at these larger mass ratios.

As far as we are aware, there has not been a systematic study of the size Wald-Zoupas CM angular momentum from numerical relativity simulations. In the PN approximation, the calculations in [26], which were reviewed in this subsection, suggest that the magnitude of the Wald-Zoupas CM angular momentum, $|\mathbf{k}^{(\alpha=\beta=1)}|$, goes as $M^2 x^{5/2}$. Thus, the magnitude of the CM angular momentum could be as large as order M^2 near the merger (thereby making the difference $|\delta\mathbf{k}^{(\alpha=\beta=1)}|$ a small effect). Further investigation is needed to have a more definitive statement about the possible importance of the term $|\delta\mathbf{k}^{(\alpha=\beta=1)}|$.

2.6.3 Super angular momentum

We now turn to understanding effect of the free parameter α ($= \beta$) on the difference of the super angular momentum from the charge of [74] for nonspinning BBH mergers. Unlike the angular momentum, the super angular momentum can have a nontrivial net change between the early- and late-time nonradiative regions of a spacetime for these systems. We thus focus on the net change in the charges $\Delta Q_Y^{\alpha=\beta}$: namely, the difference of Eq. (2.34a) between two nonradiative regions at early and late times. Thus, we will similarly be interested in the change in the difference term from the $\alpha = \beta = 1$ value of the charges; i.e., the quantity $\Delta\delta Q_Y^{\alpha=\beta=1}$, where $\delta Q_Y^{\alpha=\beta=1}$ is defined in Eq. (2.34b).

We now calculate the change in the largest (in magnitude) nonvanishing part of the super angular momentum, which appears in the $l = 2$, $m = 0$ moments of the super-CM part (in both the PN approximation and from NR simulations). First, we write the expression for this change in the charges as

$$\Delta Q^{(\alpha=\beta)}_{(e),20} = \Delta Q^{(\alpha=\beta=1)}_{(e),20} + (\alpha - 1)\Delta\delta Q^{(\alpha=\beta=1)}_{(e),20} \,. \tag{2.78}$$

The change in the term $\delta Q^{(\alpha=\beta=1)}_{(e),20}$ can be obtained by taking the difference of Eq. (2.49a) evaluated at early and late times. For nonspinning binaries, all the V_{lm} moments vanish in nonradiative regions; the change in the moments U_{lm} can be nonvanishing in nonradiative regions when there is a nontrivial GW memory effect. The largest moments are U_{20} and U_{40}, as described in Sec. 2.6.1; however, because the mode U_{40} is a factor of $60\sqrt{3}$ times smaller than the U_{20} mode, we focus here on the contribution from just U_{20}. We find that the leading change in the difference term is given by

$$\Delta\delta Q^{(\alpha=\beta=1)}_{(e),20} = \frac{3}{448\pi}\sqrt{\frac{15}{2\pi}}\Delta(U_{20})^2\,. \tag{2.79}$$

Finally, we will compute $\Delta Q^{(\alpha=\beta=1)}_{(e),20}$. The term quadratic in C_{AB} in Eq. (2.19) gives rise to a term quadratic in ΔU_{20} which is identical to the expression for $\Delta\delta Q^{(\alpha=\beta=1)}_{(e),20}$ in Eq. (2.79). The term linear in the shear does not contribute (because it involves only V_{lm} modes) and the term $-uD_A m$ does not have a contribution from nonspinning BBH mergers to this part of the charge. However, the term involving N_A in Eq. (2.19) does contribute to $\Delta Q^{(\alpha=\beta=1)}_{(e),20}$. The form of N_A is known in stationary regions that are supertranslated from the canonical frame in which $C_{AB} = 0$. It was shown in [24] that $N_A = -3mD_A\Phi/2$, where Φ is the "potential" for the electric part of the shear [as in Eq. (2.20c)], and the Bondi mass aspect m is a constant in this frame. Using the fact that $\Delta U_{20} = \sqrt{12}\Delta\Phi_{20}$, we then find that the leading $\alpha = \beta = 1$ super CM is given by

$$\Delta Q^{(\alpha=\beta)}_{(e),20} = \frac{-3}{16\pi}\frac{M}{\sqrt{2}}\Delta U_{20} + \frac{3}{448\pi}\sqrt{\frac{15}{2\pi}}\Delta(U_{20})^2\,. \tag{2.80}$$

The lowest multipole moment (consistent with the symmetries of nonprecessing BBHs) in which the change in the superspin part could appear is the $l = 3$, $m = 0$ mode. When we evaluate the contribution of the U_{20} modes in Eq. (2.48b) for $l = 3, m = 0$, we find it and the difference from the Hamiltonian charge of [74] both vanish:

$$\Delta Q^{(\alpha=\beta=1)}_{(b),30} = \Delta\delta Q^{(\alpha=\beta=1)}_{(b),30} = 0\,. \tag{2.81}$$

Note, however, that the instantaneous value of the charges (not the change in a nonradiative-to-nonradiative transition) can be nonvanishing, though we do not compute that quantity here. We next turn to the computation of the super CM using the PN approximation and the NR surrogate model discussed in the previous subsection.

PN approximation We calculate the U_{20} waveform modes associated with the GW memory effect as was described in Sec. 2.6.1. Because the PN approximation covers only the inspiral, we truncate the calculation of $\Delta U_{20}^{(\alpha=\beta=1)}$ at a finite retarded time u, at which the binary is at a PN parameter x. We thus denote the change in the PN parameter by Δx. This gives an expression for the U_{20} moment that is equivalent to the one given in [15]. We thus find that the change in the super-CM angular momentum in Eq. (2.80) and the change in the difference in Eq. (2.79) are given by

$$\Delta Q_{(e),20}^{(\alpha=\beta)} = -\frac{1}{28}\sqrt{\frac{15}{2\pi}}M^2\nu\Delta x + \frac{5}{1372}\sqrt{\frac{15}{2\pi}}M^2\nu^2\Delta(x^2)\,, \tag{2.82a}$$

$$\Delta\delta Q_{(e),20}^{(\alpha=\beta=1)} = \frac{5}{1372}\sqrt{\frac{15}{2\pi}}M^2\nu^2\Delta(x^2)\,. \tag{2.82b}$$

Thus, the different definitions of the super-CM angular momentum causes a relative 1PN-order correction to the leading-order super-CM angular momentum.

Numerical-relativity results The GW memory effect is largest not during the inspiral, but after the merger and ringdown of a BBH collision. To better understand the size of the change in the super-CM angular momentum of a BBH merger, we compute the full memory effect in the U_{20} mode as in Eq. (2.60a), and we substitute the result into Eqs. (2.79) and (2.80). We again consider nonspinning BBH mergers of different mass ratios, and we use the same hybrid surrogate model NRHybSur3dq8 [47] to compute ΔU_{20}. We take the mass M that enters into Eq. (2.80) to be the final mass, which we compute using the NR fits computed in [53].

In Fig. 2.3, we show the net change in difference in the super-CM angular momentum from the Hamiltonian super-CM angular momentum of [74], as a function of the mass ratio of nonspinning BBH mergers of different mass ratios between $1 \leq q \leq 8$. The maximum difference occurs for equal-mass BBHs and decreases with higher mass ratios, which is consistent with the amplitude of the memory effect computed from the dominant quadrupole modes, as in Eq. (2.60a). This figure illustrates that the change in the difference terms of the leading super-CM angular momentum are about one hundredth of the change in the super-CM of [74], which is itself a small effect in units of M^2. Nevertheless, the waveform modes used to compute the result are sufficiently accurate that this difference can be resolved.

2.7 Conclusions

In this paper, we investigated the freedoms in defining angular momentum and super angular momentum in asymptotically flat spacetimes and the implications of these freedoms on the values of the (super) angular momentum of nonspinning binary-black-hole mergers. The fact that such freedoms exist was recently discussed in [69], which demonstrated that there can be a two (real) parameter family of angular momenta, which encompass a few commonly used definitions of angular momentum in asymptotically flat spacetimes. All members of this two-parameter family satisfy flux balance laws and are constructed from quantities that are covariant with respect to 2-sphere cross sections of null infinity. We found, however, that for the angular momentum to vanish in flat spacetime, the two parameters must be equal; this leads to a natural requirement that the family of angular momenta should depend upon only a single real parameter. If we do not require that the angular momentum agree with the Hamiltonian definition of Wald and Zoupas, then there remained a one-parameter family of angular momentum.

We further investigated the effect of this one free parameter on the values of the angular momentum. To do so, we first derived a multipolar expansion (in terms of the radiative multipole moments of the GW strain) of the difference of the angular momentum from the Wald-Zoupas definition. The difference is constructed from the products of mass moments with current moments, unlike the flux of the Wald-Zoupas definition of angular momentum, which is written in terms of products of mass moments with themselves and current moments with themselves. This fact has an important implication for spacetimes that transition between nonradiative regions at early times and at late times, the context in which the GW memory effect is usually computed. For several types of systems of astrophysical interest, such as compact-object mergers, the GW memory effect appears in just the mass-type moments. Thus, the difference terms that arise from products of mass and current moments will vanish in these nonradiative-to-nonradiative transitions, and the net change in the angular momentum will be independent of this remaining free parameter. There will, however, be a difference in the instantaneous value of the angular momentum while the system is radiating gravitational waves.

We also proposed considering a two-parameter family of super angular momentum in analogy with the two-parameter family of angular momentum given in [69]. Choosing the two parameters to be equal does not generically make the super angular momentum vanish in flat spacetime (and it has also been argued that the super angular momentum should not necessarily vanish in this context). There is a choice of the two parameters that does manifestly make the super angular momentum vanish in flat spacetime, but it does not correspond to the analog of the Wald-Zoupas charge. We, therefore, derived a multipolar expansion of the difference in the super angular momentum from the Hamiltonian definition of [74] that involved two real parameters. We also specialized the result to have one free parameter, so that the charge reduces to the angular momentum when the symmetry vector field reduces from an infinitesimal super Lorentz transformation to a standard infinitesimal Lorentz transformation.

Next, we investigated the magnitude of the difference of the (super) angular momentum from the Wald-Zoupas charges for nonspinning, quasicircular binary-black-hole mergers. For the standard angular momentum the difference occurs only while the system is radiating GWs. In the post-Newtonian approximation, we found the difference in the intrinsic angular momentum enters at a relative 5PN-order to the

Newtonian angular momentum, while the difference in the CM angular momentum, it appears at the same PN order as the effect that cannot be set to zero through a particular choice of reference time (at 2.5PN order beyond the leading Newtonian expression). Given the high PN orders, the effects will generally be small, although they could become large near the binary's merger, when the PN approximation becomes inaccurate. During the inspiral, however, the difference in the CM angular momentum from the Wald-Zoupas value will be larger than that of the intrinsic angular momentum, because of its lower PN order. For the super angular momentum, the difference terms need not vanish after the radiation passes; thus, we focused on the net change of the charges between early times and late times. We found that the leading difference in the superspin vanishes for BBH mergers, while differences in the super-CM angular momentum cause a relative 1PN difference from the Hamiltonian super-CM angular momentum of [74].

Finally, we estimated the difference terms for the (super) angular momentum using inspiral-merger-ringdown surrogate waveforms of nonspinning BBH mergers that were fit to numerical-relativity simulation data. The intrinsic angular momentum terms are largest at equal mass, change sign at a mass ratio near two, and then take on the most negative value near a mass ratio of four before approaching closer to zero. The amplitude of the effect is small compared to the Newtonian value of the angular momentum. The maximum difference in the CM angular momentum was found to happen approximately at the mass ratio that produces the maximum kick velocity of the final black hole. The difference in the change of the super-CM angular momentum from the corresponding Hamiltonian expression of [74] in a nonradiative-to-nonradiative transition was only to a few percent correction. Although these differences in the (super) angular momentum are small compared to the values of the (super) angular momentum itself, they are able to be resolved for these systems. Thus, which definition is being used should be specified when describing the (super) angular momentum of nonspinning binary-black-hole mergers.

Acknowledgments

A.E. and D.A.N. acknowledge support from the NSF grant PHY-2011784. We thank Alex Grant for helpful discussions about the Wald-Zoupas definition of angular momentum in the covariant conformal approach to null infinity and for comments on the manuscript. We thank Geoffrey Compère and Ali Seraj for correspondence related to their work [69]; we also thank Geoffrey Compère, Adrien Fiorruci, and Romain Ruzziconi for correspondence related to their work [70].

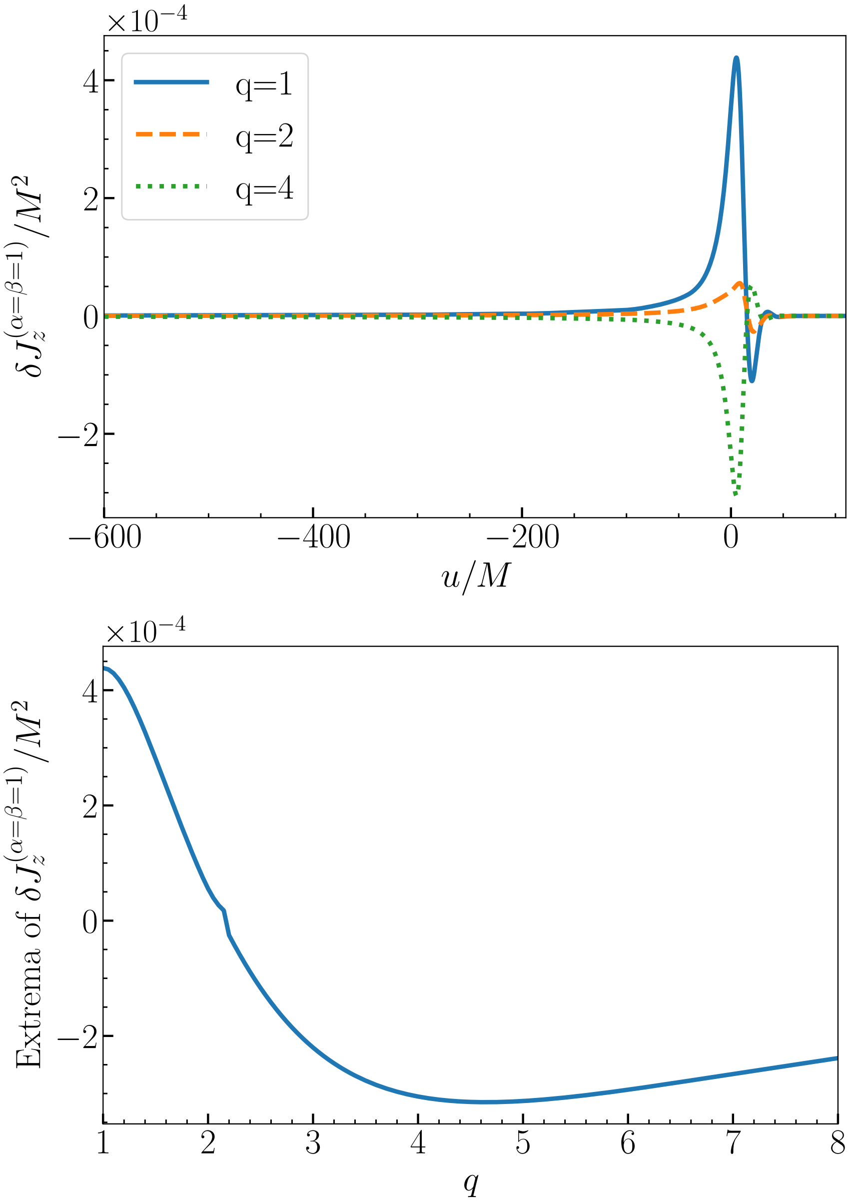

Figure 2.1: *Top*: The z component of the difference of the intrinsic angular momentum from the Wald-Zoupas values (denoted by $\delta J_z^{(\alpha=\beta=1)}$) as a function of retarded time for nonspinning BBH mergers of three mass ratios, $q = 1$, 2, and 4. Note that the extreme value switches from a maximum to a minimum as a function of mass ratio. As discussed further in the text, $\delta J_z^{(\alpha=\beta=1)}$ was computed using a NR surrogate model (where the peak of the magnitude of the waveform is at retarded time equal to zero) using Eqs. (2.70) and (2.75). *Bottom*: The extreme value of the z component of $\delta J_z^{(\alpha=\beta=1)}$ as a function the mass ratio. Consistent with the PN predictions, there is a change in the sign of the quantity $\delta J_z^{(\alpha=\beta=1)}$ that occurs near the mass ratio $q = 2$.

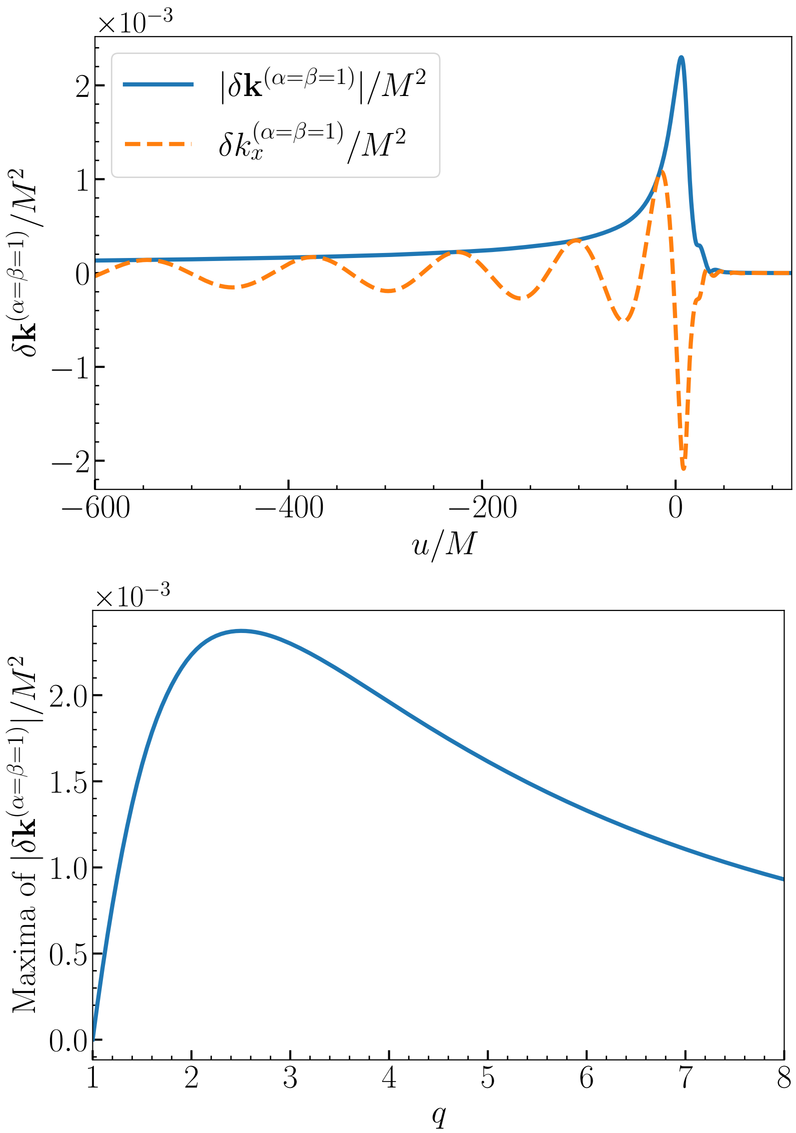

Figure 2.2: *Top*: The magnitude and the x component of the difference of the CM angular momentum from the Wald-Zoupas definition, $|\delta\mathbf{k}^{(\alpha=\beta=1)}|$ and $\delta k_x^{(\alpha=\beta=1)}$, respectively, as functions of retarded time. The system shown is a BBH merger with mass ratio $q = 3$, and the waveform modes used in Eqs. (2.74) and (2.76) were generated from a NR surrogate, where the peak magnitude of the waveform occurs at a time equal to zero. The vector $\delta\mathbf{k}^{(\alpha=\beta=1)}$ is in phase with the orbital motion of the binary during inspiral, and it grows in magnitude until the merger, after which it settles to zero. *Bottom*: The maximum of the magnitude of the difference of the CM angular momentum from the Wald-Zoupas value as a function of the mass ratio of a BBH system. Note that the maximum value as a function of q occurs at roughly the same mass ratio that produces the maximum kick velocity of the final black hole (see the text for further discussion).

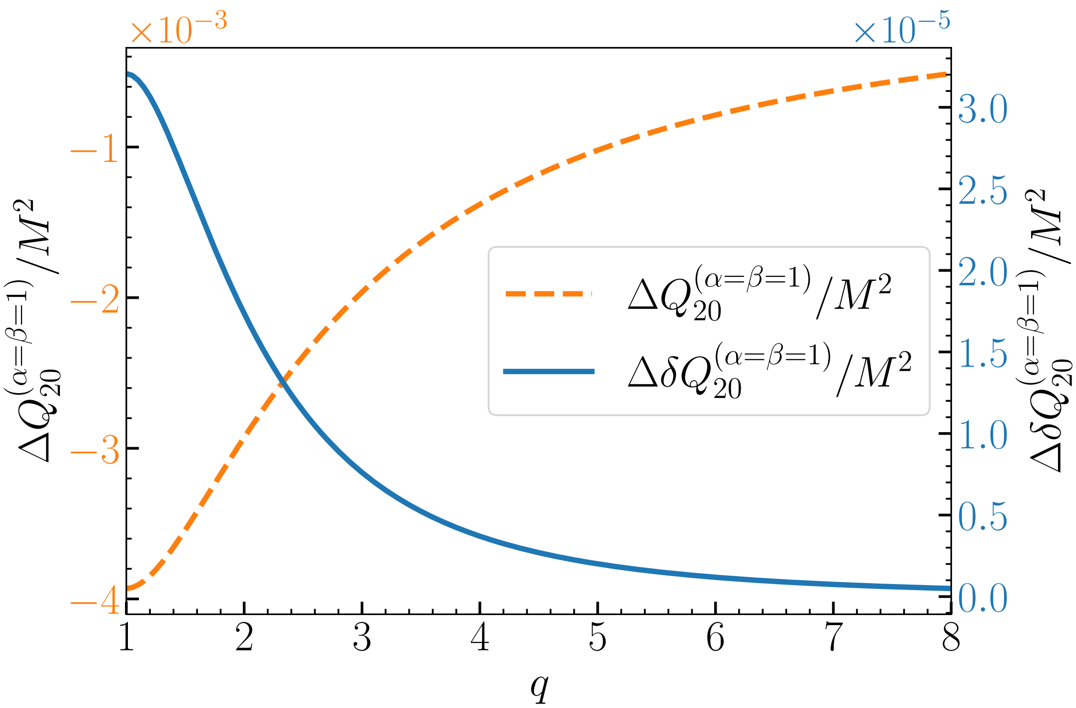

Figure 2.3: The change in the Hamiltonian super-CM angular momentum of [74], $\Delta Q_{2,0}^{(\alpha=\beta=1)}$ (scale on the left), and the change in the difference of the super-CM angular momentum from the Hamiltonian super-CM angular momentum of [74], $\Delta\delta Q_{2,0}^{(\alpha=\beta=1)}$ (scale on the right), both as a function of the mass ratio of the binary q. The difference term is about two orders of magnitude smaller that the change in the super CM.

Chapter 3

Waveform models for the gravitational-wave memory effect: I. Extreme mass-ratio limit and final memory offset

A. Elhashash and D. Nichols, Phys. Rev. D **111**, 044052 (2025),

3.1 Abstract

The gravitational-wave (GW) memory effect is a strong-field relativistic phenomenon that is associated with a persistent change in the GW strain after the passage of a GW. The nonlinear effect arises from interactions of GWs themselves in the wave zone and is an observable effect connected to the infrared properties of general relativity. The detection of the GW memory effect is possible with LIGO and Virgo in a population of binary-black-hole (BBH) mergers or from individual events with next-generation ground- and space-based GW detectors or pulsar timing arrays. Matched-filtering-based searches for the GW memory require accurate, and preferably rapid-to-evaluate waveform models of the memory effect's GW signal. One important element of such

a waveform model is a model for the final memory offset—namely, the net change in strain between early and late times. In this paper, we construct a model for the final memory offset from the merger of nonspinning BBH systems in quasicircular orbits. A novel ingredient of this model is that we first compute the memory signal for extreme mass-ratio inspirals using a high post-Newtonian-order analytic calculation, and we use this analytical result to fix the coefficient in the fit which is linear in the mass-ratio. The resulting memory-offset fit could be used for detecting the GW memory for binaries that merge on a timescale that is short relative to the inverse of the low-frequency cutoff of a GW detector. Additionally, this fit will be useful for analytic waveform models of the GW memory signals in the time and frequency domains.

3.2 Introduction

Gravitational waves (GWs) from the mergers of nearly 100 binary black holes (BBHs) have been detected during the first three observing runs of LIGO, Virgo, and KAGRA [51,52,89]. Based on the rate of detection candidates during the fourth observing run [90], the number of confirmed detections is likely to double. These GW observations have had important implications for both astrophysics and fundamental physics. The increasing number of detections has led to a more precise characterization of the population of BBH mergers, including the distributions of the masses and spins of the individual BHs and the overall merger rate [91–93]. BBH mergers also have allowed the predictions of general relativity (GR) to be tested in a strong-gravity and high-luminosity regime of the theory that is challenging to explore with other methods and systems (see, e.g., [94–96]). The high luminosities associated with BBH mergers produce sufficiently strong gravitational waves that nonlinear interactions of the waves in the asymptotic wave zone around the source become significant; this produces features in the gravitational waveforms that are absent in the linearized approximation to GR. The chief nonlinear feature that will be discussed in this paper is the nonlinear GW memory effect [21,23], which is a distinctive and observable example of this type of nonlinear gravitational interaction.

The nonlinear GW memory effect, like the linear effect [19], leads to a lasting offset in the GW strain after a burst of GWs pass by a detector far from an isolated source. This offset is produced by unbound stress energy carried by massive [19] or massless [20, 97] particles, which includes the effective stress-energy of gravitational waves [98]. The memory is also part of an "infrared triangle" that relates the memory effect to the Bondi-Metzner-Sachs supertranslation symmetries [3, 5] of asymptotically flat spacetimes [3,4] and to Weinberg's soft theorem [68] (see, e.g., [2, 99, 100]). This perspective of GW memory as being closely tied to the properties of classical and quantum gravitational scattering serves as an additional motivating factor for observational studies of the GW memory effect. Analogous memory effects and infrared triangles exist in Yang-Mills theories (see, e.g., [2]), but the gravitational-wave memory—despite the relatively weak coupling of gravity compared to other fundamental interactions in nature—may have the best chance of being detected (as we discuss in more detail in the next paragraph). Given that memory effects are generic predictions of gauge and gravitational theories which have not yet been measured, this is a compelling reason to search for these effects and verify experimentally that they are consistent with their theoretical predictions.

There are several studies of the detection prospects of the GW memory effect prior to the detection of GWs from BBH mergers (e.g., [27–31]); however, an important change in viewpoint on memory detection occurred after the detection of the first BBH merger, GW150914 [1]. Despite the GW150914 event being a relatively high signal-to-noise-ratio (SNR) event (the network SNR was roughly 24 [1]), the GW signature of the memory effect for such an event is small compared to the dominant harmonic (from which the SNR was computed) that was used in the detection and parameter estimation. After the first GW150914 event, it was proposed in [32] that one could instead search for evidence of the GW memory effect in a population of BBH mergers in which each individual memory signal was below the threshold of detection. Forecasts that take into account the now-better-constrained BBH-population properties and merger rates (namely, [91–93]) have been performed that show the memory could be detected during the fifth observing run [33–35] when the LIGO detector is in its A+ configuration [36].[1] Dedicated pipelines that search of the GW memory that make use of template-based searches have been applied to the events from all three gravitational-wave transient catalogues [33, 101, 102], and other searches using

[1] The next generation of ground-based GW detectors, Einstein Telescope and Cosmic Explorer, will likely be able to detect the memory effect from single events [35] and will be able to determine if the amplitude of the memory signal is consistent with the value predicted by general relativity to a few percent accuracy [37].

minimally modeled methods have also been implemented [103].

Measuring the GW memory requires a well-defined notion of the memory signal (see, e.g., [104, 105]), and for template-based searches, a signal model that is fast to evaluate. For comparable-mass BBH mergers most of the memory signal accumulates around the time of the merger, which typically implies that input from numerical-relativity (NR) simulations will be needed. Gravitational waveforms computed using extrapolation methods fail to extract the memory from such simulations (see, e.g., [85]), whereas those that use Cauchy-characteristic extraction [48, 49] are capable of resolving the memory signal, and different NR codes have computed the signal from BBH mergers [31, 50] (see also the review [106]). Because generating numerical-relativity waveforms is too computationally intensive to use to generate a sufficient number of GW templates for GW searches and parameter estimation, gravitational waveform models that take NR data as input and interpolate over the parameter space of binary mass ratios and spins are required.

Most efforts to compute the memory signal rely on the fact that it can be computed from oscillatory waveform modes (for which there are time- or frequency-domain waveform models) by using the effective stress-energy tensor of GWs [30, 107] or the continuity equations for charges associated with asymptotic supertranslation symmetries (the supermomenta [14, 108]), as described in [25, 86, 109]. These additional post-processing steps add a computational overhead to searches for the memory effect, and require more delicate signal processing to perform frequency-domain GW data analysis [110]. There has been recent progress in making surrogate [46] and phenomenological frequency-domain waveform models [111] that include the $l = 2$, $m = 0$ spin-weighted spherical-harmonic mode of the GWs, which is the dominant mode in which the memory signal appears for nonprecessing BBH systems. However, the $l = 2$, $m = 0$ mode contains both memory and quasinormal-mode ringing; thus, they do not represent a "memory-only" signal, that would be required to perform a (Bayesian) model comparison between a signal with or without memory, which is the currently implemented method that is used to assess the significance of the GW memory effect in GW data.

This paper is a first step towards developing a stand-alone waveform model for the GW memory effect from nonspinning BBH mergers, which contains information only about the memory effect and not other linear or nonlinear GW phenomena. Our focus will be on two aspects of the memory signal: the limit of extreme mass-ratio inspirals (EMRIs) and the final memory offset that is accumulated after the ringdown stage of a BBH merger. The two will not be independent; the EMRI calculation will feed into the calculation of the final memory offset. In addition to being an input for the full memory signal model, we can see a few other applications of both the EMRI calculation and the final memory offset.

EMRIs, in which a stellar-mass compact-object inspirals into a supermassive black hole, are one well-studied class of sources that the space-based detector LISA [112] likely will observe [113]. They are of interest both in terms of their astrophysics [113] and because the gravitational waves can map out the spacetime geometry precisely, in a result that is referred to as "Ryan's theorem" [114]. The calculations in this paper for EMRIs on quasicircular orbits will show that the memory accumulates mostly during the late inspiral, but over a timescale that for a "typical" EMRI (e.g., a $10M_\odot$ BH inspiralling into a $10^6M_\odot$ BH), will be slow compared to the longest period that the LISA detector can accurately measure. Given that EMRIs likely will form with some residual eccentricity [113], and the memory signal from highly eccentric EMRI systems has a more complicated structure [115], the nonspinning, quasi-circular assumptions that are used in this paper should be revisited for future studies of the memory signals from EMRIs.

Pulsar timing arrays are also on the cusp of a detection of the stochastic background of GWs at low frequencies (see, e.g., [116–119]). NANOGrav and the Parkes Pulsar Timing Array perform searches for GW bursts with memory, where the bursts accumulate on a timescale that is short compared with the shortest period that they can measure [41, 42]. These searches do not specify a source for the memory signal, but they constrain the amplitude of the burst with memory. While forecasts are more optimistic about detecting memory with LISA than pulsar timing arrays [38] (see also [37, 39, 40]), the final memory offset fit that we construct in this paper could be used to interpret the amplitude of a future pulsar-timing-array detection of the memory effect under the hypothesis that the burst with memory was produced by the merger of a supermassive BBH system.[2]

[2]Note that our fit for the $l = 2$, $m = 0$ spherical harmonic mode is a function of the total mass,

3.2.1 Summary and organization of this paper

We now give a brief overview of the organization and main results of this paper. In Sec. 3.3, we review our notation and our prescription for computing the memory signal in the $l = 2$, $m = 0$ spherical-harmonic mode from oscillatory ($m \neq 0$) modes. Section 3.4 then describes the waveform modes used to compute the memory. Because for comparable mass ratios, most of the memory accumulates during the late inspiral, merger and ringdown, we review in Sec. 3.4.1 the numerical-relativity hybrid surrogate model NRHybSur3dq8 [47] that we use to compute the memory signal (for nonspinning BBHs with mass ratios $1 \leq q \leq 8$) in this regime. Although the oscillatory modes of this surrogate model have been hybridized with Effective-One-Body waveforms, to speed up the calculation of the memory signal, we directly hybridize the memory computed from the NRHybSur3dq8 surrogate model with a 3PN memory waveform [87] that accounts for the memory accumulated before the first time at which we evaluate the surrogate model. This PN waveform is also discussed in Sec. 3.4.1. Section 3.4.2 contains a discussion of the 22PN-accurate resummed, factorized waveforms for EMRIs on quasicircular orbits that were computed by Fujita in [120], and an argument for why the inspiral waveforms there are sufficient to compute the memory to leading order in the mass-ratio expansion. We also discuss how we use energy balance to compute the memory signal as a function of a PN parameter

(symmetric) mass ratio and luminosity distance to the source. To use it in the context of interpreting a pulsar-timing-array detection, one would need to take into account not only the dependence of the amplitude on these three parameters in the fit, but also the dependence of the binary's inclination and the array's response to the memory signal as a function of sky position and polarization.

v and of time t along the worldline of the small compact object in the EMRI.

Section 3.5 contains some of the main results of the paper. The first part, Sec. 3.5.1, shows the result of hybridizing the surrogate NRHybSur3dq8 with a 3PN waveform that accounts for the memory accumulated from past infinity up to the starting time of the surrogate. We also provide a polynomial fit (in the symmetric mass ratio) for the offset that must be added to the surrogate at its starting time, which also is valid for mass ratios within the range of hybridization ($1 \leq q \leq 8$). There is a similar polynomial fit for the time-of-coalescence parameter t_c in the post-Newtonian waveform that is required for the hybridization, too. The next part, Sec. 3.5.2.1, gives a more detailed illustration of the increasing importance of the inspiral contribution to the memory signal as the mass ratio becomes more extreme. The remaining parts of Sec. 3.5 show the memory signal as a function of time and velocity and the impact of computing different numbers of oscillatory multipole moments and different PN orders on the memory signal and the final memory offset.

In Sec. 3.6, we construct two polynomial fits in symmetric mass ratio for the final memory offset from nonspinning BBH mergers on quasicircular orbits. The first fit uses the hybridized surrogate, whereas the second uses the same surrogate data, but also fixes the coefficient linear in the symmetric mass ratio to be the value computed from the 22PN-order EMRI calculation. The fit using comparable-mass data only overestimates the memory in the EMRI limit by about ten percent, and it performs similarly to the fit with EMRI information in the comparable-mass limit. We conclude in Sec. 3.7, and there are two Appendices B and C that contain some supplementary results about the properties of the memory integrand and the final memory offset fit.

3.3 Multipolar expansion of the memory signal

We describe how we compute the multipolar expansion of the memory signal in terms of a multipolar expansion of the oscillatory GW strain in this section. We leave the discussion of which oscillatory GW modes we use to Sec. 3.4.

We denote the multipole moments of the gravitational wave strain by h_{lm}, which are the multipoles that arise in the expansion of $h \equiv h_+ - ih_\times$ in terms of spin-weighted spherical harmonics:

$$h \equiv h_+ - ih_\times = \sum_{l=2}^{\infty} \sum_{m=-l}^{l} h_{lm}({}_{-2}Y_{lm}). \tag{3.1}$$

The coefficients h_{lm} are functions of retarded time u, and the spherical harmonics are functions of the polar and azimuthal angles (θ, ϕ), respectively. Note that waveforms are often parameterized by a time t, where t is the ordinary time at a fixed radius r from the source (for example, at the detector). Numerical-relativity waveforms often use h_{lm}, whereas those in the post-Newtonian approximation alternately use the radiative mass and current moments U_{lm} and V_{lm}, which are related to h_{lm} as follows:

$$h_{lm} = \frac{1}{r\sqrt{2}}(U_{lm} - iV_{lm}). \tag{3.2}$$

The multipoles U_{lm} and V_{lm} satisfy the relationship that $U_{l(-m)} = (-1)^m \bar{U}_{lm}$ (and similarly for V_{lm}), where the overline represents complex conjugation. For the non-spinning binaries that we will consider in this paper, the transformation of h_{lm} can also be written as $h_{l(-m)} = (-1)^l \bar{h}_{lm}$ because U_{lm} is nonvanishing when $l+m$ is even and V_{lm} is nonvanishing when $l+m$ is odd.

We compute the displacement memory signal from the balance laws for the flux of supermomentum, as described in [25,35,86,109]. The prescription used in those references calculates just the nonlinear contribution from the "oscillatory" (namely $m \neq 0$) multipole moments of the gravitational wave strain, which do not contain the memory effect. The multipole moments of the memory strain are obtained from integrating a term quadratic in the time derivative of the strain with a spin-weighted spherical harmonic. Because the strain itself is expanded in spin-weighted harmonics, the resulting angular integral will involve the integral of three spin-weighted harmonics. We use the notation of [25] (based on that in [121]) for these integrals. Specifically, we define

$$C_l(s',l',m';s'',l'',m'') \equiv \int d^2\Omega \, ({}_{s'+s''}\bar{Y}_{lm'+m''})({}_{s'}Y_{l'm'})({}_{s''}Y_{l''m''}), \tag{3.3}$$

where the coefficients are nonvanishing only for l in the set Λ defined by

$$\begin{aligned} &\Lambda \equiv \\ &\{\max(|l'-l''|,|m'+m''|,|s'+s''|),...,l'+l''-1,l'+l''\}\,. \end{aligned} \tag{3.4}$$

To compute these coefficients numerically, we will use the fact that they can be written in terms of Clebsch-Gordon coefficients:

$$C_l(s',l',m';s'',l'',m'') = (-1)^{l+l'+l''}\sqrt{\frac{(2l'+1)(2l''+1)}{4\pi(2l+1)}} \\ \times \langle l',s';l'',s''|l,s'+s''\rangle \langle l',m';l'',m''|l,m'+m''\rangle . \tag{3.5}$$

The conventions for the Clebsch-Gordon coefficients that we use are those implemented in MATHEMATICA.

When written in terms of the moments U_{lm} and V_{lm} the multipole moments of the memory strain, $h^{\rm mem}_{lm}$, are given by

$$h^{\rm mem}_{lm}(u) = \frac{1}{4r}\sqrt{\frac{(l-2)!}{(l+2)!}}\sum_{l',l'',m',m''} C_l(-2,l',m';2,l'',m'') \\ \times \int_{-\infty}^{u} du' \left[2is^{l,(-)}_{l';l''}\dot{U}_{l'm'}\dot{V}_{l''m''} + s^{l,(+)}_{l';l''}(\dot{U}_{l'm'}\dot{U}_{l'',m''} + \dot{V}_{l'm'}\dot{V}_{l'',m''})\right]. \tag{3.6}$$

We also defined the coefficients $s^{l,(\pm)}_{l';l''}$ by

$$s^{l,(\pm)}_{l';l''} = 1 \pm (-1)^{l+l'+l''}. \tag{3.7}$$

The sum over the indices l', l'', m' and m'' in Eq. (3.6) must satisfy the constraints that l', $l'' \geq 2$ as well as $|m'| \leq l'$ and $|m''| \leq l''$. However, for a fixed l and m on the left-hand side of Eq. (3.6), the coefficients $C_l(-2,l',m';2,l'',m'')$ in the sum will only be nonzero when $m = m' + m''$ as well as when l, l' and l'' satisfy the relationships in Eq. (3.4). This will decrease the number of modes required to compute the memory signal for particular values of l and m.

In this paper, we focus on the $l = 2$, $m = 0$ mode of the memory signal. This will require that $m'' = -m'$. Specializing Eq. (3.6) to this case gives

$$h_{20}^{\mathrm{mem}}(u) = \frac{\sqrt{6}}{48r} \sum_{l',l'',m'} C_2(-2, l', m'; 2, l'', -m') \int_{-\infty}^{u} du' \left\{ 2is_{l';l''}^{2,(-)} \dot{U}_{l'm'} \dot{V}_{l''(-m')} \right.$$
$$\left. + s_{l';l''}^{2,(+)} \left[\dot{U}_{l'm'} \dot{U}_{l''(-m')} + \dot{V}_{l'm'} \dot{V}_{l''(-m')} \right] \right\}. \tag{3.8}$$

We will be suppressing the "mem" superscript in the rest of this paper for the $(2, 0)$ mode, because we only use oscillatory modes with $m \neq 0$ on the right-hand side of Eq. (3.8) (and there will be no ambiguity that the $m = 0$ modes are the "memory modes"). In addition, specializing to $l = 2$ requires that the magnitude of the difference of l' and l'' (i.e., $|l' - l''|$), is at most 2 for these modes to contribute to the $(2, 0)$ memory mode.

In the discussion that follows, we find it convenient to commute the sum and integral in Eq. (3.8) to write the memory signal in the form

$$h_{20}(u) = \int_{-\infty}^{u} du' \, \dot{h}_{20}, \tag{3.9}$$

where the dot means a derivative with respect to u'.[3] The expression for the integrand $\dot{h}_{20}$ can be inferred from Eq. (3.8). We will also find it useful to introduce the notation for the "final memory offset"

$$\Delta h_{20} = \lim_{u \to \infty} h_{20}(u), \tag{3.11}$$

[3]For waveforms parameterized by t, we would instead have the analogous expression

$$h_{20}(t) = \int_{-\infty}^{t} dt' \, \dot{h}_{20}, \tag{3.10}$$

where the dot now denotes a derivative with respect to t'.

	$l=2$	$l=3$	$l=4$	$l=5$
m values	$\{\pm1,\pm2\}$	$\{\pm1,\pm2\pm3\}$	$\{\pm2,\pm3,\pm4\}$	$\{\pm5\}$

Table 3.1: m values for each l value in the NRHybSur3dq8 surrogate model that are used in the calculations for comparable mass ratios.

which represents the total memory strain accumulated over all times.

3.4 Waveform multipole moments and memory signals

We now turn to discussing which waveform multipole moments U_{lm} and V_{lm} we input into the right-hand side of Eq. (3.6) to compute the memory signal. We first cover the waveform modes required for the comparable-mass (mass ratios $q = m_1/m_2 < 8$), and then we turn to the extreme mass-ratio case.

3.4.1 Comparable mass ratios

For comparable mass ratios, the memory signal grows most rapidly and accumulates most of its offset close to the merger, when NR (or IMR waveforms fit to NR) are necessary to accurately model the gravitational waves. Thus, we will need to use an IMR waveform, and we use the NR hybrid surrogate model NRHybSur3dq8 [47], which is calibrated for aligned spin BBHs with mass ratios $1 \leq q \leq 8$ (we specialize to nonspinning systems, however). The oscillatory modes in NRHybSur3dq8 have been hybridized with EOB waveforms to produce waveforms that allow the surrogate to be evaluated for times much longer than the duration of the numerical relativity waveforms from which the surrogate model is built.

The NRHybSur3dq8 model contains only a subset of all the $l \geq 2$ modes in the waveform, which are listed in Table 3.1. We compute the contribution to the displacement memory signal in the $(2, 0)$ mode from Eq. (3.8) using the surrogate modes given in Table 3.1. The integral for the memory signal has as its lower limit negative infinity, which would require evaluating the surrogate model for an infinitely long time. This, however, is not feasible, so we instead would like to determine the appropriate initial offset to apply to the surrogate memory at a finite starting time.

We achieve this by hybridizing the memory signal computed from the surrogate with the 3PN memory waveform that had been computed by Favata [87]. The 3PN memory signal in [87] is written in terms of the post-Newtonian parameter x, which is defined to be

$$x \equiv (M\Omega)^{2/3}, \tag{3.12}$$

where $M = m_1 + m_2$ is the total mass of the binary with primary mass m_1 and secondary mass m_2, and Ω is the orbital frequency of the circular orbit. At Newtonian order, the PN parameter can be written in terms of time as

$$x(t) = \frac{1}{4}\left[\frac{\eta}{5M}(t_c - t)\right]^{-1/4}. \tag{3.13}$$

The time here, is usually the coordinate time in the near zone of the source, but one can write it in terms of retarded time u, too. We introduced the parameter t_c, which is the time of coalescence, and the symmetric mass ratio, η. There are several equivalent expressions for it in terms of the individual masses, m_1 and m_2 or the mass ratio $q = m_1/m_2$:

$$\eta = \frac{q}{(q+1)^2} = \frac{m_1 m_2}{M^2} = \frac{\mu}{M}. \tag{3.14}$$

The last equality used the definition of the reduced mass $\mu = m_1 m_2/M$. We reproduce Favata's expression for the 3PN memory in terms of x:

$$\begin{aligned} h_{20}^{\rm PN}(x) = \frac{4}{7}\sqrt{\frac{5\pi}{6}}\eta x \Bigg\{ 1 + x\left(- \frac{4075}{4032} + \eta\frac{67}{48}\right) + x^2\Bigg(- \frac{151877213}{67060224} - \eta\frac{123815}{44352} \\ + \eta^2\frac{205}{352}\Bigg) + \pi x^{5/2}\left(- \frac{253}{336} + \eta\frac{253}{84}\right) + x^3\Bigg[- \frac{4397711103307}{532580106240} \\ + \eta\left(\frac{700464542023}{13948526592} - \frac{205}{96}\pi^2\right) + \eta^2\frac{69527951}{166053888} + \eta^3\frac{1321981}{5930496}\Bigg]\Bigg\}. \end{aligned} \tag{3.15}$$

We introduced the "PN" superscript on the $(2,0)$ mode of the strain to indicate it is only valid in the regime of validity of the PN expansion.

As a brief comment, the $(2,0)$ mode at 3PN order is computed from oscillatory waveform modes with $l \leq 6$, so it contains information about additional l modes that are not included in the surrogate modes in Table 3.1. However, it is still possible to achieve a robust hybridization between the surrogate and the 3PN memory as we show in more detail in Sec. 3.5.1.

3.4.2 Extreme mass ratios

For extreme mass-ratio systems, we will work to linear order in the symmetric mass ratio $\eta \ll 1$. Such systems tend to be strongly relativistic, so that perturbation theory about a Schwarzschild background (for non-spinning binaries) is a more suitable approach for modeling these systems (see, e.g., [122]). Thus, neither the 3PN memory waveform nor the NRHybSur3dq8 surrogate will be well suited for computing the memory signal from EMRIs. There is a more recent EMRI surrogate, EMRISur1dq1e4 [45], which is calibrated up to a mass ratio of $q = 10^4$, and spans a duration of time of order $10^4 M$. This could be used for EMRI systems with more comparable mass ratios. The rest of this subsection covers how we compute a memory signal that is accurate to linear order in η.

3.4.2.1 Contributions of different stages of an extreme mass-ratio binary coalescence to the memory signal

The coalescence of an extreme mass-ratio binary takes place in several stages. The stage with the longest duration is referred to as the "adiabatic inspiral"; in this stage, the small compact object can be modeled as adiabatically evolving between a sequence of bound orbits because of gravitational radiation reaction. As the small compact object approaches the innermost stable circular orbit (ISCO) radius, the adiabatic approximation becomes inaccurate, and the binary's evolution is better described by a "transition from inspiral to plunge" [123] (such a transition also occurs for comparable-mass binaries; see, e.g., [124]). As described in [125–127], the characteristic timescale to cross the ISCO in this transition stage goes as $M/\eta^{1/5}$. Naturally, the "plunge" stage follows this transition stage, which occurs after the small compact object crosses the ISCO and before it enters the event horizon of the more massive black hole. At the end of the plunge, the small compact object passes through the event horizon of the primary black hole, and there will be a merger and ringdown signal in the gravitational waveform associated with this process.

We now argue that for extreme mass-ratio systems, the majority of the GW memory offset will accumulate during the adiabatic inspiral. For any mass ratio (and all stages of the binary's evolution), the radiative moments U_{lm} and V_{lm} are proportional to the reduced mass $\mu = M\eta$. Their time derivatives are proportional to $\eta\Omega \sim \eta x^{3/2}/M$ times the radiative moments. The integrand that arises in the memory integral in Eq. (3.6) is proportional to η^2. However, the memory accumulates over the radiation-reaction timescale, which goes like M/η, which is why the memory strain scales as $\eta M/r = \mu/r$ (up to other dimensionless factors). This can be determined more quantitatively by assuming energy balance, writing the integral with respect to the PN parameter x, and converting the integration measure from dt [as in Eq. (3.10)] to dx. This procedure introduces a factor of dt/dx, which scales as $1/\eta$, thereby canceling one factor of η in the integrand for the memory. The integral is evaluated over an η-independent range of x, which leads to the scaling linear in η.

The transition from inspiral to plunge, followed by the plunge, takes place over the timescale $M/\eta^{1/5}$ described in [125–127]. Given that the memory integrand scales with η^2, a scaling argument (like that described for the adiabatic inspiral above) suggests that the contribution to the GW memory offset should scale as $\eta^{9/5}$ from these stages. It would be beneficial to perform a more quantitative calculation of the memory signal during these stages in future work, however. In our subsequent calculations, we will ignore the contributions to the memory from the transition from inspiral to plunge and the plunge itself, because we expect that they contribute to the memory signal at higher orders in the mass ratio than the leading linear effect during the adiabatic inspiral. We will also truncate the adiabatic inspiral at the ISCO radius rather than at the ISCO radius plus a correction term of order $\eta^{2/5}$ times the ISCO radius (which is the length scale over which the transition takes place), for simplicity.

Finally, during the merger and ringdown stages of the waveform (where we are referring to the merger as the end of the plunge phase), the integrand for the memory signal, with dt as the measure [as in Eq. (3.10)], should scale as η^2, but the memory accumulates on a timescale of order the dynamical time of the massive black hole, which is determined by M and is independent of η. Thus, the memory signal generated during the merger and ringdown will be of order $\eta^2 M/r = \eta\mu/r$. In the comparable mass-ratio limit, the symmetric mass ratio is of order $\eta \sim 1/4$, and the more relativistic speeds during the merger and higher radiative losses compensate for the shorter timescale over which the memory accumulates (and in fact, the merger and ringdown stages produce the largest part of the memory signal). For extreme mass-ratios however, the η^2 scaling of the memory during the merger and ringdown also makes its contribution negligible (when working to linear order in η), despite the fact that the merger is nominally the most relativistic stage of a binary merger. We will provide a visual illustration of the relative importance of the adiabatic inspiral over the other stages of the coalescence in Sec. 3.5.2.1. Because the adiabatic inspiral waveforms will be sufficient to compute the memory signal at the accuracy in η at which we are working, we next describe the waveform modes that we use during the adiabatic inspiral.

3.4.2.2 Factorized post-Newtonian waveform

We use the 22PN analytical waveforms computed by Fujita [120] for Schwarzschild black holes based on the expansion of analytical solutions to the Teukolsky equation [16] in the low-frequency limit [128]. These waveforms are typically written in terms of $v \equiv \sqrt{x} = (M\Omega)^{1/3}$, as defined in Eq. (3.12). The waveforms are written in a resummed, factorized form (see [129, 130]) which we now review. The discussion below will focus on modes with $m > 0$; modes with $m < 0$ can be obtained from the fact that for nonprecessing binaries $h_{l(-m)} = (-1)^l \bar{h}_{lm}$. The multipole moments of the strain are written as

$$h_{lm} = h_{lm}^{(\mathrm{N},\epsilon_p)} \hat{S}_{\text{eff}}^{(\epsilon_p)} T_{lm} e^{i\delta_{lm}} (\rho_{lm})^l. \tag{3.16}$$

The label ϵ_p is the parity of the mode for nonprecessing binaries:

$$\epsilon_p = \begin{cases} 0 & \text{for } l+m \text{ even} \\ 1 & \text{for } l+m \text{ odd} \end{cases}. \tag{3.17}$$

The first term $h_{lm}^{(\mathrm{N},\epsilon_p)}$ in Eq. (3.16) is the leading-order "Newtonian" part of the waveform. It is linear in η and given by

$$h_{lm}^{(\mathrm{N},\epsilon_p)} = \frac{\mu}{r} n_{lm}^{(\epsilon_p)} (-v)^{l+\epsilon_p} Y_{l-\epsilon_p,-m}(\pi/2, \phi), \tag{3.18}$$

where the coefficient $n_{lm}^{(\epsilon_p)}$ is defined in the parity even and odd cases, respectively by

$$n_{lm}^{(0)} = \frac{8\pi(im)^l}{(2l+1)!!}\sqrt{\frac{(l+2)(l+1)}{l(l-1)}}, \tag{3.19a}$$

$$n_{lm}^{(1)} = -\frac{16\pi i(im)^l}{(2l+1)!!}\sqrt{\frac{(l+2)(2l+1)(l^2-m^2)}{(2l-1)(l+1)l(l-1)}}. \tag{3.19b}$$

The effective source $S_{\text{eff}}^{(\epsilon_p)}$ is given by the relativistic reduced energy ($\tilde{E} > 0$) of stable circular geodesics in the Schwarzschild spacetime for the even-parity case and v times the reduced angular momentum per total mass of the geodesics for the odd-parity case:

$$S_{\text{eff}}^{(\epsilon_p)} = \begin{cases} \tilde{E} = \dfrac{1-2v^2}{\sqrt{1-3v^2}} & \text{for } \epsilon_p = 0 \\ \dfrac{v\tilde{L}}{M} = \dfrac{1}{\sqrt{1-3v^2}} & \text{for } \epsilon_p = 1 \end{cases}. \tag{3.20}$$

This effective source can be expanded straightforwardly in a PN series in v. The factor T_{lm} is a "resummed tail factor" that is defined to be

$$T_{lm} = \frac{\Gamma(l+1-i2mv^3)}{\Gamma(l+1)} e^{m\pi v^3} e^{i2mv^3 \ln(4mv^3/\sqrt{e})}. \tag{3.21}$$

The natural log of T_{lm}, for small v, can be expanded in terms of a Taylor series with coefficients involving the polygamma function; it is then straightforward to use the Taylor series for the exponential of this series to obtain the post-Newtonian expansion of T_{lm}. The terms $(\rho_{lm})^l$ and $e^{i\delta_{lm}}$ represent the remaining amplitude and phase of the waveform that cannot be expressed in terms of the Newtonian, effective source, or resummed tail factors. Post-Newtonian series for ρ_{lm} and δ_{lm} can be downloaded in a format adapted for MATHEMATICA on the webpage [131]. The expressions are quite lengthy series in powers of v (including terms of the form $v^j(\ln v)^k$, for whole numbers j and k with $j > k$); thus, we do not give their explicit expressions here.

The result of this calculation is that we can obtain the radiative modes U_{lm} and V_{lm} or h_{lm} as PN series in the parameter v, which can then be used to compute the EMRI memory signal, as described next.

3.4.2.3 Evaluating the memory signal

Once the modes h_{lm} are computed, we can use Eqs. (3.2) and (3.8) to compute the memory signal h_{20} for the inspiral of an EMRI. Because $\dot{v}$ is an order η term, whereas $\dot{\phi} = \Omega = v^3/M$ is η independent, then the time derivatives of h_{lm} satisfy

$$M\dot{h}_{lm}(v) = imv^3 h_{lm}(v) + O(\eta^2)\,. \tag{3.22}$$

Thus, the time derivatives are straightforward to compute analytically. For the expression for the memory signal h_{20} to be accurate to 22 PN order, we need to include oscillatory memory terms up to $l' = 25$ (and similarly for l'') in Eq. (3.8); as a result, there are hundreds of terms in the sum that contribute to the memory signal.

The product of two modes in the integrand in Eq. (3.8) always involves one mode with positive m multiplying one with $-m$. Because the modes h_{lm} satisfy the complex-conjugate relationship $h_{l(-m)} = (-1)^l \bar{h}_{lm}$, then the terms in the integrand of the form $|\dot{U}_{lm}|^2$ (or similarly for V_{lm}) depend only on the amplitudes of all the terms in the factorized waveform in Eq. (3.16). However, for the terms that involve products of

U_{lm} (or V_{lm}) modes with different l, then the phases will contribute, but only the parts of the phases that are dependent on both l and m (for example, the ϕ dependence in the Newtonian part of the waveform will not contribute). This phase dependence then comes from the supplementary phase δ_{lm} and the phase in the resummed tail $T_{lm} = |T_{lm}|e^{i\tau_{lm}}$. The resulting memory waveform can then be expressed in terms of the amplitude and the cosine of the difference in the phase $\delta_{lm} + \tau_{lm}$ for modes with different values of l. These cosine terms must be expanded in a PN series, as well.

Finally, to compute the integral in Eq. (3.8), we postulate the energy balance holds adiabatically. More specifically, we assume the secondary evolves from a circular geodesic parameterized by a relativistic specific energy $\tilde{E} > 0$ to another circular orbit with a smaller such energy in response to radiative losses from gravitational-wave emission. The specific energy $\tilde{E}$ is that of a test particle moving on a circular orbit around a Schwarzschild black hole, which is the same expression as that given in the $\epsilon_p = 0$ case of Eq. (3.20), namely

$$\tilde{E} = \frac{1 - 2v^2}{\sqrt{1 - 3v^2}}. \tag{3.23}$$

We will consider losses that include the GW luminosity radiated to infinity and into the horizon:

$$\frac{dE_{\rm GW}}{dt} = \left(\frac{dE}{dt}\right)_\infty + \left(\frac{dE}{dt}\right)_H. \tag{3.24}$$

The time variable t on the left-hand is the time that parameterizes the geodesic followed by the secondary in the EMRI system. While it is most natural to parameterize the power radiated to infinity by retarded time $u = t - r_*$ (where r_* is the tortoise coordinate in the Schwarzschild spacetime) and the power radiated into the horizon by advanced time $t + r_*$, they can equivalently be parameterized by d/dt.[4] For our main computations, we will use 22PN-accurate expressions for the GW luminosity at infinity and 22.5PN-accurate expression for the power radiated into the horizon, which are given in [120]; the lengthy expressions for these GW luminosities also can be obtained from [131]. We will perform convergence tests in which we consider lower PN orders (which also involves fewer multipole moments of the luminosities—see Sec. 3.5.2.4 for further detail).

We next rewrite the integral with respect to u in Eq. (3.9) instead as an integral with respect to the velocity $v = (M\Omega)^{1/3}$. To do so, we use the chain rule, the fact that $E = \mu\tilde{E}$ (for the reduced mass μ), the fact that $dt = du$ (as discussed above), and the notion of energy balance,

$$\frac{dE}{dt} = -\frac{dE_{\mathrm{GW}}}{dt}. \tag{3.25}$$

This then allows us to write

$$h_{20}(v) = \int_0^v dv' \frac{dh_{20}}{dv'}, \tag{3.26}$$

[4]This follows because a differential change in retarded time is related to the coordinate time at the location of the geodesic by $du = dt - dr_* = dt(1 - \dot{r}_*) \approx dt$, where $\dot{r}_*$, the change in radial position of the particle is an order η change on circular orbits, and we have been consistently ignoring such higher order in η terms throughout this paper.

where we defined

$$\frac{dh_{20}}{dv} = \frac{dE}{dv}\left(\frac{dE}{dt}\right)^{-1}\dot{h}_{20} = -\frac{dE}{dv}\left(\frac{dE_{\rm GW}}{dt}\right)^{-1}\dot{h}_{20}, \tag{3.27}$$

We also assumed that the parameter v goes to zero as u goes to minus infinity.

The derivative of the energy $d\tilde{E}/dv$ is

$$\frac{d\tilde{E}}{dv} = \frac{v(6v^2-1)}{(1-3v^2)^{3/2}}, \tag{3.28}$$

and can be expanded to the necessary PN order. When combined with the PN expansions for the inverse of $dE_{\rm GW}/dt$ in Eq. (3.24) and the PN expansion of $\dot{h}_{20}$, which can be obtained from Eqs. (3.8) and (3.9), we can obtain the 22PN accurate expansion of dh_{20}/dv or $h_{20}(v)$.

We will also find it convenient to have a PN expression for $h_{20}(t)$, which we obtain by integrating dt from some reference time t_i in the past:

$$t - t_i = \int_{t_i}^{t} dt' = -\int_{v_i}^{v} dv' \left(\frac{dE}{dv'}\right)\left(\frac{dE_{\rm GW}}{dt}\right)^{-1} \tag{3.29}$$

to give t as a function of v. We then can construct an interpolation function for $v(t)$, which we substitute into our expression for $h_{20}(v)$ to obtain $h_{20}(t)$.

3.4.2.4 Required multipole and PN orders for computing the memory

We described in Sec. 3.4.2.3 how the memory signal can be computed from the factorized waveform modes h_{lm} (see Sec. 3.4.2.2 for a discussion of how the factorized modes are computed) and the GW luminosity dE/dt. For the luminosity, the part radiated to infinity (into the horizon) has been computed in [120] (respectively, [132]) up to 22PN (22.5PN) order beyond the leading quadrupole formula for the power at infinity,

$$\left(\frac{dE}{dt}\right)_N = \frac{32}{5}\eta^2 v^{10}. \tag{3.30}$$

To compute the power at infinity, Ref. [120] uses the expression

$$\left(\frac{dE}{dt}\right)_\infty = \frac{r^2}{16\pi}\sum_{l,m}\left|\dot{h}_{lm}\right|^2, \tag{3.31}$$

where the sum runs over integers $l \geq 2$ and $m \in [-l, l]$. Having an accuracy of 22PN order requires including $\dot{h}_{lm}$ modes with l running from 2 to 24, because the scaling of $\dot{h}_{lm}$ with v is $\dot{h}_{lm} \sim v^{l+\epsilon_p+3}$. For the $l = 24$ modes, only the leading "Newtonian" part of the mode contributes for $\epsilon_p = 0$; however, for the $l = 2$, $m = 2$ mode, the full 22PN accuracy of the mode is required.[5] To compute the radiated power at infinity, we use the per mode data that is available on the website [131]. That data is given in terms of modes $\eta_{lm}^{(\infty)}$ (which are *not* related to our symmetric mass ratio, $\eta = \mu/M$). The $\eta_{lm}^{(\infty)}$ related to the GW luminosity at infinity by

$$\left(\frac{dE}{dt}\right)_\infty = \left(\frac{dE}{dt}\right)_N \sum_{l=2}^{24}\sum_{m=1}^{m=l}\eta_{lm}^{(\infty)}. \tag{3.32}$$

Note that the normalization of the expression is such that only positive values m are summed over.

[5]Interpolating between these two values, one can find that for a mode h_{lm} with $l \geq 2$ and $m \in [-l, l]$ that $(24 - l - \epsilon_p)$ PN orders beyond the leading "Newtonian" part of the mode would be required to obtain the contribution of $\dot{h}_{lm}$ to the radiated power at 22PN orders beyond the quadrupole formula.

The radiated power into the future event horizon involves a different number of multipole modes, because it is determined by the magnitude squared of different radiative degrees of freedom: specifically, the shear of the generators of the horizon, which is related to the time integral of the Weyl scalar Ψ_0 [133]. The PN scaling of each (l, m) mode of the flux goes as $v^{10+4l+2\epsilon_p}$ on the horizon, rather than $v^{6+2l+2\epsilon_p}$ at infinity; this implies that the maximum required value of l will be $l = 11$. In the paper [132] and the data available on the website [131], the per mode contributions to the GW luminosity down the horizon, $\eta_{lm}^{(H)}$, are normalized such that

$$\left(\frac{dE}{dt}\right)_H = v^5 \left(\frac{dE}{dt}\right)_N \sum_{l=2}^{11} \sum_{m=1}^{m=l} \eta_{lm}^{(H)}, \tag{3.33}$$

where again the sum runs over positive values of m only.

For computing the memory signal h_{20}, the required multipoles $\dot{h}_{lm}$ are those at infinity, but they are not precisely the same as those used for calculating the GW luminosity at infinity, in the following sense. Because the expression in Eq. (3.8) involves products of modes with $|l' - l''| \leq 2$, then this requires computing many of the oscillatory h_{lm} modes at 0.5 or 1PN order higher than is required for the radiated power. In addition, it also requires some of the $l = 25$ modes (specifically those with $\epsilon_p = 0$). The 1PN-order-higher maximum here is specific to the $l = 2$ memory signal; for $l > 2$, this would require evaluating some modes at $(l/2)$PN orders higher than is required for the equivalent PN order in the radiated power.

3.5 Memory signals for comparable and extreme mass ratios

In this section, we summarize the main features of the memory signals from nonspinning binary black-hole mergers in comparable, intermediate and extreme mass-ratio binaries.

3.5.1 Memory for comparable mass ratios

For comparable mass-ratio binaries, we argued in Sec. 3.4.1 that to obtain a memory signal of arbitrary length in time, it is more efficient computationally to hybridize the NRHybSur3dq8 signal to a 3PN memory waveform than to evaluate the surrogate for a long stretch and compute the memory waveform from just the surrogate model. We now describe how we perform this hybridization.

Given the different coordinate conditions in PN and NR calculations, and the fact that there is not a known mapping between the two conditions, some postulates must be made to hybridize the two waveforms. We assume that the individual masses m_1 and m_2 have the same values in the PN and NR contexts and that the time variables t have the same meaning in both cases. With these assumptions, we can perform the hybridization by adding two free parameters, one in the NR surrogate and one in the PN memory waveform, and specifying the initial and final times over which the hybridization takes place. Specifically, for the NR surrogate, we add a positive, undetermined constant, h_0 to the memory signal, which represents the amount of memory that has accumulated in the surrogate at the earliest time at which it is evaluated ($t/M = -10^4$ here). For the PN waveform, we treat the time of coalescence t_c as an undetermined parameter in Eq. (3.13); when this is substituted into Eq. (3.15), this gives a time-domain PN signal for the memory effect, $h_{20}^{\rm PN}(t)$. There is also freedom in the choice of the initial and final times over which the hybridization takes place, denoted by t_1 and t_2 respectively. We choose $t_1 = -5000M$ and $t_2 = -4000M$, because the interval $t \in [t_1, t_2]$ was the same range used by the NRHybSur3dq8 surrogate model to hybridize between the NR and the EOB waveforms for the oscillatory $m \neq 0$ waveform modes.

We then hybridize the 3PN and NR surrogate waveforms by solving a nonlinear optimization problem for the two free parameters h_0 and t_c by minimizing the cost function

$$C[h_{\rm surr}, h_{\rm PN}] = \frac{\displaystyle\int_{t_1}^{t_2} dt |h_{20}^{\rm surr}(t) - h_{20}^{\rm PN}(t)|^2}{\displaystyle\int_{t_1}^{t_2} dt |h_{20}^{\rm surr}(t)|^2} \,. \tag{3.34}$$

We hybridize at 50 different values of q between $q = 1$ and $q = 8$, which we distribute uniformly in η (this corresponds to η in the range $[0.01, 0.25]$, accurate to the hundredths digit place). An example of the result of the hybridization procedure is shown in Fig. 3.1 for an equal-mass ($q = 1$) nonspinning BBH merger. We truncate the 3PN memory at the peak time of the waveform $t/M = 0$, because it starts to deviate significantly from the NR surrogate memory waveform signal at larger values (which is not surprising, because $t/M = 0$ is the time at which the $l = 2$, $m = 2$ mode of the surrogate waveform reaches its peak value). The hybridized memory, in principle,

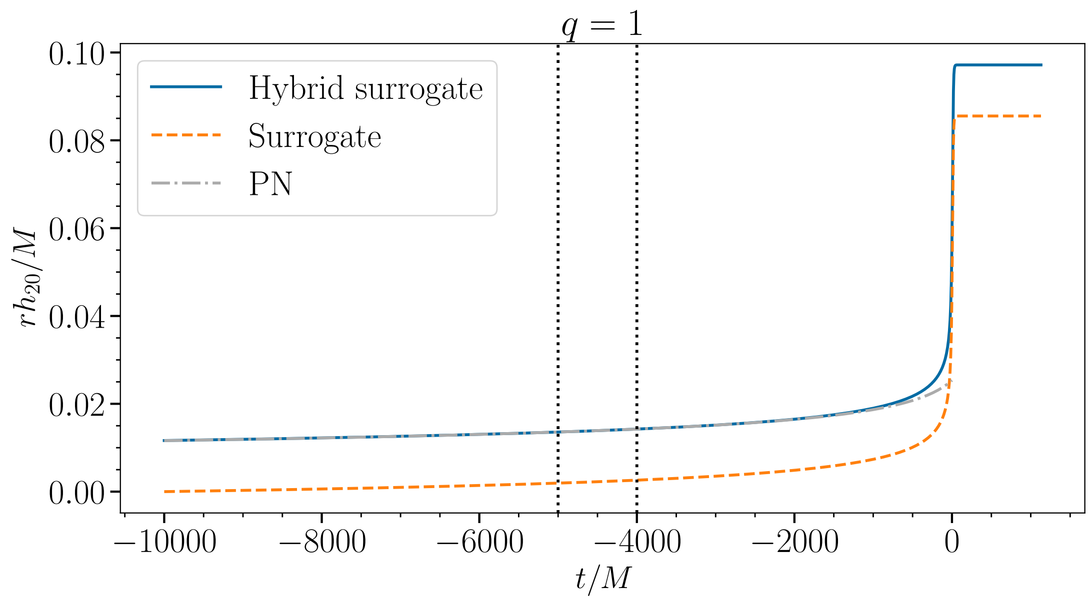

Figure 3.1: **Hybridized memory signal versus time**: The three curves shown are the PN memory signal (dashed-dotted gray curve), surrogate memory signal computed from the NRHybSur3dq8 surrogate waveform modes (dashed orange curve), and hybridized surrogate memory signal (solid blue curve) for an equal-mass non-spinning BBH merger. The vertical black dotted lines indicate the region over which the PN and surrogate memory signals were hybridized (specifically, from $t_1 = -5000M$ to $t_2 = -4000M$; see the main text for why this region was selected). The surrogate memory has been computed using the modes summarized in Table 3.1.

could be obtained for an arbitrarily long duration by evaluating the PN waveform from the desired initial time to t_2 and the surrogate from t_1 to the desired final time. Over the hybridization interval $t \in [t_1, t_2]$, the PN and surrogate waveforms could be smoothly "blended" to obtain the memory signal $h_{20}(t)$.[6]

When we require the surrogate memory signal with the final memory offset ac-

[6]The term "blended" means that the PN signal should be multiplied by a smooth function that is one at $t \leq t_1$ and goes to zero at $t \geq t_2$, whereas the surrogate should be multiplied by one minus this function. Adding the PN and surrogate waveforms scaled by these functions then will yield a smooth memory signal, up to small errors in the hybridization procedure that cause differences in the values of the PN and surrogate waveforms over the hybridization interval $[t_1, t_2]$.

cumulated in the limit as the past time goes to minus infinity, but only for a short time interval, we do not need to perform the blending described above. In the hybridization procedure, we evaluated the surrogate waveforms for the same length of time between $t/M = -10^4$ and $t/M = 130$ for the different values of η that we used. Therefore, the value of h_0 that comes out of the hybridization is the value of the memory signal at the starting time ($t/M = -10^4$) that should be added to the surrogate waveform to compute the remainder of the memory signal consistent with one that extends indefinitely into the past. For this reason, it is convenient to have an expression for $h_0(q)$ that can be evaluated for any value of q in the range $[1, 8]$, rather than the fixed values of q at which the hybridization was performed. Thus, we find it useful to have a fit for h_0 over this range.

In fact, we construct polynomial fits for both h_0 and t_c as a function of η over this range using a quartic polynomial.[7] We write the polynomials in the form

$$h_0^{\text{fit}}(\eta) = \frac{\mu}{r}\sum_{j=0}^{4} a_j \eta^j, \tag{3.35a}$$

$$t_c^{\text{fit}}(\eta) = M\sum_{j=0}^{4} b_j \eta^j, \tag{3.35b}$$

where the coefficients a_j and b_j are given in Table 3.2. The t_c fit could be used to determine the correct 3PN time-domain waveform to match with the surrogate model, though we do not use it for that purpose in this paper.

[7]The choice of quartic order was determined empirically. The residuals between the fit and the values of h_0 determined through hybridization (at specific values of η) improved as the polynomial fit order was increase from linear to quartic, but did not decrease as dramatically for higher-order polynomial fitting functions.

Coefficient	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$
a_j	5.67×10^{-4}	5.81×10^{-2}	-9.29×10^{-2}	2.02×10^{-1}	-2.04×10^{-1}
b_j	2.11×10^{3}	-2.04×10^{4}	1.03×10^{5}	-2.85×10^{5}	3.27×10^{5}

Table 3.2: Coefficients for the h_0 and t_c polynomial fits in Eq. (3.35). The fits were constructed by using 50 BBH systems with mass ratios equally spaced in η for a range of mass ratios with the range of validity of the NRHybSur3d18: $1 \leq q \leq 8$. The values of h_0 and t_c were obtained using the hybridization procedure described in Sec. 3.5.1.

Note that in Fig. 3.1, the hybridized memory starts at $t = -10^4 M$ at a value of around one tenth of the final memory offset, rather than zero for the surrogate memory without hybridization. Using the PN expression for the memory in Eq. (3.15) and the PN parameter as a function of time in Eq. (3.13), we can estimate that to decrease the initial value of the memory from the surrogate by a factor of ten (to around one hundredth of the final offset), the surrogate would need to be evaluated for an amount of time 10^4 times longer. Although the computation of the memory from the surrogate for one waveform takes of order one second, increasing the length by this factor of 10^4 would make the calculation take hours. The fitting function in Eq. (3.35a) can be evaluated in negligible time. This indicates that there is a considerable advantage for using this fitting function when it is important to capture the initial offset in the memory.

Note also that the calculation of the 3PN memory signal in [87] [our Eq. (3.15)] uses additional multipole moments that are not those given in Table 3.1. We computed a PN memory signal using just the multipole modes in Table 3.1 evaluated at 3PN accuracy. The relative difference between this PN memory signal and the full 3PN result in Eq. (3.15) was of order 10^{-6}. Hybridizing the surrogate to this limited-multipole PN signal led to relative difference of the same order. The smallness of this difference is why we opted to use the full 3PN result for the hybridization.

3.5.2 Memory for extreme mass ratios

We discuss several aspects of the results for the memory from nonspinning EMRI systems in this part. We first demonstrate with the EMRISur1dq1e4 model [45] that as the mass ratio becomes more extreme, a greater fraction of the memory signal accumulates during the inspiral rather than the merger and ringdown (consistent with the analytical argument given in Sec. 3.4.2). We then study the convergence of the EMRI memory signal that is calculated using the factorized oscillatory waveforms described in Sec. 3.4.2 as a function of the PN order and the multipole index l.

3.5.2.1 EMRI surrogate results

The results in this section using the EMRISur1dq1e4 surrogate model [45] are intended to be illustrative of how the morphology of the memory signal changes between comparable and extreme mass ratios. Thus, we will not use the results here in the construction of the memory fit in the following section, nor will we try to hybridize the EMRISur1dq1e4 surrogate memory signal with an appropriate PN model during the inspiral here either.[8]

[8]The EMRI surrogate is not hybridized and starts at a time $t/M = -10^4$ earlier than the merger time (the peak of the $l = 2$, $m = \pm 2$ waveform). As the mass ratio becomes more extreme, the 3PN memory waveform in Eq. (3.15) begins to lose accuracy at the starting time of the surrogate waveform. Instead, one could use the high-PN-order memory for EMRIs to hybridize, but that could also run into errors because it neglects higher-order terms in the symmetric mass ratio η. We are not aware of oscillatory waveform modes to second order in mass ratio and at high PN order, though

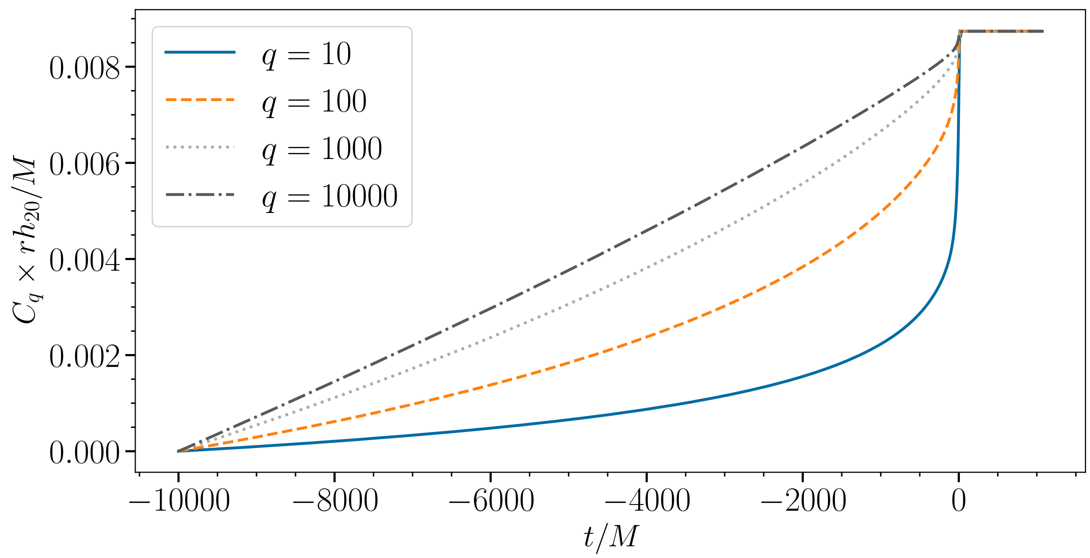

Figure 3.2: **Memory signal for comparable to extreme mass ratios versus time**: The time-domain memory signal computed using the EMRISur1dq1e4 surrogate model is shown for nonspinning BBH mergers with mass ratios $q = 10$ (the solid blue curve), $q = 10^2$ (the dashed orange curve), $q = 10^3$ (the dotted light-gray curve) and $q = 10^4$ (the dark-gray dashed-dotted curve). As the mass ratio increases, a larger fraction of the final memory offset accumulates during the inspiral in a fixed range of time. The memory signals have been scaled by the corresponding factors of C_q defined in Eq. (3.36), because the memory becomes smaller as the mass ratio becomes more extreme. The values were chosen to make the final memory strain equal for all mass ratios, thereby allowing all the curves to be plotted on the same scale and illustrating the accumulation of the memory more clearly over this fixed time interval. For reference, the values for the C_q factors are $C_{10} = 1$, $C_{10^2} = 2.84 \times 10^1$, $C_{10^3} = 1.13 \times 10^3$, and $C_{10^4} = 7.08 \times 10^4$. The oscillatory waveform modes used to compute the memory signal are those listed in Table 3.1.

The EMRISur1dq1e4 surrogate model contains the same waveform modes given in Table 3.1, and it also includes modes for $l = 5$ and $m = \pm 3$ and ± 4. In this section, we do not use these additional waveform modes in the EMRISur1dq1e4 surrogate to compute the h_{20} memory signal via Eq. (3.8), we use just the modes listed in Table 3.1. In Fig. 3.2, we compute the memory signal with the EMRISur1dq1e4 surrogate for nonspinning BBH systems with the mass ratios $q = 10$, 10^2, 10^3 and 10^4. The merger (understood as the peak amplitude of the $l = 2$, $m = \pm 2$ waveform modes) occurs at the time $t = 0$ as with the comparable mass surrogate, and the surrogate model is not hybridized and does not extend earlier than the time $t/M = -10^4$, which is the earliest time at which the memory is plotted in Fig. 3.2. The EMRI surrogate does not extend beyond a time $t/M = 130$, so we extended the length of time after $t = 0$ by padding the memory waveform with the final value attained at $t/M = 130$.

The memory accumulated during a fixed time interval decreases with increasing the mass ratio. Therefore, we scale the more extreme mass ratios shown in Fig. 3.2 by a factor of C_q, which we defined as the ratio between the final memory strain Δh_{20} for a BBH system with mass ratio $q = 10$ and the final strain for the other with mass ratio q:

$$C_q = \frac{\Delta h_{20}(q = 10)}{\Delta h_{20}(q)}. \tag{3.36}$$

Because the mass of the primary differs by at most ten percent in these different cases, the timescale of the merger is roughly the same for different mass ratios, and

once they are available, our work could be extended to include them.

corresponds to the short range around $t = 0$, where the memory accumulates most rapidly for the $q = 10$ case. This then allows one to see how much more memory accumulates during the inspiral phase as the mass ratio becomes more extreme. This trend in Fig. 3.2 becomes more pronounced for the common convention of EMRIs ($q \gtrsim 10^5$), where the memory offset during the merger and ringdown is suppressed from that during the inspiral by an additional factor of η, as discussed in Sec. 3.4.2. Figure 3.2 provides a visual justification of why we will use a high-order PN approximation for the memory accumulation during the adiabatic inspiral phase to compute the memory signal from EMRIs and ignore the contribution from the transition to plunge, the plunge, and the merger and ringdown.

3.5.2.2 Memory signal and memory offset

Finally, we can compute the memory signal after using the prescription outlined in Sec. 3.4.2.3 with the required oscillatory multipole modes at the relevant PN orders discussed in Sec. 3.4.2.4. The result can be written schematically as

$$h_{20} = \sum_{j=2}^{46} \sum_{k=0}^{k_{\max}(j)} h_{20,jk} v^j (\ln v)^k, \tag{3.37}$$

where $h_{20,jk}$ are numerical coefficients and $k_{\max}(j)$ is the maximum order of the powers of $\ln v$ that appear in the PN series at a given PN order $(j/2-1)$ relative to the leading v^2 term. Note that only the powers of v, not $\ln v$, determine the PN order. We do not list the coefficients $h_{20,jk}$ explicitly in this paper (there are 184 nonzero values at 22PN order), but we make this series data available in MATHEMATICA format as an ancillary file associated with the arXiv version of this paper and on Zenodo [134]. However, we do give the value of the final memory offset computed using all relevant $l \leq 25$ and up to 22PN order:

$$\Delta h_{20}^{(25,22)} \approx 0.102414\,\frac{\mu}{r}. \tag{3.38}$$

The notation with the superscript $(25, 22)$ indicates the multipole and PN accuracy of the expression, which we will use subsequently: i.e., $h_{20}^{(\bar{\ell},n)}$ denotes a memory signal evaluated with oscillatory l modes up to $l = \bar{\ell}$ with these modes at the necessary accuracy to compute h_{20} to 22PN order.[9] We discuss why we quote the result to six digits of accuracy in more detail in Secs. 3.5.2.3 and 3.5.2.4.

We also show how the memory signal accumulates as a function of time in Fig. 3.3. We use the method for evaluating $v(t)$ described in Sec. 3.4.2.3, which we then substitute into the final result in the form of Eq. (3.37). We show the memory signal starting from a time $t_i = -10^6 M/\eta$ until it reaches the ISCO, which is defined to be $t = 0$. For an EMRI with a primary BH mass of $10^6 M_\odot$ and a secondary BH with mass $10 M_\odot$, so that $\eta \approx 10^{-5}$, then the range of Fig. 3.3 covers about 5×10^{11} s or about 1.6×10^4 years. While this is the time scale over which the memory accumulates

[9] There is a potential ambiguity regarding the labeling of the multipole index $\bar{\ell}$ of the memory signal in the following sense. As described in Sec. 3.4.2.4, different elements of the calculation of h_{20} at a fixed PN order require a different number of spherical-harmonic modes. For example, the horizon term $(dE/dt)_H$ at 22PN order requires only harmonics up to $l = 11$, the luminosity at infinity requires modes up to $l = 24$, and the sum of modes that enter the memory expression in Eq. (3.8) requires modes up to $l = 25$. The label $\bar{\ell} = 25$, therefore, refers to the largest value of l needed in *some* aspect of the calculation to achieve the required PN order n; however, it does not imply that all the components of the calculation need to be evaluated at this multipole order to achieve the necessary PN accuracy.

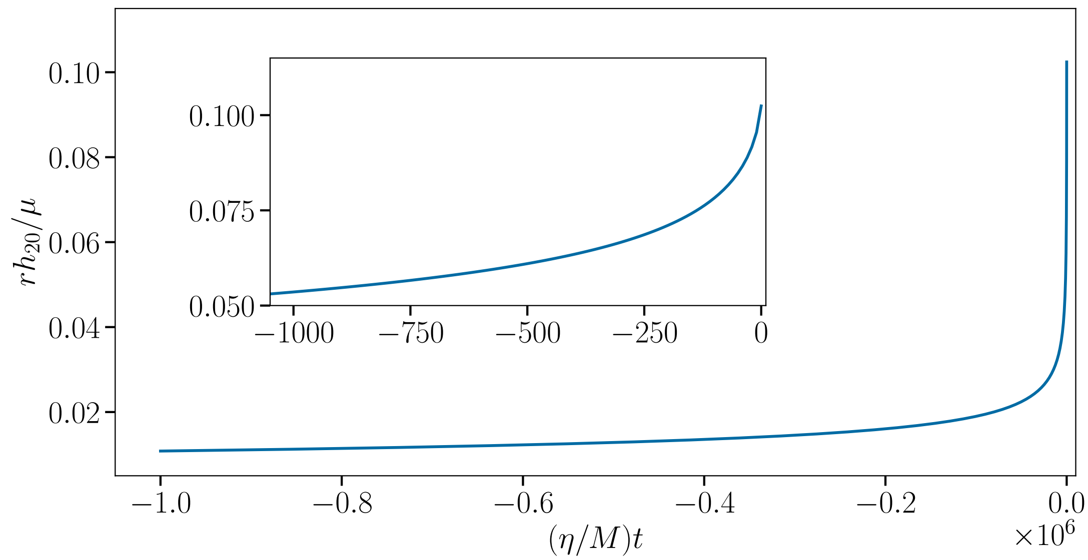

Figure 3.3: **Memory signal over different timescales**: The memory signal h_{20} is shown as a function of the normalized time $(\eta/M)t$ for an EMRI with nonspinning components. The memory is computed up to 22PN order relative to the Newtonian memory using oscillatory modes with $l \leq 25$. The main panel shows a time span running from $-10^6 M/\eta$ to 0, whereas the inset focus on a shorter range from $-10^3 M/\eta$ to 0. The signal stops at ISCO (the time $t = 0$), after which the memory accumulated is a small (order η^2) correction. Further discussion of the figure is given in the text of Sec. 3.5.2.2.

about ninety-percent of its final value, it is also a long timescale from the perspective of GW detection. We, therefore, also show the last $10^3 M/\eta$ of the signal in the inset, which corresponds to a roughly 16-year timescale for the same system (and is near the lower limit of the periods that pulsar timing arrays currently measure). It also illustrates that roughly half of the final memory offset accumulates in this last decade and a half. Note that for the same secondary BH of the same mass ($10M_\odot$) but a primary of mass ten times smaller (respectively, larger), the relevant time spans would be 100 times shorter (respectively, longer), namely order a month (respectively, a millennium) for the $10^3 M/\eta$ duration.

Given the large number of terms that contribute to the mode h_{20} at 22PN order [there are several thousand nonzero terms when evaluating the sums in Eq. (3.8)], and the fact that the PN series for the oscillatory waveform modes and the GW-luminosity have hundreds of terms when written in a form similar to the expression for the memory signal in Eq. (3.37), it is useful to perform some consistency checks of our result. The first was verifying that our result agrees with the linear in η terms in the 3PN expression derived by Favata [87] given in Eq. (3.15). Next, we also perform some convergence analyses, in which we study the contributions to the memory signal as a function of increasing multipole index l and PN order n. The results of these two investigations are discussed in next in Secs. 3.5.2.3 and 3.5.2.4 for the multipole and PN cases, respectively.

3.5.2.3 Memory signal at different multipole orders

To understand how much the different oscillatory multipoles contribute to the memory effect, we compare the complete memory signal at 22PN order with that computed by keeping l values less than or equal to a given value $\bar{\ell}$. We define an absolute difference

$$\delta_{\bar{\ell}} h_{20} = |h_{20}^{(25,22)} - h_{20}^{(\bar{\ell},22)}|, \tag{3.39}$$

which represents the contribution to the memory from the modes with $\bar{\ell} < l \leq 25$, at 22PN order. For ease of notation, we will often drop the $(25, 22)$ superscript on the memory signal with the highest PN order and number of multipoles, so that $h_{20} \equiv h_{20}^{(25,22)}$. It will also be helpful to compute a fractional contribution of these higher l modes, $\delta_{\bar{\ell}} h_{20}/h_{20}$, as well. We computed $\delta_{\bar{\ell}} h_{20}$ for all $\bar{\ell}$ values from 2 to 24, but show only a subset of them in the figures below. Because h_{20} depends on the GW luminosities at infinity and at the horizon, we also truncate the sum over l in Eqs. (3.32) and (3.33) at the corresponding value of $\bar{\ell}$.

As discussed in Sec. 3.4.2.4 and Footnote 9 especially, different terms in the computation of the memory require different numbers of multipoles to be accurate to a given PN order. For example, the power radiated down the future event horizon, $(dE/dt)_H$ only requires multipoles with $l \leq 11$. Thus, when we consider values of $\bar{\ell} < 11$, we sum over multipoles in Eq. (3.33) with $\bar{\ell}$ as the upper limit of the l sum, as noted above, but when $\bar{\ell}$ satisfies $\bar{\ell} \geq 11$, we use the full sums in Eq. (3.33) without any truncation nor with the addition of higher l terms that would be of a higher PN

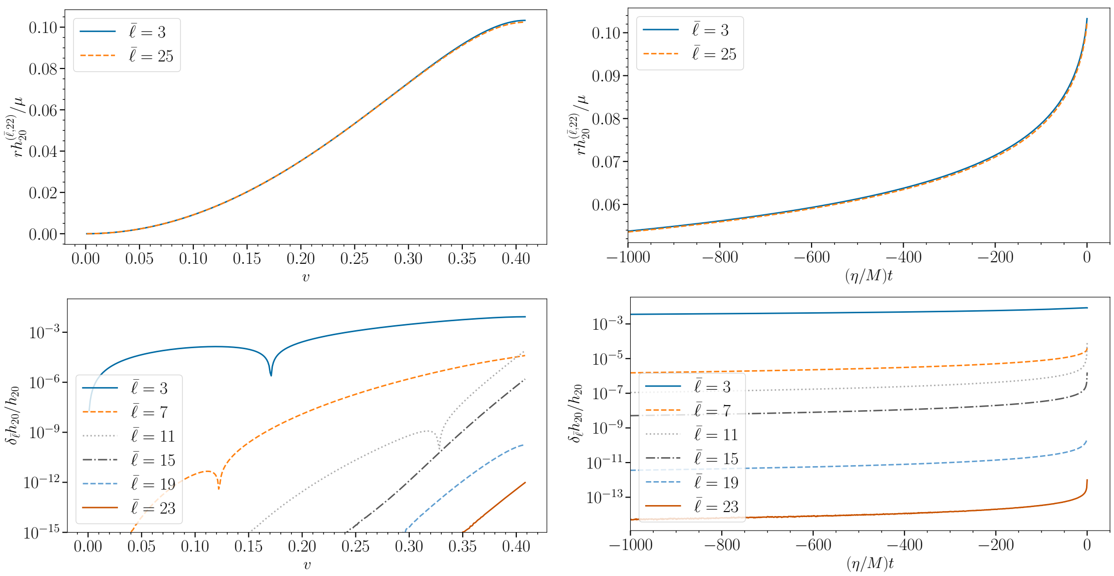

Figure 3.4: **Memory signals at different multipole orders versus velocity or time**: The top row shows the scaled memory signal $rh_{20}^{(\bar{\ell},22)}/\mu$ for different total numbers of multipoles included. Including just up to octopole order ($\bar{\ell} = 3$) is shown as solid blue curves and all multiples up to the 25-pole ($\bar{\ell} = 25$) is shown as dashed orange. The bottom row shows the fractional contribution to the full 22PN expression for the memory from modes with $l > \bar{\ell}$ for six cases, $\bar{\ell} = 3$, 7, 11, 15, 19, and 23, which are the solid dark-blue, dashed orange, dotted light-gray, dotted-dashed black, dashed light-blue, and solid brown curves, respectively. The left column shows these results versus velocity, with a lower velocity of 10^{-3} and a maximum velocity of $v_{\rm ISCO} = 1/\sqrt{6}$. The right column shows the quantities versus time, $(\eta/M)t$ over an interval spanning from -10^3 to 0. The initial time shown on the right corresponds to a velocity of $v \approx 0.25$. Further discussion of the implications of these figure panels are given in the text of Sec. 3.5.2.3.

order.

Our focus in this part will be on $h_{20}(v)$, for all values of v from 0 to $v_{\rm ISCO} = 1/\sqrt{6}$, on $h_{20}(t)$ for the last $10^3 M/\eta$ before the particle crosses the ISCO, and for the final memory offset Δh_{20}. We also include some intermediate results on the different multipolar contributions to $\dot{h}_{20}$ and dh_{20}/dv as a function of v in Appendix B.

In Fig. 3.4, we show in the top row the scaled memory signal $rh_{20}^{(\bar{\ell},22)}/\mu$, for different values of $\bar{\ell}$. The solid blue curve, $\bar{\ell} = 3$, contains only up to octopole order, and the dashed orange, $\bar{\ell} = 25$, contains all the relevant multipoles. The left panel shows this quantity versus velocity, from 10^{-3} to $v_{\rm ISCO}$, whereas the right panel shows it versus time over a range of $(\eta/M)t \in [-10^3, 0]$, which corresponds to a velocity range of approximately $[0.25, v_{\rm ISCO}]$. On the scale shown in the upper panels, the two different multipole orders are difficult to distinguish, which suggests that most of the memory signal comes from just the quadrupole and octopole oscillatory terms in the sum in Eq. (3.8).

To determine the contributions of higher modes more quantitatively, we show in the bottom row the fractional contribution of the modes with $l > \bar{\ell}$ to the full memory signal (rh_{20}/μ), which is the label on the vertical axis, $\delta_{\bar{\ell}} h_{20}/h_{20}$. The left again shows the results as function of velocity, whereas the right shows them as a function of time; the ranges of both velocity and time are the same as in the respective panels above. We show $\delta_{\bar{\ell}} h_{20}/h_{20}$ for six cases, $\bar{\ell} = 3$, 7, 11, 15, 19, and 23, which correspond to, respectively, the solid dark-blue, dashed orange, dotted light-gray, dotted-dashed black, dashed light-blue, and solid brown curves. The solid blue curves correspond to the relative contributions from hexadecapole and higher moments, and they demonstrate that the contributions from all modes higher than quadrupole and octopole terms contribute at most a few percent to the total memory signal. For certain applications, using just the quadrupole and octopole would be a justified approximation. Even a more stringent requirement, such as requiring the fractional contributions from higher multipoles be $\lesssim 10^{-6}$ for all velocities or times, requires computing up to $\bar{\ell} = 15$, rather than $\bar{\ell} = 25$.

As the PN approximation is most accurate in the limit of small v and higher multipoles enter at higher PN orders, it is to be expected that the contributions of higher multipoles become larger as v becomes larger (or t approaches 0, for the time-domain curves). The sharp v-shaped dips in the curves correspond to values of the velocity where the memory $rh_{20}^{(\bar{\ell},22)}/\mu$ crosses the full 22PN (rh_{20}/μ) expression (i.e., for different values of v or t, the contribution of the higher multipoles with $l > \bar{\ell}$ can be either larger or smaller than the full 22PN-accurate expression with all the relevant multipoles included). The time-domain results on the right illustrate the anticipated result that the majority of time during the EMRI's inspiral is spent at lower velocities (larger separations), so that the contributions from higher multipoles are most important for the shorter times near when the EMRI is close to the ISCO.

We next turn to the memory offset at the ISCO in Fig. 3.5. The value of $\Delta h_{20}^{(\bar{\ell},22)}$ is shown in the top panel for $\bar{\ell}$ values from $\bar{\ell} = 2$ up to $\bar{\ell} = 25$. Including higher multipoles tends to decrease the final value of the memory offset, though one should note that the overall vertical scale is small, and spans roughly a three percent relative difference in the values of the memory. The change at higher multipoles is hard to resolve on the scale of the top panel; thus, we also plot the fractional change $\delta_{\bar{\ell}}(\Delta h_{20})/\Delta h_{20}$ in bottom panel of Fig. 3.5, where $\delta_{\bar{\ell}}(\Delta h_{20})$ is defined analogously to that for the time- or velocity-dependent memory signals in Eq. (3.39). The range

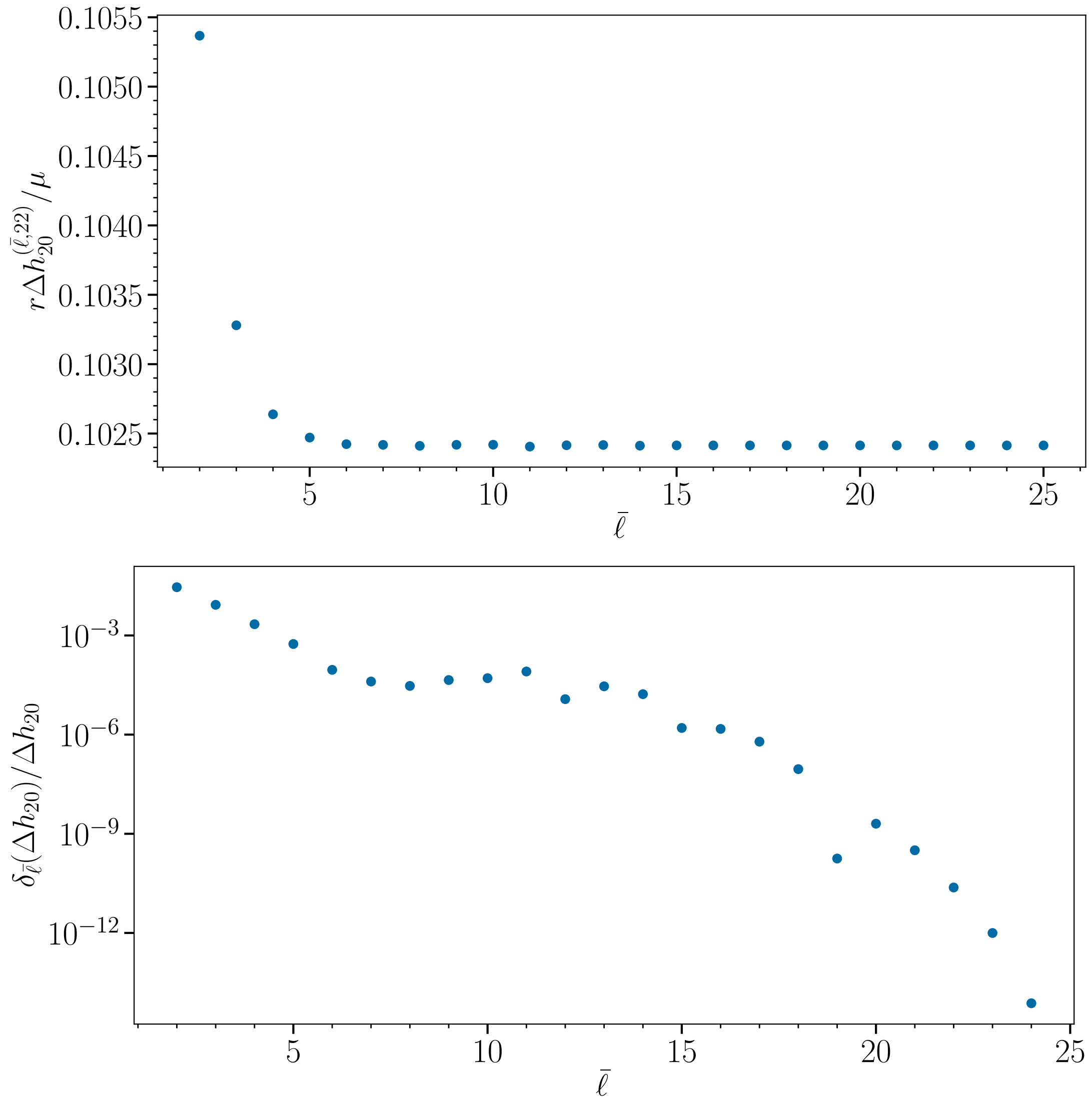

Figure 3.5: **Memory offset at ISCO versus highest multipole**: *Top*: The total accumulated EMRI memory offset, $\Delta h_{20}^{(\bar{\ell},22)}$, computed from different $l \leq \bar{\ell}$ modes is shown versus $\bar{\ell}$, for $\bar{\ell} = 2$ to $\bar{\ell} = 25$. The values for $\bar{\ell} = 3$ and $\bar{\ell} = 25$ are the same as those shown in Fig. 3.4 at the final velocity $v_{\rm ISCO} = 1/\sqrt{6}$. *Bottom*: The residual contribution to Δh_{20} is shown for different values of $\bar{\ell}$ as a fraction of the total Δh_{20}. Further discussion of these panels is given in the text of Sec. 3.5.2.3.

of $\bar{\ell}$ in the lower panel runs from $\bar{\ell} = 2$ to $\bar{\ell} = 24$, because $\delta_{\bar{\ell}}(\Delta h_{20})$ involves a difference from the highest ($\bar{\ell} = 25$) mode signal. The quadrupole contribution to the memory offset is about three percent larger than the final value with all multipoles, which could be sufficient for low-accuracy applications. The trace of the points has a nontrivial structure. There is a rapid decrease in the fractional contribution for the next few $\bar{\ell}$, up to about $\bar{\ell} = 7$. From $\bar{\ell} = 7$ until $\bar{\ell} = 14$, there a plateau-like structure in the trace of the points. This feature arises because successive multipole contributions have comparable amplitudes; however, above $\bar{\ell} = 14$ the amplitudes of successive multipoles decrease rapidly leading to the faster fall-off with $\bar{\ell}$. To obtain a more stringent accuracy requirement, then including a larger number of $\bar{\ell}$ is required.

Both Figs. 3.4 and 3.5 indicate that the highest l contributions to the memory at 22PN order contribute a relatively small amount to the memory signal and final offset. As there are $2l + 1$ modes for each l, the number of terms in Eq. (3.8) grows as a function of l, but the increase in terms is negated by a larger decrease in the amplitudes of these terms.

3.5.2.4 Memory signal at different post-Newtonian orders

Because the PN series for each of the oscillatory modes becomes increasingly complex as the PN order increases, it is also useful to determine the contributions of the higher PN terms to the total memory signal and the final memory offset, both to better understand the convergence of the PN series and to establish how many PN orders are required to compute the memory to a given precision.

Computing the memory to a given PN order also limits the multipolar order at which the memory is computed. Specifically, because an (l, m) mode of the strain has a time derivative that scales as $v^{l+\epsilon_p+3}$ for nonspinning binaries, then from Eq. (3.8), the lowest PN-order contribution from a given l to the memory signal h_{20} will enter at a PN order n of $\lfloor n \rfloor = l-3$ (for $l > 3$) and $\lfloor n \rfloor = l-2$ (for $l \leq 3$). The notation $\lfloor x \rfloor$ is the floor function of x, which is necessary to account for half-integer PN orders. Thus, multipoles with $l > \lfloor n \rfloor + 3$ (for $n > 1$) and $l > 2\lfloor n+1 \rfloor$ (for $n \leq 1$) will not contribute when working at n-PN order, which implies that limiting the PN order simultaneously limits the multipole order. To denote the required corresponding truncation in l, we define

$$\bar{\ell}_n \equiv \begin{cases} 2\lfloor n+1 \rfloor & n \leq 1 \\ \lfloor n \rfloor + 3 & n > 1 \end{cases} . \tag{3.40}$$

We then introduce a notation for the residual contribution to h_{20} from PN orders greater than n:

$$\delta_n h_{20} = |h_{20} - h_{20}^{(\bar{\ell}_n, n)}|, \tag{3.41}$$

where, as before, h_{20} is the 22PN memory computed from $\bar{\ell} = 25$ and $h_{20}^{(\bar{\ell}_n, n)}$ is the memory computed up to $\bar{\ell} = \bar{\ell}_n$ and accurate to a relative n-PN order.[10]

In Fig. 3.6, we show the memory-signal quantities analogous to those in Fig. 3.4, but the top panels now show the memory signal $h_{20}^{(\bar{\ell}_n, n)}$ and the bottom panels show the fractional contribution $\delta_n h_{20}/h_{20}$ for different PN orders n (rather than the fixed

[10]As discussed in Footnote 9, the $\bar{\ell}_n$ should be interpreted as the largest l value needed in Eq. (3.8), not that needed in the GW luminosities at the horizon or at infinity involved in the calculation of the memory, which require a lower order in l to achieve the same PN accuracy.

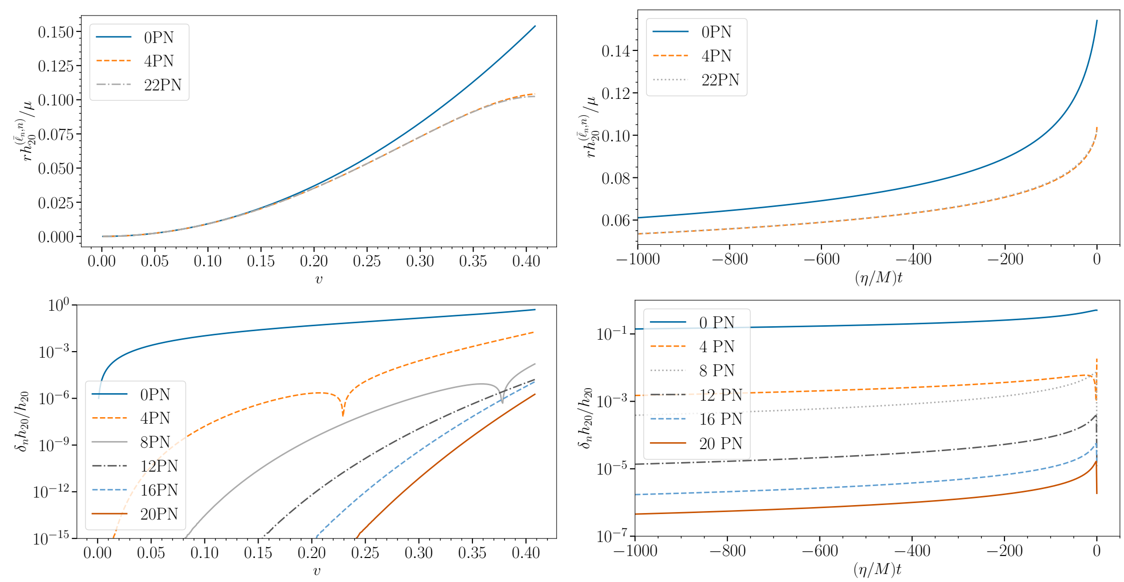

Figure 3.6: **Memory signals at different PN orders versus velocity or time**: This figure is analogous to Fig. 3.4, but the different curves now represent the memory signal computed up to a fixed PN order, which also requires truncating at a multipolar order given in Eq. (3.40). The top panels now show $h_{20}^{(\bar{\ell}_n,n)}$ at three different PN orders, $n = 0$, 4, and 22, which are the solid blue, dashed orange, and dotted-dashed light gray curves, respectively. The time or velocity ranges are identical to those in Fig. 3.4. The bottom panels now show the fractional contribution to the full memory signal h_{20} that come from PN orders greater than $n = 0$, 4, 8, 12, 16, and 20, which are depicted as the solid dark-blue, dashed orange, dotted light-gray, dotted-dashed black, dashed light-blue, and solid brown curves, respectively. Further discussion of the panels in this figure is given in the text of Sec. 3.5.2.4.

PN order of 22 and different multipole orders $\bar{\ell}$ in Fig. 3.4). The top panels show $h_{20}^{(\bar{\ell}_n,n)}$ versus velocity and time for the PN orders $n = 0$, 4, and 22 as the solid blue, dashed orange, and dotted-dashed light-gray curves, respectively. The ranges of velocity on the left and times on the right are identical to those in Fig. 3.4.

The Newtonian (0PN) curve is much larger than the 22PN memory signal, especially at velocities close to $v_{\rm ISCO}$. The reason for this can be understood from dh_{20}/dv, which is a linear function of v in the Newtonian limit, but which has a peak near $v \approx 0.3$ for higher PN orders (see Appendix B). The 4PN-accurate memory signal, however, differs from the 22PN signal by at most a few percent, which could make it a useful approximation for lower-accuracy applications. To understand the role of higher PN orders more quantitatively, it is useful to consider the bottom panels.

The bottom panels of Fig. 3.6 display the fractional contribution to the full memory signal h_{20} that come from a PN orders greater than $n = 0$, 4, 8, 12, 16, and 20, which are depicted as the solid dark-blue, dashed orange, dotted light-gray, dotted-dashed black, dashed light-blue, and solid brown curves, respectively. Similarly to the equivalent panels in Fig. 3.4, the errors grow as velocity and time increase; however, the contributions at larger v or $(\eta/M)t$ at higher PN orders are more significant than the high multipole terms. Even for $n = 20$ (PN orders greater than 20, namely), the contribution to the 22PN memory signal near $v_{\rm ISCO}$ is an order 10^{-6} correction (to be compared with the 10^{-12}-level contribution for $\bar{\ell} = 23$ in Fig. 3.4). Thus, neglecting higher PN orders is likely a larger source of error than is neglecting higher multipoles at a fixed PN order.

In Fig. 3.7, we show the analogue for different PN orders of the results in Fig. 3.5 for different multipole orders. Specifically, in the top panel, we show $r\Delta h_{20}^{(\bar{\ell}_n,n)}/\mu$, the final memory offset at the ISCO, as a function of different PN orders n, relative to the leading $O(v^2)$ term. We include all half PN orders from the Newtonian memory (0PN) and up to 22PN. Note that the 0.5PN has the same value as the 1PN, because there are no 0.5PN terms in the leading $l = 2$, $m = \pm 2$ oscillatory waveform, and the 1.5PN term is the same as the 1PN term, because of a cancellation between terms in the GW luminosity at infinity and the tail terms in the leading $l = 2$, $m = \pm 2$ oscillatory waveform when evaluating the memory integral (this is the reason there is no 1.5PN term in the 3PN memory signal in [87]). Above roughly 7PN order, the difference between the next successive PN orders is difficult to resolve on the scale of the figure.

To better understand the importance of the higher PN-order contributions to the memory, we show $\delta_n(\Delta h_{20})/h_{20}$, the relative residual contribution to the full 22PN-order memory from PN orders greater than n in the bottom panel of Fig. 3.7. Comparing this with the analogous bottom panel in Fig. 3.5, one again can see that higher PN contributions tend to be more significant than higher multipole-order terms. For example, considering the furthest-right point at 21.5PN order, this point can be interpreted as showing that all 22PN order terms contribute a relative contribution to the memory of order 10^{-6}. Following the general trend in the bottom panel of Fig. 3.7, we

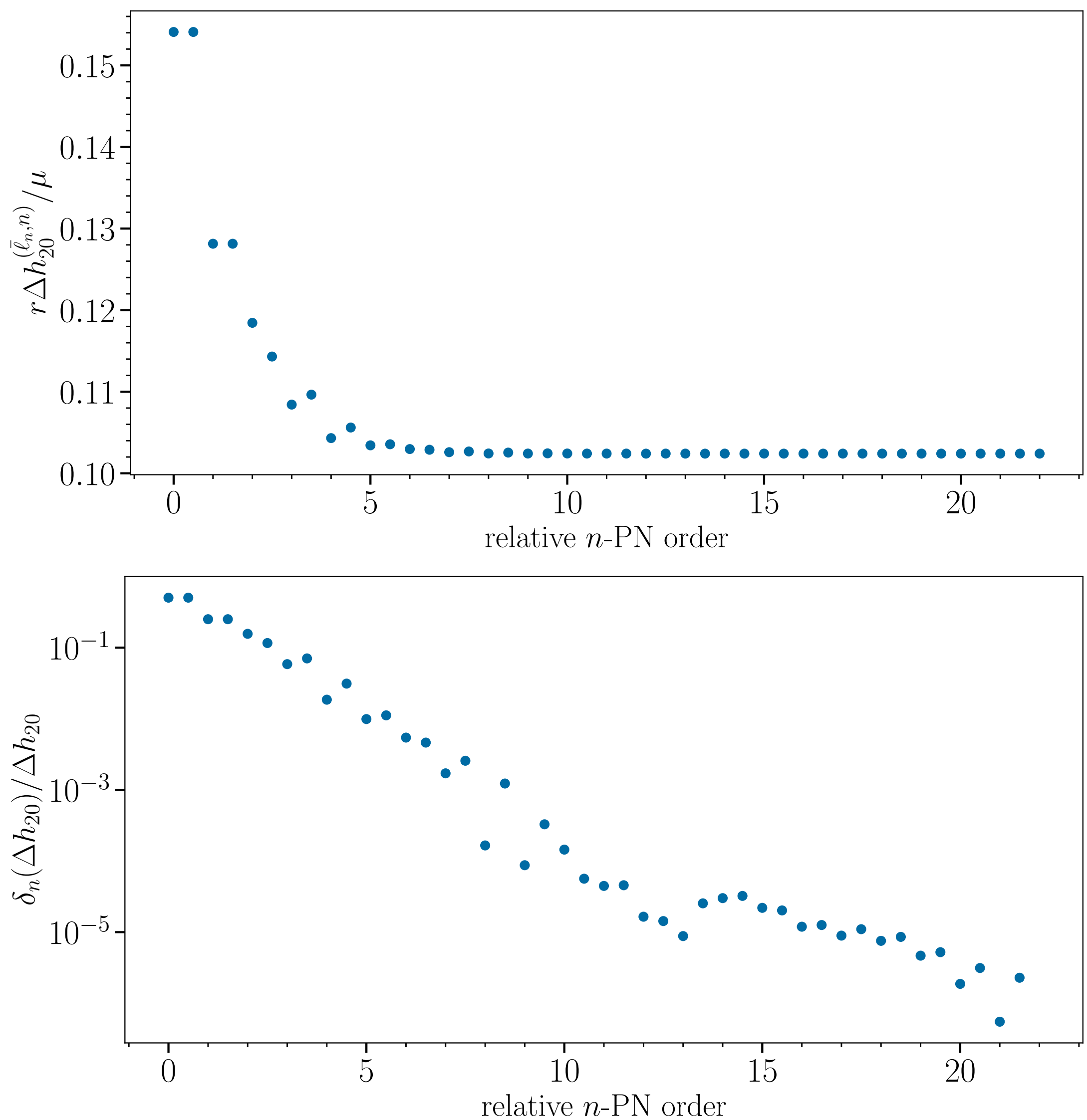

Figure 3.7: **Memory offset at ISCO versus PN order**: This figure is analogous to Fig. 3.5, but it focuses on the PN order rather than the multipole index. memory offset for different PN orders. *Top*: The total accumulated EMRI memory offset at ISCO, $r\Delta h_{20}^{(\bar{\ell}_n,n)}/\mu$ is shown for all n-PN orders from Newtonian ($n = 0$) to $n = 22$ in increments of 1/2 PN orders (single powers of v). The $n = 0$, 4 and 22 cases can be obtained from the final value of the memory as a function of velocity or time in Fig. 3.6. *Bottom*: The fractional residual contribution to the total accumulated EMRI memory from PN orders greater than n, $\delta_n(\Delta h_{20})/h_{20}$, which is computed from Eq. (3.41) evaluated at ISCO. Further discussion of this figure is given in the text of Sec. 3.5.2.4.

can estimate that the contributions of the 22.5PN and 23PN terms would be of order 10^{-7} at the smallest. This is the reason why we only quote six digits of accuracy for the coefficient of the final memory offset at ISCO in Eq. (3.38). The bottom panel also shows that if one requires a relative accuracy of the memory of order only 10^{-4}, one could instead work at 10PN order.

3.6 Fit for the final memory strain

Fit type	c_1	c_2	c_3	c_4	c_5	c_6
comp	0.113875	0.421532	2.44125	-5.90547	33.6768	-23.0496
EMRI	0.102414	0.770384	-1.70081	18.1139	-34.4687	52.7548

Table 3.3: Coefficients for the $\Delta h_{20}(\eta)$ polynomial fits in Eq. (3.42). Similar to the fits described in Table 3.2, these fits were constructed through a least-squares procedure using 50 BBH systems with mass ratios equally spaced in η for a range of mass ratios with the range of validity of the NRHybSur3dq8: $1 \leq q \leq 8$. The data in the first row, the comparable-mass-ratio fit labeled "comp," allows the term linear in η to be a free coefficient determined by least-squares fitting. The data in the second row, the fit labeled "EMRI," fixes the linear term to be the value computed in the EMRI limit, and given in Eq. (3.38). It is a five-parameter rather than a six-parameter fit.

In this section, we will make least-squared polynomial fits of the final memory strain for nonspinning BBH mergers that can be applied to different ranges of mass ratios.

The first fit is computed using just comparable mass-ratio ($1 \leq q \leq 8$) data from the hybridized 3PN memory signal with the NRHybSur3dq8 surrogate model (which follows the procedure described in Sec. 3.5.1). We compute the hybridized memory for 100 nonspinning BBH systems with mass ratios between $q = 1$ and $q = 8$, uniformly spaced in the symmetric mass ratio parameter η. We then fit fifty of the hybridized final memory values to a sixth-order polynomial in the symmetric mass ratio η, using a linear least-squares fit, and we use the other fifty to test how well the fit compares with points not used to construct the fit. We denote the coefficients in this fit by c_j, so that the final memory offset can be written as

$$\Delta h_{20}(\eta) = \frac{M}{r} \sum_{j=1}^{6} c_j \eta^j. \tag{3.42}$$

We do not include a constant (η-independent) term in the fit, because the memory offset should go to zero in the test-particle limit of $\eta \to 0$. The values of the coefficients are given in the first row of Table 3.3, and plots of the fit and its residual are given in Appendix C.

Note, however, that the comparable-mass fit has a term linear in η that does not agree with the numerical coefficient of the EMRI final memory offset in Eq. (3.38). This means that the fit, which we denote $\Delta h_{20}^{\mathrm{comp}}(\eta)$, should not be extrapolated to the EMRI limit, because it will be larger than the output of the EMRI calculation by roughly ten percent. We can construct another fit, which we denote by $\Delta h_{20}^{\mathrm{EMRI}}$, that has the correct EMRI limit if we instead fix the linear in η coefficient in the fit to be the value of the coefficient in Eq. (3.38), $c_1 = 0.102414$, (similarly to how we fixed the coefficient $c_0 = 0$); then, we can still fit for the remaining five coefficients. The results of the fit are shown in the second row of Table 3.3. This fit has the correct EMRI limit and agrees with the comparable-mass-ratio results (as we show in Fig. 3.8

below); thus, it could be used to interpolate between the EMRI and comparable-mass-ratio limits to give an estimate of the final memory strain from IMRI systems. We discuss this further in Appendix C.

To verify that the fit for the final memory offset that includes EMRI data works well in the comparable mass ratio regime, we plot the memory offset computed from the hybridized surrogate described in Sec. 3.5.1 with fifty comparable mass-ratio values between $q = 1$ and $q = 8$ that were not used to construct the fit. These are the blue points in the top panel of Fig. 3.8, and the solid orange curve is the polynomial fit that includes the EMRI data, $\Delta h_{20}^{\mathrm{EMRI}}$. The residuals for the same fifty values of η are shown in the bottom panel of Fig. 3.8. Adding the EMRI data slightly made very little difference in the accuracy of the fit (see Appendix C for the analogous figure for the fit $\Delta h_{20}^{\mathrm{comp}}$), but it now recovers the correct EMRI limit.

The difference between the two fits is shown in Fig. 3.9 for two different ranges of η. The top panel spans the same range of η as shown in Fig. 3.8; comparing with the bottom panel there, shows that the difference between the two fits is actually smaller than the difference between the hybridized-surrogate final memory offset and the fit. For comparable mass ratios, either fit would work equally well. The bottom panel of Fig. 3.9 shows the same difference as the top panel, but for a different range of η corresponding to mass ratios greater than $q = 8$. The fit with EMRI data consistently gives smaller values for the memory offset than the comparable-mass-only fit. Given the small number of NR simulations or second-order self-force calculations, there are

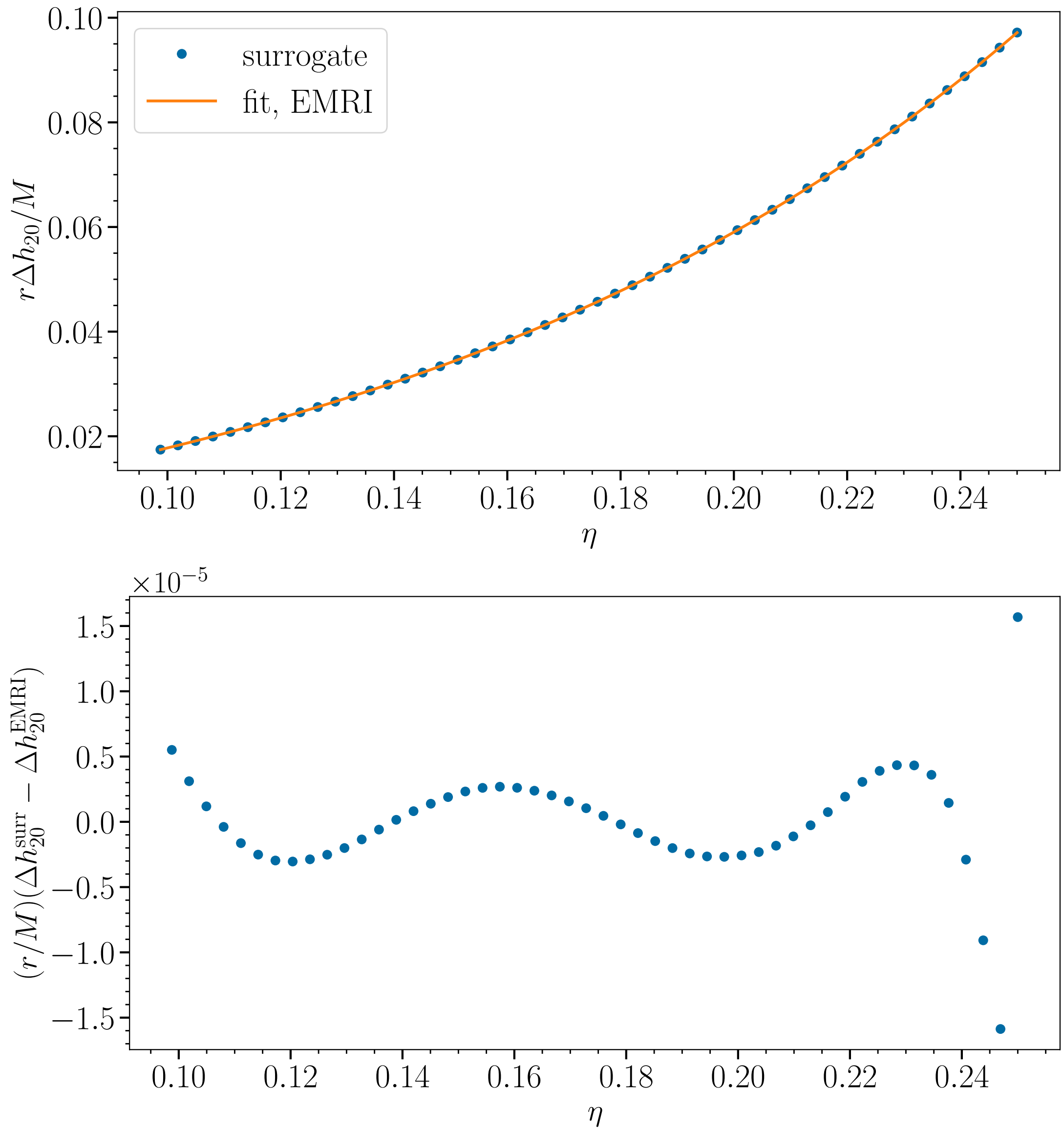

Figure 3.8: **Final memory offset fit and residual**: The blue points in both the top and bottom panels are for fifty values of the mass ratio between $q = 1$ and $q = 8$ that are distinct from the 50 values used to construct the fit. *Top*: The final memory computed from the hybridized surrogate in Sec. 3.5.1, $\Delta h_{20}^{\rm surr}$, is depicted with blue dots, whereas the fit for the final memory strain, $\Delta h_{20}^{\rm EMRI}$, is the solid orange curve. This fit fixes the term that is linear in η, so that it agrees with the EMRI calculation in that limit. *Bottom*: The blue points here depict the residual between the fit and the fifty values of the mass ratio that are used to test the fit. The maximum absolute error is of order 10^{-5} and occurs near the equal-mass-ratio case ($\eta = 0.25$); however given the smaller value of the memory at smaller η, the largest relative error occurs near the smallest values of η shown ($q = 8$).

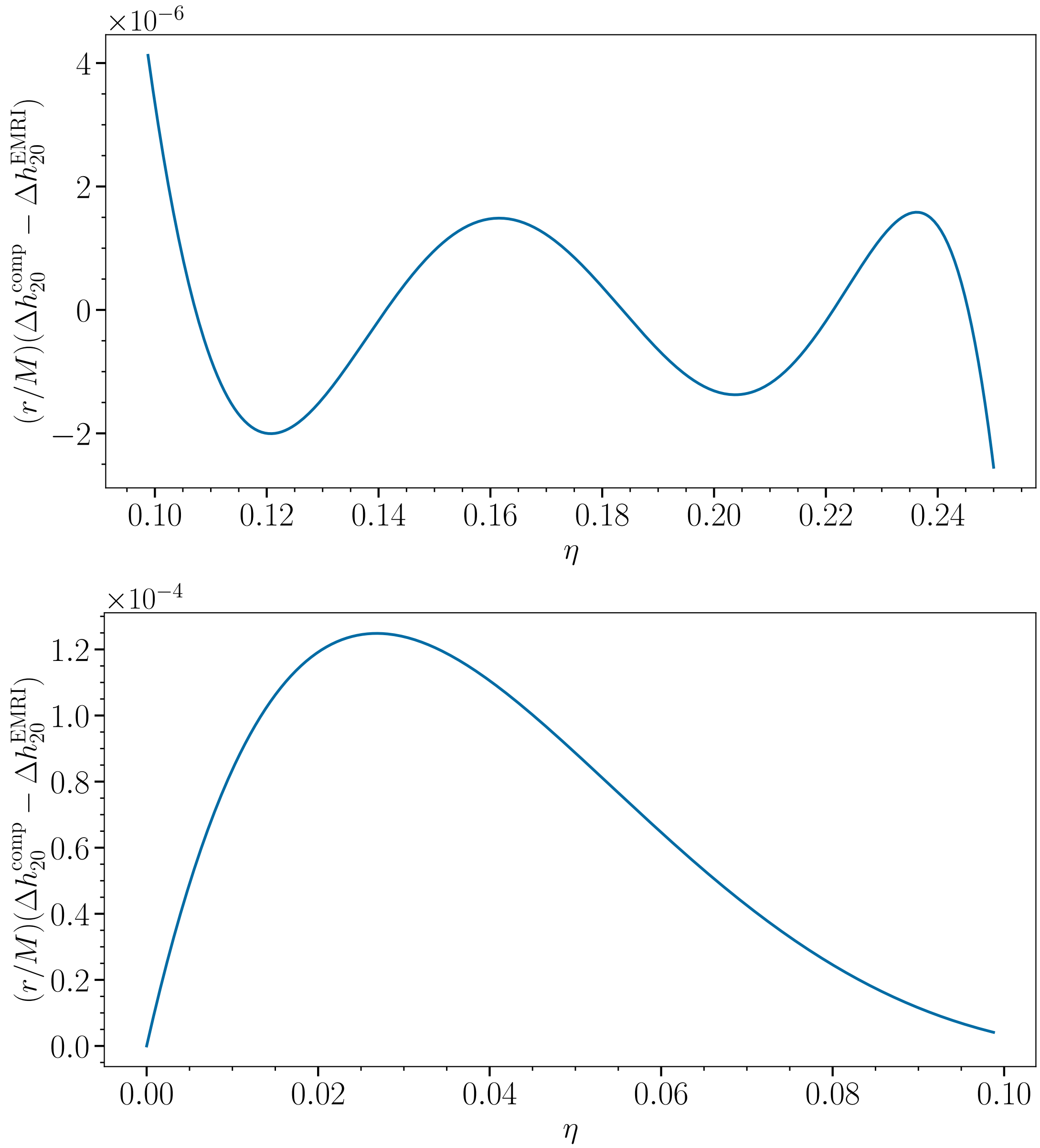

Figure 3.9: **Difference between the two polynomial memory-offset fits**: *Top*: The difference between the final memory strain fit $\Delta h_{20}^{\mathrm{comp}}$ that does not use the EMRI calculation, and the fit $\Delta h_{20}^{\mathrm{EMRI}}$ that does, is plotted versus η as a continuous function of η for values of q between 1 and 8. The differences between the two fits are smaller than the differences between the fit and the surrogate itself. *Bottom*: The same difference between the final memory strain fits $\Delta h_{20}^{\mathrm{comp}}$ and $\Delta h_{20}^{\mathrm{EMRI}}$ as in the top panel, but now for values of η in the IMRI regime. The comparable-mass fit consistently estimates a larger value of the final memory offset.

not many robust waveforms that could be used to determine which fit performs better in this regime. However, we do give in Appendix C an estimate for the memory in this regime by hybridizing the 3PN inspiral memory signal with the EMRISur1dq1e4.

3.7 Conclusions

This paper focused on modeling two aspects of the GW memory signal from nonspinning BBH mergers on quasicircular orbits:

(i) a 22PN-order calculation of the memory signal as a function of time or PN parameter for EMRIs

(ii) and two polynomial fits for the final memory offset after ringdown that used numerical-relativity data in one case, exclusively, and in the other case, NR data in combination with the result of the EMRI calculation.

We focused on the nonlinear memory effect, and we used the continuity equation for the supermomentum to compute the memory signal from oscillatory ($m \neq 0$) waveform modes without the memory effect. To obtain these results, we had to make some improvements in calculating the memory offset for comparable mass ratios systems, and we needed to perform a new calculation of the memory effect from EMRI systems.

For comparable mass ratios, most of the memory accumulates during the late inspiral, merger and ringdown; however, there is a nontrivial contribution from the earlier inspiral. While the numerical-relativity hybrid surrogate model NRHybSur3dq8 has oscillatory waveform multipoles that were hybridized with Effective-One-Body waveforms to produce arbitrary-length signals, evaluating the surrogate for long enough to resolve the offset accumulated during the early inspiral is time consuming. Instead, we hybridized the surrogate memory signal with a 3PN memory waveform. The 3PN waveform accounts for the memory accumulated at all times before the starting time of the surrogate signal, and it is faster to evaluate over long times. We computed a polynomial fit for this initial offset that must be added to the surrogate memory at the starting time (and also a polynomial fit for the PN time of coalescence, which is a free parameter in the 3PN waveform used in the hybridization). This fit is valid for mass ratios from one to eight. The data contained within this fit feeds into the other polynomial fits that we construct for the final memory offset.

For EMRIs, the memory offset, to leading order in the symmetric mass ratio, accumulates just during the inspiral. We computed the EMRI memory up to 22PN order relative to the Newtonian memory, using resummed, factorized waveforms for EMRIs that were computed by Fujita in [120] for a test particle orbiting a Schwarzschild BH on a quasicircular orbit. To have an expression for the $l = 2$, $m = 0$ harmonic of the memory at 22PN order, that requires using oscillatory modes up to $l = 25$. To compute the time dependence of the memory signal, we assumed that the test particle evolves adiabatically between geodesics with different energies such that the change in the geodesic energy is equal to the energy radiated to infinity and into the horizon (which made use of the GW luminosities computed in [120]). Our 22PN order result, restricted to 3PN order, agrees with the 3PN expression used in the hybridization procedure discussed above, when the 3PN expression is restricted to linear order in the symmetric mass ratio η.

We also investigated how the EMRI memory signal behaves as a function of the number of oscillatory modes l modes and of the post-Newtonian order. The largest contribution to the memory comes from the quadrupole ($l = 2$) and octopole ($l = 3$) oscillatory modes, with the higher oscillatory modes contributing at most a few percent to the total memory signal; for certain applications, the quadrupole and octopole modes would be sufficient. Different multipole orders did not contribute uniformly: the final memory offset, for example, adding modes between $l = 4$ and $l = 7$ caused the memory to converge rapidly towards the full $l = 25$ expression; including modes between $l = 8$ and $l = 14$ produced little change in the relative accuracy; the convergence sped up above this value of l again. We performed a similar analysis of the contributions of different PN orders to the full 22PN expression (though note that truncating the series at lower PN orders also truncates the highest multipole order needed in the computation). The memory signal and the final memory offset do converge to the 22PN expression as higher PN orders are included, but the rate at which it converges as a function of PN order is slower than the analogous convergence with the highest multipole l included in the calculation. In this sense, ignoring higher PN orders has a larger effect than ignoring higher l oscillatory modes.

We then used the EMRI and comparable mass calculations to construct two polynomial fits in η for the final memory strain for nonspinning BBH mergers. Both fits used comparable mass-ratio data with mass ratios within the range $1 \leq q \leq 8$, which were computed from our hybridization of numerical-relativity surrogate. The first fit used only this data, and it performed well in its range of validity, but it overestimated the memory in the EMRI limit when it was extrapolated to small η. To incorporate the result of the EMRI calculation into a new fit, we used the same comparable-mass data, but we fixed the coefficient linear in η to the computed value of the EMRI memory offset. This polynomial fit has the correct EMRI limit and it performed similarly to the first fit in the comparable-mass regime. Incorporating information about EMRIs into the fit that largely used comparable mass-ratio data from numerical-relativity simulations allowed the memory to be interpolated, rather than extrapolated into the IMRI regime, where there are fewer reliable waveforms (though see, e.g., [135–138] for notable exceptions).

We can foresee a few applications of the results of this paper. When it is necessary to know the amount of memory accumulated during the inspiral (for example, in setting initial data in Cauchy-characteristic-extraction simulations) the polynomial fit of the memory accumulated during the early inspiral could be useful. While EMRIs are a target GW source for the LISA detector, the long timescale over which the memory accumulates for most EMRI systems would make the signal fall out of LISA's frequency range; however for lighter IMRI systems the frequency ranges are more consistent with what LISA could measure. The study [139] focused on very light IMRIs (which were challenging for LIGO and Virgo to detect, but could be detected by Einstein Telescope and Cosmic Explorer); it would be interesting to revisit this study in the context of LISA. To use the analytical waveforms here, however, it would be useful to generalize the calculations to spinning primaries, and also eccentric systems, to have a wider coverage of the parameter space of binaries.[11] The fit for the memory offset could also be useful for interpreting the results of a potential pulsar-timing-array detection of a burst with GW memory, as discussed in Sec. 3.2. Finally, this paper is a first step towards developing a stand-alone waveform model for the nonlinear gravitational wave memory effect for BBH mergers. The fit will feed into the construction of both time- and frequency-domain waveform models, which could be used for searches for the memory effect.

Acknowledgments

A.E. and D.A.N. acknowledge support from NSF Grants No. PHY-2011784 and No. PHY-2309021. They thank Niels Warburton for discussions about the EMRI calculation during the early stages of this project. They also thank an anonymous referee for helpful feedback on this paper.

[11] While this paper was under review, a pre-print [140] appeared which computed the memory for spinning binaries and included post-adiabatic corrections, too. Reference [140] also provides further discussion about computing the memory signal from EMRIs on eccentric orbits.

Chapter 4

Waveform models for the gravitational-wave memory effect: II. Time-domain and frequency-domain models for nonspinning binaries

A. Elhashash and D. Nichols, arXiv:2504.18635,

4.1 Abstract

The nonlinear gravitational-wave (GW) memory effect—a permanent shift in the GW strain that arises from nonlinear GW interactions in the wave zone—is a prediction of general relativity which has not yet been observed. The amplitude of the GW memory effect from binary-black-hole (BBH) mergers is small compared to that of primary (oscillatory) GWs and is unlikely to be detected by current ground-based detectors. Evidence for its presence in the population of all the BBH mergers is more likely, once thousands of detections are made by these detectors. Having an accurate and computationally efficient waveform model of the memory signal will assist detecting

the memory effect with current data-analysis pipelines. In this paper, we build on our prior work to develop analytical time-domain and frequency-domain models for the dominant nonlinear memory multipole signal ($l = 2$, $m = 0$) from nonspinning BBH mergers in quasicircular orbits. The model is calibrated for mass ratios between one and eight. There are three parts to the time-domain signal model: a post-Newtonian inspiral, a quasinormal-mode-based ringdown, and a phenomenological signal during the late inspiral and merger (which interpolates between the inspiral and ringdown). The time-domain model also has an analytical Fourier transform, which we compute in this paper. We assess the accuracy of our model using the mismatch between our waveform model and the memory signal computed from the oscillatory modes of a numerical-relativity surrogate model. We use the advanced LIGO sensitivity curve from the fourth observing run and find that the mismatch increases with the total mass of the system and is of order 10^{-2}–10^{-4}.

4.2 Introduction

Gravitational waves (GWs) from the mergers of nearly 100 binary black holes have been announced by the Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo collaborations [51, 52, 141]. Over 200 more detection candidates have been announced during the fourth observing run of the LIGO-Virgo-KAGRA (LVK) collaboration to date [90]. The detections have been used to test the predictions of general relativity in the strong-gravity and high-luminosity regime of the theory [94–96]. The works [94–96] contain a large suite of tests, most of which look for parametrized (or, in some cases, unparametrized) deviations from general relativity, which are often described as being "agnostic" about the underlying theory that produces any such deviations.

An alternate approach to testing general relativity is to verify that all components of the GW signal predicted by general relativity do indeed appear in the observed signals. One such example of such a test was the confirmation of the presence of higher harmonics (i.e., multipole moments) of the dominant quadrupolar GW signal with the exceptional event GW190814 [142]. As GW measurements improve, more of these predictions of general relativity will become accessible to observational study. One such prediction of general relativity that has not yet been detected, is a nonlinear gravitational effect that arises from a nonlinear interaction of GWs in the wave zone far from an isolated source. It is known as the nonlinear or Christodoulou GW memory effect [21, 23]. This paper will focus on this GW phenomenon and aspects of the effort to detect this effect.

The nonlinear GW memory effect (as well as the linear memory effect [19]) produces a lasting offset in the GW strain following a burst of GWs from an isolated system. Many decades prior to the first detection of gravitational waves, the memory effect was identified as a possible source that GW interferometers could detect [27–29]. Interest in detecting the memory effect has grown because of the detection of GWs by LIGO and Virgo, and because of theoretical investigations of the memory effects that showed its close connections to infrared properties of gravitational physics. Specifically, the realization that the memory effect is closely connected to the Bondi-Metzner-Sachs supertranslation symmetries [3–5] (and their conserved charges the supermomentum [108]) and Weinberg's soft graviton theorem [68] (see, e.g., [2, 24, 99, 100]) in an "infrared triangle" has provided more compelling theoretical reasons to search for the GW memory effect.

The limited frequency response of ground-based GW interferometers makes the lasting (i.e., zero-frequency) memory signal challenging to detect with interferometers, such as LIGO and Virgo. There is, however, a time-dependent signal (which contains signal power at higher frequencies), which is associated with the nonlinear GW memory effect and which GW interferometers can measure more readily [35]. Its detection will be challenging because the memory signal still remains small compared to both the dominant quadrupolar waves and even some of the higher harmonics that LIGO and Virgo have measured. Thus, it is unlikely that the LVK collaboration will detect the memory effect from an individual black hole merger [31], given what is currently known about the population of merging binary black holes [91, 92] and the detectors' observation plans and timelines to be upgraded [36].

It is much more likely, however, that the presence of the nonlinear memory effect can be inferred in the population of all the binary-black-hole (BBH) mergers [32–35] measured by the LVK collaboration. A Bayesian search for the nonlinear memory effect already has been implemented and applied to the data from the first three observing runs [33,101,102]. No significant evidence for the nonlinear memory effect in the population of BBH mergers has been found, which is consistent with forecasts [33–35]. The searches for the GW memory effect require a well-defined notion of the nonlinear GW memory signal from a BBH merger, as discussed above. Computing this signal from a set of multipole moments of the GW strain requires numerically differentiating the modes, integrating different products of the modes, and summing the different modes to obtain the final result. While the works [33, 101, 102] have shown that it is possible to perform such searches with existing GW memory waveform models, it would be advantageous for the search for the GW memory effect to have a waveform model that does not require as many steps to compute.[1]

In the paper [143] (henceforth, Paper I), the authors made a first step towards building such a model. Paper I contained a fit of the final memory strain offset, which included results from a high post-Newtonian (PN) order calculation in the extreme mass-ratio limit. These results will be used in this current work. The main aim of this paper is to build a time-domain model for the $l = 2$, $m = 0$ spin-weighted spherical-harmonic mode of the memory signal from nonspinning BBH mergers. We use a PN model of the memory signal during the inspiral stage and a superposition of products of quasinormal modes (QNMs) during the ringdown. Unlike the minimal waveform model for nonspinning equal-mass binaries in [30], we need to add a phenomenological

[1] Having a stand-alone model for the GW memory signal can also allow the signal to be evaluated more efficiently. Numerically differentiating and integrating the oscillatory ringdown modes to compute the memory signal requires sampling in time many times per orbital period to accurately evaluate the derivatives. The memory model for nonprecessing binaries evolves on a slower timescale (namely, that of radiation reaction), so its waveform will not need to be evaluated as frequently in time as the oscillatory modes to obtain the same accuracy requirement on the signal.

"intermediate" model that bridges between the PN inspiral model and the QNM ringdown model. Similar to the minimal waveform model of [30], we find that the time-domain model has an analytic Fourier transform in terms of transcendental and other special functions. This allows us to obtain an analytical frequency-domain model of the nonlinear GW memory effect.

While this work was in progress, a time-domain numerical-relativity (NR) surrogate model using Cauchy-characteristic extraction (CCE) [46] and a phenomenological frequency-domain waveform model of the $l = 2$, $m = 0$ spin-weighted spherical-harmonic mode of the gravitational waveform [111] were developed. Both of these waveforms contain the memory signal as part of the $l = 2$, $m = 0$ mode, but the mode also includes other features in this signal, such as the QNMs of the remnant Kerr black hole formed during the merger. These QNMs are not related to the nonlinear GW memory effect, so they would need to be removed from these models, if they were to be used in searches for the GW memory effect alone. Our time-domain and its Fourier transform does not need any such modifications to be used in searches for the GW memory signal.

4.2.1 Summary and organization of this paper

We close the introduction by summarizing the organization and main results in this paper. First, in Sec. 4.3, we review some results from Paper I, which describe how we compute the memory signal from multipole moments of the gravitational-wave strain and how we compute the late-time memory offset strain. In Sec. 4.4, we discuss the modeling techniques we use for the inspiral, ringdown and intermediate stages of the memory signal, respectively. The inspiral model uses results from PN theory, the ringdown model uses multimode QNM fitting, and the intermediate model is a phenomenological approach that enforces continuity and some smoothness between the inspiral and ringdown. The resulting time-domain model has a relative error of a few percent compared with a calculation of the memory signal from a NR surrogate model. Section 4.5 covers several topics related to the frequency-domain representation of our time-domain model: namely, the analytical calculation of the Fourier transform of the time-domain model, a method to remove windowing artifacts from the calculation of the fast Fourier transform (FFT) of the time-domain model, and finally the performance of the model, as quantified by the mismatch. We conclude in Sec. 4.6. Several more technical results related to the phase of the complex frequency-domain signal, a more general ringdown model, the fitting coefficients in parts of our model, and the calculation of the FFT are given in four Appendices D–G.

4.3 Review of results from Paper I

We begin by reviewing some aspects of the calculation of the memory signal from Paper I. Specifically, we summarize the expansion of the memory signal in spin-weighted spherical harmonics, and computation of the memory signal from the NR surrogate model. We note the small differences from Paper I in some aspects of these calculations where applicable.

4.3.1 Multipolar expansion of the memory

We start by writing the multipolar expansion of the complex strain $h \equiv h_+ - ih_\times$ in terms of spin-weighted spherical harmonics. We denote the multipole moments by h_{lm}, so that

$$h \equiv h_+ - ih_\times = \sum_{l=2}^{\infty} \sum_{m=-l}^{l} h_{lm} ({}_{-2}Y_{lm}) \,. \tag{4.1}$$

The multipoles h_{lm} are functions of the retarded time u, and the spin-weighted spherical harmonics are functions of the polar and azimuthal angles (θ, ϕ), respectively. The spin-weighted spherical harmonics with fixed spin s form a complete basis for functions on the two-sphere of spin weight s. They are orthonormal,

$$\int d^2\Omega \, ({}_s\bar{Y}_{lm})({}_sY_{l'm'}) = \delta_{ll'}\delta_{mm'}, \tag{4.2}$$

where the overline represents complex conjugation.

Evaluating the multipole moments for the GW memory signal involves computing the integral of three spin-weighted spherical harmonics (see, e.g., [25], for more details). We use the notation of [25] for these integrals:

$$C_l(s', l', m'; s'', l'', m'') \equiv \int d^2\Omega \, ({}_{s'+s''}\bar{Y}_{lm'+m''})({}_{s'}Y_{l'm'})({}_{s''}Y_{l''m''}). \tag{4.3}$$

These coefficients are nonvanishing when the index l is in the set Λ,

$$\Lambda \equiv$$

$$\{\max(|l' - l''|, |m' + m''|, |s' + s''|), ..., l' + l'' - 1, l' + l''\}. \tag{4.4}$$

The coefficients can also be written in terms of Clebsch-Gordon coefficients:

$$C_l(s', l', m'; s'', l'', m'') = (-1)^{l+l'+l''} \sqrt{\frac{(2l'+1)(2l''+1)}{4\pi(2l+1)}} \times \langle l', s'; l'', s''|l, s'+s''\rangle \langle l', m'; l'', m''|l, m'+m''\rangle . \tag{4.5}$$

We use the same conventions for the Clebsch-Gordon coefficients as those implemented in MATHEMATICA. One useful transformation property of the coefficients $C_l(s', l', m'; s'', l'', m'')$ that we will use is

$$C_l(s', l', m'; s'', l'', m'') = (-1)^{l+l'+l''} C_l(s', l', -m'; s'', l'', -m'') . \tag{4.6}$$

It was shown in [35], for example, that the nonlinear memory signal, when written in terms of the GW strain multipole moments h_{lm}, is given by

$$h_{lm}^{\rm mem}(u) = r\sqrt{\frac{(l-2)!}{(l+2)!}} \sum_{l',l'',m',m''} \int_{-\infty}^{u} du'\, \dot{h}_{l'm'} \dot{\bar{h}}_{l''-m''} (-1)^{m''} C_l(-2, l', m'; 2, l'', m''). \tag{4.7}$$

The sum over the indices l', l'', m' and m'' in Eq. (4.7) must satisfy the constraints that l', $l'' \geq 2$ as well as $|m'| \leq l'$ and $|m''| \leq l''$. For fixed values of l and m on the left-hand side of Eq. (4.7), the coefficients $C_l(-2, l', m'; 2, l'', m'')$ in the sum will only be nonzero when $m = m' + m''$ as well as when l, l' and l'' satisfy the relationships in Eq. (4.4).

As in Paper I, we will specialize to the $l = 2$, $m = 0$ mode of the memory signal.

	$l=2$	$l=3$	$l=4$	$l=5$
m values (memory model)	$\{\pm1,\pm2\}$	$\{\pm2\}$	—	—
m values (NR surrogate)	$\{\pm1,\pm2\}$	$\{\pm1,\pm2\pm3\}$	$\{\pm2,\pm3,\pm4\}$	$\{\pm5\}$

Table 4.1: The m values (for each l) of the oscillatory modes that use used in constructing the time-domain memory model (first row) and that are generated by the NRHybSur3dq8 surrogate model (second row).

This requires that $m'' = -m'$, and Eq. (4.7) reduces in this case to

$$h_{20}^{\rm mem}(u) = \frac{r}{2\sqrt{6}} \sum_{l',l'',m'} (-1)^{m'} C_l(-2,l',m';2,l'',-m') \int_{-\infty}^{u} du'\, \dot{h}_{l'm'}\dot{\bar{h}}_{l''m'}. \tag{4.8}$$

For the nonspinning binaries that we will consider in this paper, the complex conjugate of h_{lm} is given by $\bar{h}_{lm} = (-1)^l h_{l-m}$. We can use this transformation of h_{lm} for nonspinning binaries along with the transformation of the coefficients $C_l(s',l',m';s'',l'',m'')$ in Eq. (4.6) to rewrite the sum over m' in Eq. (4.8) to be over only positive values of m':

$$h_{20}^{\rm mem}(u) = \frac{r}{\sqrt{6}} \sum_{l',l''} \sum_{m'=1}^{l'} (-1)^{m'} C_l(-2,l',m';2,l'',-m') \int_{-\infty}^{u} du'\, \Re\Big[\dot{h}_{l'm'}\dot{\bar{h}}_{l''m'}\Big]. \tag{4.9}$$

Because the integrand is the real part of $\dot{h}_{l'm'}\dot{\bar{h}}_{l''m'}$, this implies that the memory signal $h_{20}^{\rm mem}$ is also real.

4.3.2 Surrogate memory signal and hybridization

As in Paper I, we will construct our memory model using the spin-weighted spherical-harmonic moments h_{lm} that are output from the NR hybrid surrogate model NRHybSur3dq8 [47]. To obtain the initial offset for the memory during the long inspiral, we found it more efficient to hybridize the memory computed from the surrogate model with a PN memory waveform than to evaluate the surrogate model for very long periods of time. We perform a hybridization procedure similar to the one done in Paper I, except here we hybridize using the 3.5PN memory waveform computed in [140], rather than the 3PN memory waveform used in Paper I. The PN memory waveforms are written in terms of the PN parameter x, which is defined to be

$$x \equiv (M\Omega)^{2/3} \,. \tag{4.10}$$

We use the notation $M = m_1 + m_2$ for the total mass of the binary, where the primary mass is m_1 and secondary mass is m_2, and the orbital frequency is Ω. As in Paper I, we write the PN parameter x in terms of the coordinate time t at Newtonian order, for simplicity:

$$x(t) = \frac{1}{4}\left[\frac{\eta}{5M}(t_c - t)\right]^{-1/4} . \tag{4.11}$$

The parameter t_c is the time of coalescence. The symmetric mass ratio η has several equivalent expressions, which can be given in terms of the total mass M and the individual masses (m_1 and m_2), the mass ratio ($q = m_1/m_2$), or the reduced mass ($\mu = m_1 m_2/M$):

$$\eta = \frac{q}{(q+1)^2} = \frac{m_1 m_2}{M^2} = \frac{\mu}{M} \,. \tag{4.12}$$

The 3.5PN memory waveform, when written in terms of the PN parameter x, is given by the lengthy expression in [140]:

$$h_{20}^{\rm PN}(x) = \frac{4M}{7r}\sqrt{\frac{5\pi}{6}}\eta x\left\{1 + x\left(-\frac{4075}{4032} + \eta\frac{67}{48}\right) + x^2\left(-\frac{151877213}{67060224} - \eta\frac{123815}{44352}\right.\right.$$

$$+\eta^2\frac{205}{352}\bigg) + \pi x^{5/2}\bigg(-\frac{253}{336}+\eta\frac{253}{84}\bigg) + x^3\bigg[-\frac{4397711103307}{532580106240}$$
$$+\eta\bigg(\frac{700464542023}{13948526592}-\frac{205}{96}\pi^2\bigg)+\eta^2\frac{69527951}{166053888}+\eta^3\frac{1321981}{5930496}\bigg]$$
$$+\pi x^{7/2}\bigg(\frac{38351671}{28740096}-\eta\frac{3486041}{598752}-\eta^2\frac{652889}{598752}\bigg)\bigg\}. \tag{4.13}$$

Following the procedure in Paper I, we hybridize over a time interval of time $t \in [t_1, t_2]$. We again choose the time interval of the hybridization to be between $t_1 = -5000M$ and $t_2 = -4000M$ (which was the time interval over which the oscillatory modes of the surrogate were hybridized). To hybridize, we allow t_c to be a free parameter in the PN memory waveform, and we add a constant h_0 to the surrogate memory waveform, which represents the memory signal accumulated from before the initial time of $t/M = -10^4$, at which we first evaluate the surrogate model. The main difference between this paper and Paper I, is that in addition to computing the memory signal from all the oscillatory waveform modes in the surrogate model as in Paper I (see the second row of Table 4.1), here we also hybridize the surrogate with the three lowest l and m modes that contribute to the memory signal (see the first row of Table 4.1). This latter choice is related to our modeling of the ringdown part of the signal, which we discuss in more detail in Sec. 4.4.2.

To perform the hybridization we minimize the following cost function involving the surrogate waveform $h^{(A)} = h_{20}^{\rm surr}$ and the PN waveform $h^{(B)} = h_{20}^{\rm PN}$ for a specific

(l, m) mode cases	c_1	c_2	c_3	c_4	c_5	c_6
Memory model	0.102414	0.824195	-2.3413	22.486	-58.276	105.885
NR surrogate	0.102414	0.770384	-1.70081	18.1139	-34.4687	52.7548

Table 4.2: Coefficients for the $\Delta h_{20}(\eta)$ polynomial fits in Eq. (4.15). These fits were constructed through a least-squares procedure using 50 BBH systems with mass ratios equally spaced in η for a range of mass ratios with the range of validity of the NRHybSur3dq8: $1 \leq q \leq 8$. The data in the first row are for the memory fit computed from the three plus-minus pairs of oscillatory modes in the first row of Table 4.1. The data in the second row are for the memory fit computed from all oscillatory modes available in the NR surrogate model. They are similar to the results given in Paper I, except for the fact that a 3.5PN memory signal was used here rather than the 3PN signal used in Paper I. Both fits have the linear term fixed to be the value computed in the EMRI limit, which was derived in Paper I.

mass ratio q:

$$C_q[h_{(A)}, h_{(B)}] \equiv \frac{\displaystyle\int_{t_1}^{t_2} dt\, |h_{20}^{(A)}(t;q) - h_{20}^{(B)}(t;q)|^2}{\displaystyle\int_{t_1}^{t_2} dt\, |h_{20}^{(A)}(t;q)|^2} . \tag{4.14}$$

The free parameters that can be tuned in this cost function are t_c and h_0. We use the `SciPy` *minimize* function to perform the optimization and to obtain estimates of the best-fit parameters h_0 and t_c for each mass ratio.

As in Paper I, we use the values of t_c and h_0 to evaluate the final memory offset at 50 equally spaced points in η. We similarly assume a sixth-order polynomial function in the symmetric mass ratio η as our fitting ansatz. We do not include an coefficient with no powers of η, and we fix the coefficient linear in η to be given by the value of the memory offset in the extreme mass-ratio limit, which was computed in Paper I. This leaves five undetermined parameters in our fitting functions. Specifically, we write the final memory offset as

$$\Delta h_{20}(\eta) = \frac{M}{r} \sum_{j=1}^{6} c_j \eta^j \,, \tag{4.15}$$

where $c_1 = 0.102414$ is fixed. The values of the coefficients c_j from the least-squares fit, using just l and m modes in the first line of Table 4.1, are summarized in the first row of Table 4.2. The coefficients of the fit using all surrogate modes (as in Paper I, except using the 3.5PN memory signal to hybridize) is given in the second row of Table 4.2 for comparison. Using the 3.5PN rather than the 3PN waveform for performing the hybridization did not change the values of the coefficients c_j at the accuracy at which we give the results. Note that there is a more significant difference in the coefficients from using just the modes in the first line of Table 4.1 rather than the modes in the second line.

4.4 Time-domain model

As discussed in Sec. 4.2, we divide the memory signal into three parts: an inspiral part (during which the memory slowly grows in amplitude on the inspiral time scale), an intermediate part (which includes the late inspiral and merger), and ringdown part (which follows the peak of the amplitude of the $l = 2$, $m = 2$ mode). We use different analytical or phenomenological models to model the memory signal in each part of the binary's evolution. During the inspiral phase, we use the PN approximation. In the late inspiral phase, however, the PN approximation begins to deviate more from the NR surrogate memory; therefore, we stop using the PN waveform before any significant differences arise. During the ringdown part, we use a superposition of QNMs with different overtone numbers to model oscillatory (l, m) modes, and their products to model the memory signal. We also require that the final accumulated memory should match the value computed from our fit for the final memory strain Δh_{20} defined in Eq.(4.15). This implies that our time-domain memory model can be written naturally as a piecewise function as follows:

$$h_{20}(t) = \begin{cases} h_{20}^{\text{insp}}(t) & \text{for } t < t_{\text{int}}, \\ h_{20}^{\text{int}}(t) & \text{for } t_{\text{int}} \leq t \leq t_{\text{rd}}, \\ \Delta h_{20} - h_{20}^{\text{rd}}(t) & \text{for } t \geq t_{\text{rd}}. \end{cases} \tag{4.16}$$

The times t_{int} and t_{rd} define the start of the intermediate and ringdown memory signals, respectively. Notice that we wrote the memory accumulated during the ringdown as the difference between the final memory Δh_{20} and the memory accumulated from some time t up to infinity $h_{mem}^{\text{rd}}(t)$, so as to enforce that the final time-domain memory signal matches our fit for the final memory for t much greater than t_{rd}.

4.4.1 PN memory signal during inspiral

During the inspiral phase, we use the 3.5PN memory waveform given in Eq. (4.13) as our model for the time-domain memory signal, as we did for our hybridization procedure in Sec. 4.3.2. To make the inspiral signal faster to evaluate, we would like to avoid having to perform a hybridization procedure with the surrogate at each mass ratio to obtain the optimal t_c for that mass ratio. Instead, we construct a fitting function for t_c as a function of mass ratio (for $q \in [1, 8]$), which minimizes a residual between the hybridized surrogate and the 3.5PN waveform.

As in Paper I, we will assume that t_c can be fit using a quartic polynomial in the symmetric mass ratio, η,

$$t_c = \sum_{i=0}^{4} t_{c,i}\, \eta^i. \tag{4.17}$$

The five coefficients $t_{c,i}$ in the polynomial function of η are the parameters for which we will fit. In this paper, however, we will use a somewhat different approach to find the values of the $t_{c,i}$ parameters from what was performed in Paper I. Here, we determine these coefficients by minimizing the cost function

$$C[h_{20}^{\rm surr}, h_{20}^{\rm insp}] = \sum_{q=1}^{8} C_q[h_{20}^{\rm surr}, h_{20}^{\rm insp}], \tag{4.18}$$

where the cost function $C_q[h_{20}^{\rm surr}, h_{20}^{\rm insp}]$ was defined in Eq. (4.14). We obtain the optimized values of the parameters $t_{c,i}$ by using the `SciPy` *minimize* function again. Note that unlike the hybridization in Sec. 4.3.2, which determined the value of t_c by minimizing a cost function for a single mass ratio over a time window of duration $1000M$, here we modify the procedure in two ways. First, the hybridization is done over a longer time interval between $t_1 = -10^4 M$ and $t_2 = -2000M$. Second, the cost function is defined as the sum over eight integer mass ratios $q = 1, 2, \ldots, 8$ of the cost functions used for the hybridization for a single mass ratio in Sec. 4.3.2.

The result of this optimization are the five coefficients $t_{c,i}$, the values for which are given in Table 4.3. The inspiral memory waveform model can be computed using

Coefficient	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$
$t_{c,i}$	2.1558×10^3	-2.2124×10^4	1.1760×10^5	-3.3106×10^5	3.80724×10^5

Table 4.3: Coefficients for the t_c polynomial fit in Eq. (4.17). More details about how the fits were constructed is given in the text of Sec. 4.4.1.

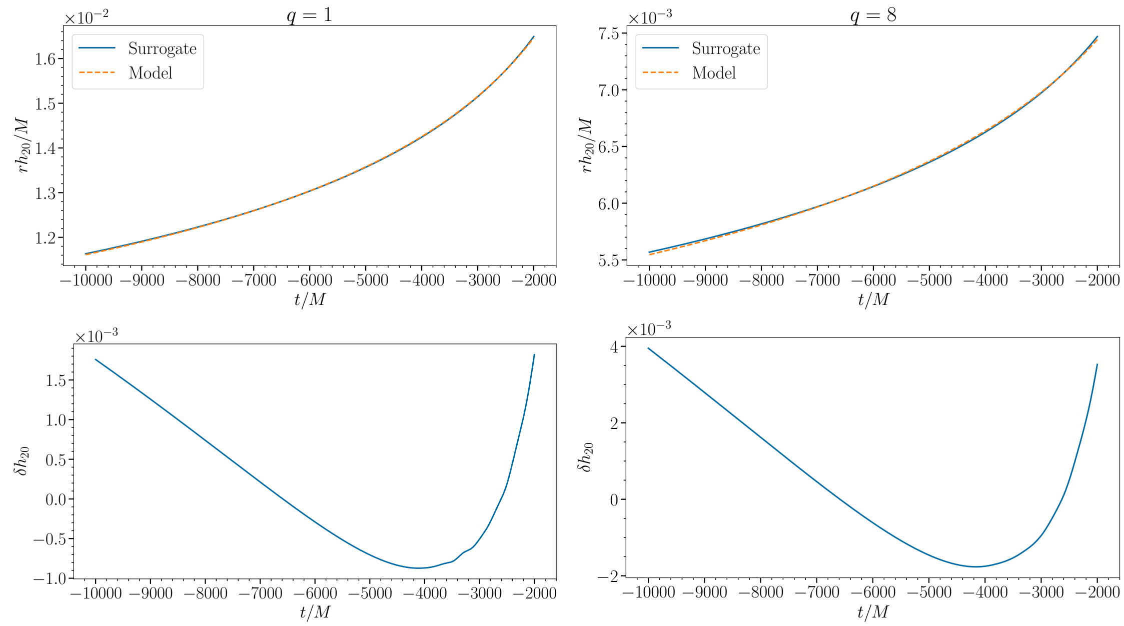

Figure 4.1: **Inspiral memory model and its relative error versus time**: *Top*: The hybridized surrogate memory signal computed from the NRHybSur3dq8 surrogate waveform modes (solid, blue curve) and the inspiral time-domain memory model (dashed, orange curve) for an equal-mass non-spinning BBH merger (left) and for a mass ratio $q = 8$ (right). *Bottom*: The relative error $\delta h_2 0$ in Eq. (4.19) the inspiral memory model for each of the corresponding mass ratios shown above.

the values of the coefficients $t_{c,i}$ in the expansion of t_c in Eq. (4.17) and substituting by t_c into the 3.5PN memory waveform in Eq. (4.13) through the expression for $x(t)$ in Eq. (4.11). This allows the inspiral memory signal to be evaluated rapidly for $q \in [1, 8]$ without additional hybridization, but to still have the appropriate offset at the initial time at which it is evaluated.

The inspiral memory signal model is shown in the top row of Fig. 4.1 for two

BBH systems with mass ratios $q = 1$ (left) and $q = 8$ (right). The solid blue curve shows the memory computed from the surrogate model, and the dashed orange curve shows the memory computed from the inspiral model in Eq. (4.13). The accuracy of the model is measured by computing the relative error between the inspiral memory model and the inspiral part of the hybridized surrogate memory

$$\delta h_{20} = \frac{|h_{20}^{\rm surr} - h_{20}^{\rm model}|}{h_{20}^{\rm surr}} \,. \tag{4.19}$$

The relative error at the two smallest and largest mass ratios in our mode is representative of the results for all mass ratios in the model ($q \in [1, 8]$). Namely, the relative error is the largest positive value near the times t_1 and t_2, and is most negative near $-4000M$, which was the t_2 value over which the surrogate was hybridized with the 3.5PN waveform. For larger q values, the relative error increases, because the PN approximation becomes less accurate at a fixed time from merger.

4.4.2 Ringdown memory signal modeling

During the ringdown phase, recent work has showed that a superposition of QNMs of different overtones provides a good fit to the waveform (l, m) modes after the peak amplitude of the mode (see, e.g., [144–146]). We will use this approach to model the ringdown phase of the oscillatory modes that are the inputs to our calculation of the nonlinear memory effect. We will ignore the nonlinearities in the oscillatory ringdown modes [86], because they will be a further nonlinear correction to the nonlinear memory effect. To determine the QNM frequencies, we use the `qnm` package [147].[2]

The ringdown QNM waveforms are often expressed in terms of spin-weighted spheroidal harmonics ${}_sS_{lm}(\theta, \phi; a\omega)$. The spheroidal harmonics can be expanded in terms of the spin-weighted spherical harmonics ${}_sY_{lm}(\theta, \phi)$, which are the basis used for the NR surrogate waveform modes, as follows:

$$
{}_sS_{lm}(\theta, \phi; a\omega) = \sum_{l'=l_{\min}}^{\infty} A_{l'lm}(a\omega)\, {}_sY_{l'm}(\theta, \phi)\,. \tag{4.20}
$$

Here $l_{\min}$ is given by $l_{\min} = \max\left(|m|, |s|\right)$ and the coefficients $A_{l'lm}(a\omega)$ are complex-valued coefficients that are referred to as "mixing" coefficients. The QNMs are functions of the final mass M_f and final spin χ_f of the remnant BH, which are nontrivial functions of the initial masses m_1 and m_2 that are used during the inspiral phase. We use the `SurfinBH` Python package to compute the final mass and spin parameters from the individual mass parameters.

The QNM frequencies are parametrized by three integers: the spheroidal harmonic indices (l, m) and the overtone number $n = 0, 1, \ldots$. We will typically denote them by ω_{lmn}. They generically have nonzero real and imaginary parts. The real part $\Re[\omega_{lmn}]$ is the angular frequencies and minus the imaginary part $-\Im[\omega_{lmn}]$ is the inverse of the damping time (we use the convention that the imaginary part is negative). Higher overtones have a faster damping rate, so the fundamental mode $(n = 0)$ is longer lived

[2] This package uses a spectral method for solving the angular Teukolsky equation by Cook and Zalutskiy [17] (which has some advantages over Leaver's original method [18]). In addition to the QNM frequencies it also gives the angular separation constant and the spherical-spheroidal mixing functions (discussed below) for any dimensionless spin parameter χ that satisfies $0 \leq \chi < 1$.

than the $n > 0$ overtones. For each (l, m, n) triple, there is a QNM with a positive and a negative real part, which are referred to as the "ordinary" ($\Re[\omega_{lmn}] > 0$) and the "mirror" ($\Re[\omega_{lmn}] < 0$) modes. For rotating black holes, for each (l, m, n) value with $m \neq 0$, there are two modes, the "prograde" and the "retrograde" modes, for which m has the same or opposite sign, respectively, as the sign of the energy. There is a discrete symmetry of the modes that relates ordinary, prograde modes with the mirror, retrograde modes (and a similar symmetry for the ordinary, retrograde and mirror, prograde modes). For nonspinning black-hole binaries, the case we are focusing on in this paper, the waveform (l, m) modes can be well fit by just the prograde modes (see, e.g., [144]), so we also use just prograde modes. The `qnm` package computes the ordinary modes and thus prograde modes for positive m and retrograde modes for negative m; the other cases can be obtained through the discrete symmetry described above.

4.4.2.1 Multimode ringdown fitting

Equation 4.20 implies that a single spheroidal-harmonic QNM mode can contribute to multiple spherical-harmonic strain modes. As is described in [145], this implies that to fit for the QNM amplitudes from the spherical-harmonic modes of the NR surrogate waveform, then one would generically need to fit for multiple amplitudes simultaneously in the different spherical-harmonic modes rather than fitting sequentially in each mode. For a black hole of a known final mass and spin, the frequencies are known, and the fitting problem is linear in the complex amplitudes of the modes.

We will use the "eigenvalue method" introduced in [145] to fit the strain modes. This method involves constructing overlap integrals between the NR-surrogate spherical-harmonic strain modes and a model of the ringdown formed from a superposition of QNM overtones written in terms of spheroidal harmonics. The least-squares solution can be determined from maximizing an overlap between the normalized surrogate model and the waveform constructed from the superposition on QNMs with different overtones. We describe the procedure more quantitatively below.

First, we write the ringdown GW signal as a linear combination of QNMs by

$$h^{\rm rd}_{\rm model} = \frac{M}{r} \sum_{l,m,n} C_{lmn} e^{-i\omega_{lmn}(t-t_0)} {}_{-2}S_{lm}(\theta, \phi; a\omega). \tag{4.21}$$

Here the coefficients C_{lmn} are the complex amplitudes of the different QNMs and t_0 is the start time of the QNM mode, which is chosen to be some time after the peak amplitude of the $l = 2$, $m = 2$ waveform mode.[3] We will set the maximum overtone index n in the sum to be 7, based on the findings of prior studies [144,145]. Following the notation in [145], we will rewrite the superposition of QNMs in $h^{\rm rd}_{\rm model}$ as

$$h^{\rm rd}_{\rm model} = \frac{M}{r} \sum_{K\in\{l,m,n\}} C_K h_K \, . \tag{4.22}$$

Here the index K is a short-hand for the triple index lmn of a QNM. The correspond-

[3]Note that we will use as our time, t, the time used in the NR surrogate model, which is measured in terms of the initial total mass of the system, M. The QNM frequencies are more naturally expressed in terms of the mass of the final black hole, M_f, which differs from M. Thus, one should rescale the frequencies ω_{lmn} by M/M_f so that they are expressed in units consistent with the time used in the NR surrogate. To avoid introducing factors of M/M_f, we absorb this factor in the values of the frequencies ω_{lmn} (i.e., we assume that t and ω_{lmn} are scaled to be in the same units).

ing waveform for a QNM is

$$h_K \equiv e^{-i\omega_{lmn}(t-t_0)}\,{}_{-2}S_{lm}(\theta,\phi;a\omega)\,. \tag{4.23}$$

The eigenvalue method in [145] is the solution that arises from maximizing the normalized overlap between the ringdown model and the surrogate waveform. The overlap is defined by

$$\rho^2 = \frac{|\langle h^{\rm rd}_{\rm model}|h_{\rm surr}\rangle|^2}{\langle h_{\rm surr}|h_{\rm surr}\rangle\langle h^{\rm rd}_{\rm model}|h^{\rm rd}_{\rm model}\rangle}\,, \tag{4.24}$$

where the inner product of two complex functions is given by

$$\langle h_1|h_2\rangle \equiv \int_{t_i}^{t_f} dt \int d^2\Omega\,\bar{h}_1(t,\Omega)h_2(t,\Omega)\,. \tag{4.25}$$

Substituting $h^{\rm rd}_{\rm model}$ in Eq. (4.22) into the overlap function in Eq. (4.24) gives for ρ^2

$$\rho^2 = \frac{\sum_K |\bar{C}_K A_K|^2}{\langle h_{\rm surr}|h_{\rm surr}\rangle \sum_{I,J}\bar{C}_I B_{IJ} C_J}\,. \tag{4.26}$$

Here, we introduced the notation A_k and B_{ij} to represent the following inner products:

$$A_K \equiv \langle h_K|h_{\rm surr}\rangle\,, \tag{4.27a}$$

$$B_{IJ} \equiv \langle h_I|h_J\rangle\,. \tag{4.27b}$$

The indices I and J, like K, denote a triple of lmn indices for a QNM. We will introduce a matrix-vector notation, $\mathbf{A}_\eta$ and $\mathbb{B}_\eta$, to denote these collections of inner products for a specific value of η. By maximizing the overlap function ρ with respect to the unknown coefficients $\mathbf{C}_\eta$, the solution is given by [145]

$$\boldsymbol{C}_\eta = \mathbb{B}_\eta^{-1} \cdot \boldsymbol{A}_\eta \,. \tag{4.28}$$

The η subscript is used here to represent solving this fitting problem for a specific BBH system with symmetric mass ratio η.

We next give more explicit expressions for the components of $\boldsymbol{A}_\eta$ by substituting h_K in Eq. (4.23) and the spin-weighted spherical harmonic expansion of $h_{\rm surr}$ into the definition of A_K in Eq. (4.27a).

$$A_K = \sum_{l'm' \in \{\rm surr\}} \int dt \, e^{i\bar{\omega}_{lmn}(t-t_0)} h^{\rm surr}_{l'm'} \int d^2\Omega \, {}_{-2}\bar{S}_{lm}(\theta,\phi; a\omega_{lmn}) \, {}_{-2}Y_{l'm'}(\theta,\phi) \,. \tag{4.29}$$

The notation $l', m' \in \{\rm surr\}$ in the sum means that the sum takes place over (a subset of) the modes that are included in the surrogate waveform. Using the expansion of the spin-weighted spheroidal harmonics in terms of the spin-weighted spherical harmonics in Eq. (4.20), and the orthogonality of the spin-weighted spherical harmonics in Eq. (4.2) gives

$$A_K = \sum_{l'} \bar{A}_{l'lm}(a\omega_{lmn}) \int dt \, e^{i\bar{\omega}_{lmn}(t-t_0)} h^{\rm surr}_{l'm}(t) \,. \tag{4.30}$$

The quantities $A_{l'lm}$ are the mixing coefficients that arise from expanding the spin-weighted spheroidal harmonics in Eq. (4.20) in terms of spherical harmonics. Similarly, the components of $\mathbb{B}_\eta$ are given by

$$B_{IJ} = \int dt \int d^2\Omega \, e^{i(\bar{\omega}_{lmn} - \omega_{l'm'n'})(t-t_0)} \, {}_{-2}\bar{S}_{lm}(\theta,\phi; a\omega_{lmn}) \, {}_{-2}S_{l'm'}(\theta,\phi; a\omega_{l'm'n'}) \,. \tag{4.31}$$

Using the expansion of the spin-weighted spheroidal harmonics in Eq. (4.20) and using the orthogonality of the spin-weighted spherical harmonics, the result can be written as

$$B_{IJ} = \int dt\, e^{i(\bar{\omega}_{lmn} - \omega_{l'mn'})(t-t_0)} \sum_{l''} \bar{A}_{l''lm}(a\omega_{lmn}) A_{l''l'm}(a\omega_{l'mn'})\,. \tag{4.32}$$

The sum over l'' runs over all allowed values of l'' above the larger of $l_{\min}$ and $l'_{\min}$. In practice, we must truncate the sum at a finite upper value of l'', for which the coefficients $A_{l''lm}$ are sufficiently small.

We compute the components of $\boldsymbol{A}_\eta$ and $\mathbb{B}_\eta$ by evaluating the time integrals in Eq. (4.30) numerically and those in (4.32) analytically. We integrate over the time interval $[t_i, t_f] = [0, 130M]$. We choose the start of the QNM modes t_0 to be at the merger time, which we define to be at the peak time of the $l = 2$, $m = 2$ mode of the waveform (i.e., $t_0 = 0$). After constructing $\boldsymbol{A}_\eta$ and $\mathbb{B}_\eta$, we can solve Eq. (4.28) for the QNM amplitudes $\mathbf{C}_\eta$.

4.4.2.2 Multiple mass-ratio, multimode fitting

As with the inspiral model, we will calculate a fit for the ringdown amplitudes at a particular mass ratio that can be computed from a fitting function in η, so that it is not necessary to solve the multimode fitting problem in Eq. (4.28) at each mass ratio. To make this fitting problem more tractable, we restrict the number of waveform modes that we include in the spherical-harmonic expansion of the surrogate model. Specifically, we fit the (l, m) modes $(2, \pm 1)$, $(2, \pm 2)$, and $(3, \pm 2)$ given in the first row of Table 4.1. We henceforth refer to just the positive m values, because the negative m values can be obtained through complex conjugation.

For our fitting ansatz, we expand the QNM amplitude coefficients as quadratic polynomials in η (the form of these quadratic fits is given in Eq. (4.33)). To account for the vanishing of the $l = 2$, $m = 1$ mode for equal-mass binaries (with mass ratio $q = 1$ and symmetric mass ratio $\eta = 1/4$), we multiply the coefficients of the $(2, 1)$ mode by a factor of $\sqrt{1 - 4\eta}$:

$$C_{22n} = \sum_{j=0}^{2} C_{22nj}\, \eta^j \, , \tag{4.33a}$$

$$C_{32n} = \sum_{j=0}^{2} C_{32nj}\, \eta^j \, , \tag{4.33b}$$

$$C_{21n} = \sqrt{1 - 4\eta} \sum_{j=0}^{2} C_{21nj}\, \eta^j \, . \tag{4.33c}$$

Because we fit for 24 mode amplitudes—three (l, m) mode pairs with the fundamental mode and seven overtones—our fitting problem contains 72 undetermined parameters. We determine these coefficients by fitting for these coefficients using three mass ratios. For the $(2, 2)$ and $(3, 2)$ modes we include the smallest and biggest mass ratios available plus one in the middle of the range (specifically $q = 1, 5, 8$). However, because the h_{21} mode vanishes for equal mass ratio binaries, we fit over the three mass ratios $q = 2, 5, 8$. Because there is no mode mixing between the $m = 1$ and $m = 2$ modes, these two fitting problems decouple. Thus, we solve two fully determined problems for the 24 coefficients C_{21nj} and for the 48 coefficients C_{22nj} and C_{32nj}. We introduce N to denote the number of QNMs in the multimode fitting (i.e., 8 for the $l = 2$, $m = 1$ case and 16 for the joint $l = 2, 3$, $m = 2$ case).

The eigenvalue method discussed in Sec. 4.4.2.1 requires some modifications to obtain the coefficients C_{Kj} rather than $\mathbf{C}_\eta$. The problem is still linear, and the solution can be written as the solution to a linear system of equations,

$$\boldsymbol{C} = \mathbb{B}^{-1} \cdot \boldsymbol{A} \,. \tag{4.34}$$

Here $\boldsymbol{A}$ is constructed by combining three vectors $\boldsymbol{A}_\eta$ for the three different mass ratios used in the fitting procedure of a specific (l, m) mode.

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{\eta_1} \\ \boldsymbol{A}_{\eta_2} \\ \boldsymbol{A}_{\eta_3} \end{pmatrix}_{3N\times 1} . \tag{4.35}$$

The components of $\boldsymbol{A}_{\eta_i}$ are computed from Eq. (4.30) for the i^{th} mass ratio. The matrix $\mathbb{B}$ can be written in terms of the product of two matrices. The first one is a block diagonal matrix constructed from the matrices $\mathbb{B}_\eta$ for each of the three mass ratios. The second matrix is similar to a Vandermonde matrix, but it differs in that it includes the terms η^i that appear in the expansion of the coefficients in terms of the symmetric mass ratio in Eq. (4.33). For fitting the $l = 2$, $m = 2$ and $l = 3$, $m = 2$ modes, the $\mathbb{B}$ matrix is defined as

$$\mathbb{B} = \begin{pmatrix} \mathbb{B}_{\eta_1} & 0 & 0 \\ 0 & \mathbb{B}_{\eta_2} & 0 \\ 0 & 0 & \mathbb{B}_{\eta_3} \end{pmatrix}_{3N\times 3N} \cdot \begin{pmatrix} \mathbb{I} & \mathbb{S}_{\eta_1} & \mathbb{S}^2_{\eta_1} \\ \mathbb{I} & \mathbb{S}_{\eta_2} & \mathbb{S}^2_{\eta_2} \\ \mathbb{I} & \mathbb{S}_{\eta_3} & \mathbb{S}^2_{\eta_3} \end{pmatrix}_{3N\times 3N} , \tag{4.36}$$

where η_i represents the i^{th} mass ratio over which the fitting is done. The matrix $\mathbb{S}_j$ is a diagonal matrix with the symmetric mass ratio as its diagonal elements, defined for the i^{th} mass ratio as

$$\mathbb{S}_{\eta_i} = \text{diag}(\eta_i, \ldots, \eta_i)_{N\times N} \,. \tag{4.37}$$

For the $l = 2$, $m = 1$ mode, the factor $\sqrt{1-4\eta_i}$ can be absorbed into the matrix $\mathbb{B}$ by adding another block-diagonal matrix

$$\mathbb{B} = \begin{pmatrix} \mathbb{B}_{\eta_1} & 0 & 0 \\ 0 & \mathbb{B}_{\eta_2} & 0 \\ 0 & 0 & \mathbb{B}_{\eta_3} \end{pmatrix} \cdot \begin{pmatrix} \mathbb{F}_{\eta_1} & 0 & 0 \\ 0 & \mathbb{F}_{\eta_2} & 0 \\ 0 & 0 & \mathbb{F}_{\eta_3} \end{pmatrix} \cdot \begin{pmatrix} \mathbb{I} & \mathbb{S}_{\eta_1} & \mathbb{S}^2_{\eta_1} \\ \mathbb{I} & \mathbb{S}_{\eta_2} & \mathbb{S}^2_{\eta_2} \\ \mathbb{I} & \mathbb{S}_{\eta_3} & \mathbb{S}^2_{\eta_3} \end{pmatrix} , \tag{4.38}$$

where the matrix of additional factors $\mathbb{F}_{\eta_i}$ is defined to be

$$\mathbb{F}_{\eta_i} = \text{diag}(\sqrt{1-4\eta_i}, \ldots, \sqrt{1-4\eta_i})_{N\times N} \,. \tag{4.39}$$

The vector of QNM amplitude coefficients $\boldsymbol{C}$ has the form

$$\boldsymbol{C} = \begin{pmatrix} C_{K0} \\ C_{K1} \\ C_{K2} \end{pmatrix}_{3N\times 1} , \tag{4.40}$$

where K runs over all QNM indices lmn included in the multimode fit. The coefficients C_{Kj} were defined in Eq. (4.33).

We solve the system of linear equations in Eq. (4.34) by computing the inverse of the constructed matrix $\mathbb{B}$ using its singular value decomposition. The components in the vector of coefficients $\boldsymbol{C}$ in Eq. (4.34) are written in terms of their amplitude and phase in Tables F.1 and F.2 in Appendix F.

The model for the GW strain modes can be computed from Eq. (4.21) with the coefficients C_{Kj}, the spherical-spheroidal mixing coefficients $A_{l'lm}(a\omega)$, and the QNM frequencies ω_{lmn}. Substituting the expansion of the spin-weighted spheroidal harmonics in Eq. (4.20) into the strain model $h^{\rm rd}_{\rm model}$ in Eq. (4.21) gives

$$h^{\rm rd}_{\rm model} = \frac{M}{r}\sum_{l,m,n,j} C_{lmnj}\eta^j e^{-i\omega_{lmn}(t-t_0)}\sum_{l'} A_{l'lm}(a\omega_{l'mn})_{-2}Y_{l'm}\,. \tag{4.41}$$

It then follows that the (l,m) mode ringdown strain expanded in spherical harmonics can be written as

$$h^{\rm rd}_{lm} = \frac{M}{r}\sum_{l',n,j} C_{l'mnj}\eta^j A_{ll'm}(a\omega_{lmn})e^{-i\omega_{l'mn}(t-t_0)}\,. \tag{4.42}$$

It can be seen from the form of the ringdown strain modes in Eq. (4.42) that the mixing of different modes arises for different l with the same m, because of the coefficients $A_{ll'm}$.

We first show the results of the fitting procedure for the $l = 2$, $m = 1$ mode in Fig. 4.2. We illustrate just the mass ratio $q = 4$, as a representative mass ratio which we did not use in constructing the ringdown fit. The top panel shows the surrogate model as a solid, blue curve. The ringdown model of the mode is the orange, dashed curve. The residual (difference between the QNM ringdown model and the NR surrogate) is shown in the bottom panel. For the $l = 2$, $m = 1$ mode, the difference between the ringdown model and the surrogate is at most a few percent.

Next, we show the similar results for the $l = 2, 3$ and $m = 2$ modes in Fig. 4.3 for

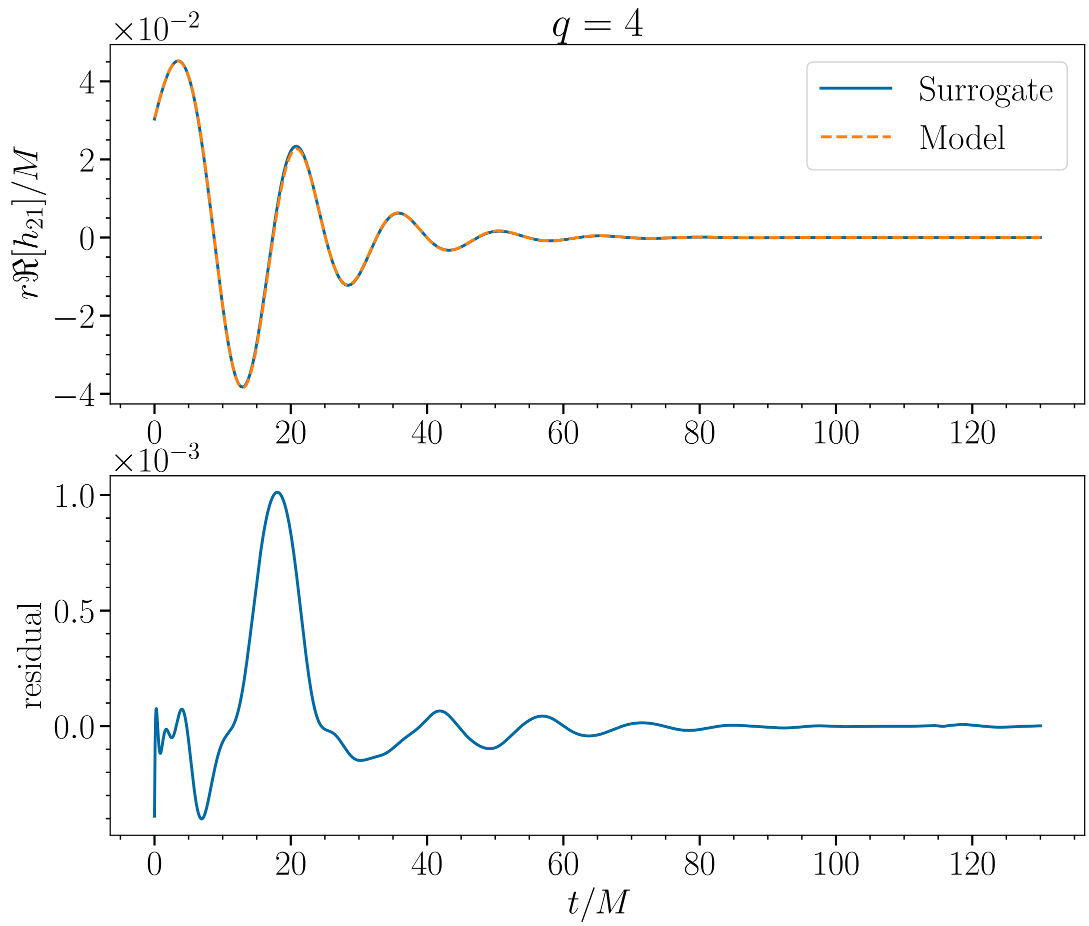

Figure 4.2: **Waveform for the h_{21} mode**: *Top*: The real part of the h_{21} mode computed from the surrogate model shown in solid, blue curve, and the same mode computed from our ringdown model in Eq. (4.42) is shown as an orange, dashed line. *Bottom*: Tjhe residuals between the h_{21} waveform evaluated from the surrogate model and our ringdown model waveform.

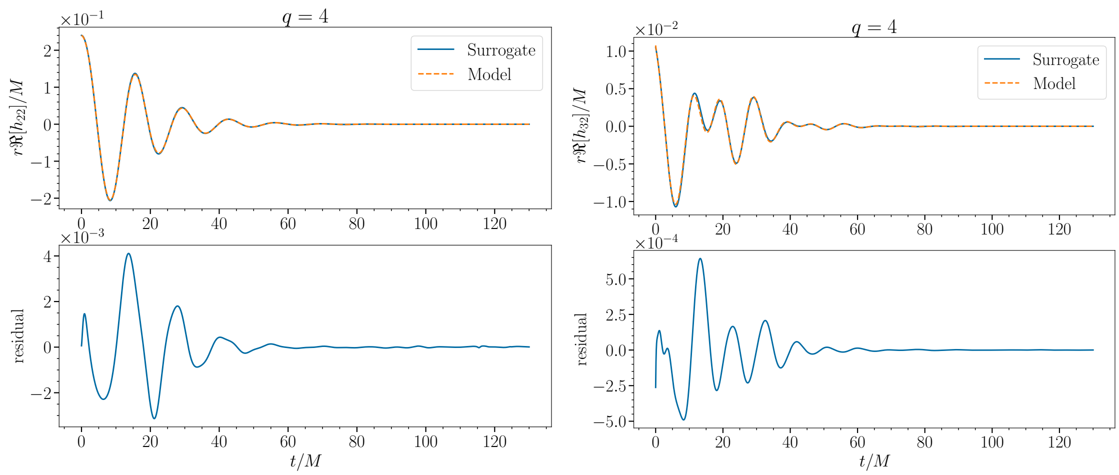

Figure 4.3: **Waveform for the h_{22} and h_{32} modes**: *Top*: The real part of the h_{22} mode (left) from the surrogate model shown in solid, blue curve. The same mode computed from our ringdown model is the orange, dashed curve. On the right are the equivalent results for the h_{32} mode. *Bottom*: The residuals between the h_{22} waveform evaluated from the surrogate model and our ringdown model waveform (left) and the same quantity for the h_{32} mode (right).

the mass ratio $q = 4$. The results for $l = 2$ are shown on the left and $l = 3$ are on the right. The depiction of the curves and the residuals are completely analogous to those in Fig. 4.2 for the $l = 2$, $m = 1$ mode. The $l = 2$, $m = 2$ mode is the largest of the three (more than an order of magnitude larger than the $l = 3$, $m = 2$ mode and a factor of five larger than the $l = 2$, $m = 1$ mode). The multimode ringdown model again reproduces the NR surrogate results with an accuracy of roughly a few percent. The $l = 3$, $m = 2$ mode of the ringdown has a more complicated oscillatory structure than the $l = 2$, $m = 2$ mode, because the $l = 2$, $m = 2$ QNMs mix into the $l = 3$ and $m = 2$ harmonic with a larger relative amplitude than the mixing of the $l = 3$, $m = 2$ modes into the $l = 2$, $m = 2$ harmonic.

Note that we did not fit for the QNM frequencies ω_{lmn} or the mixing coefficients

$A_{l'lm}$ for different mass ratios. To evaluate the model for the ringdown memory modes in Eq. (4.42), we use the *Python* package `qnm` for a certain final mass M_f and spin parameter χ_f for a given mass ratio. We obtain the values of the final mass and spin from the `SurfinBH` package rather than refitting those values.

4.4.2.3 The ringdown memory model

We now compute the model for the ringdown memory strain signal using the modes $h^{\rm rd}_{lm}$ in Eq. (4.42). To obtain the h_{20} memory mode, we substitute the ringdown strain modes computed in Eq. (4.42) into Eq. (4.9). As we described near Eq. (4.16), during the ringdown, we compute the memory signal for some $t \geq t_0$ by integrating from late times ($t \to \infty$) back to a time t and adding the result from the final memory offset. This is equivalent to subtracting the integral from this time $t \geq t_0$ to infinity from Δh_{20}. We can explicitly evaluate this integral by using the form of the ringdown modes in Eq. (4.42). We find that

$$\int_t^\infty dt'\, \dot{h}_{l'm'}(t')\dot{\bar{h}}_{l''m'}(t') = \frac{M^2}{r^2} \sum_{\bar{l},\bar{\bar{l}},\bar{n},\bar{\bar{n}},j,j'} \left(\frac{i\omega_{\bar{l}m'\bar{n}}\bar{\omega}_{\bar{\bar{l}}m'\bar{\bar{n}}}}{\omega_{\bar{l}m'\bar{n}} - \bar{\omega}_{\bar{\bar{l}}m'\bar{\bar{n}}}} \right) C_{\bar{l}m'\bar{n}j'} \bar{C}_{\bar{\bar{l}}m'\bar{\bar{n}}j} \eta^{j+j'}$$
$$\times A_{l'\bar{l}m'} \bar{A}_{l''\bar{\bar{l}}m'} e^{-i(\omega_{\bar{l}m'\bar{n}} - \bar{\omega}_{\bar{\bar{l}}m'\bar{\bar{n}}})(t-t_0)} \,. \tag{4.43}$$

Because the imaginary part of the QNM frequency is negative (i.e. $\Im[\omega_{lmn}] < 0$), the exponential term vanishes at infinity. We have also used the shorter notation of $A_{\bar{l}lm}(a\omega_{lmn}) \equiv A_{\bar{l}lm}$ for conciseness. By substituting Eq. (4.43) into the expression for the memory signal Eq. (4.9), we get the following form of the ringdown memory model for nonprecessing BBH mergers

$$h_{20}^{\rm rd}(t) = \frac{M^2}{\sqrt{6}r}\sum_{l',l''}\sum_{m'=1}^{l'}(-1)^{m'}C_l(-2,l',m';2,l'',-m')$$
$$\times \sum_{\bar{l},\bar{\bar{l}},\bar{n},\bar{\bar{n}},j,j'} \Im\left[\left(\frac{\omega_{\bar{l}m'\bar{n}}\bar{\omega}_{\bar{\bar{l}}m'\bar{\bar{n}}}}{\omega_{\bar{l}m'\bar{n}}-\bar{\omega}_{\bar{\bar{l}}m'\bar{\bar{n}}}}\right)C_{\bar{l}m'\bar{n}j}\bar{C}_{\bar{\bar{l}}m'\bar{\bar{n}}j'}\eta^{j+j'}A_{l'\bar{l}m'}\bar{A}_{l''\bar{\bar{l}}m'}e^{-i(\omega_{\bar{l}m'\bar{n}}-\bar{\omega}_{\bar{\bar{l}}m'\bar{\bar{n}}})(t-t_0)}\right]. \tag{4.44}$$

Equation (4.44) can be used to compute the ringdown memory signal, because we obtain the QNM frequencies ω_{lmn} and spherical-spheroidal mixing coefficients $A_{l'lm}(a\omega_{lmn})$ from the `qnm` package (using the final mass and spin from the `SurfinBH` package), and we have fit for the polynomial coefficients of the QNM amplitudes C_{lmnj} (see Tables F.1 and F.2). Given that our modeling of the oscillatory ringdown modes contains the three largest (l,m) modes $[(2,2),(3,2),(2,1)]$, we compute the memory model for the $(2,0)$ mode from Eq. (4.44) from these three modes.

We show the full ringdown memory signal in Fig. 4.4. We show the extreme two mass ratios $q=1,8$ for our model, along with the relative error between our ringdown model and the ringdown memory computed from the NR surrogate waveforms. The left panels are for the mass ratio $q=1$ and the right panels are for $q=8$. Note that the relative error is at least an order of magnitude smaller at mass ratios $q=1$ and $q=8$ than the corresponding relative error associated with the ringdown modes for $q=4$ (described Sec. 4.4.2.2). The error is significantly smaller for these mass ratios because these were two of the mass ratios used to perform the multimode fitting for the $l=2,3$ and $m=2$ oscillatory ringdown modes. The relative error for the memory

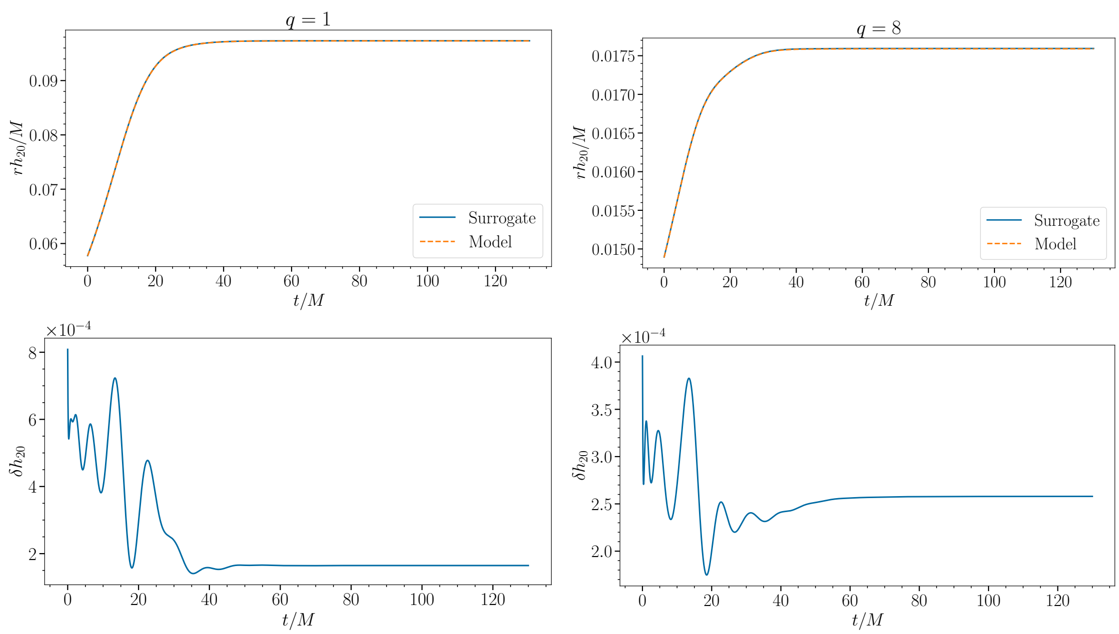

Figure 4.4: **Ringdown memory model and relative error versus time**: *Top*: The hybridized surrogate memory signal computed from the NRHybSur3dq8 surrogate waveform modes (solid, blue curve) and the ringdown memory model (dashed, orange curve) for non-spinning BBH mergers. The left panel contains the results for the mass ratio $q = 1$ and the right shows the mass ratio $q = 8$. *Bottom*: The relative error of the ringdown memory model and the memory signal computed directly from the NR surrogate for the corresponding mass ratios in the panels above.

ringdown signal for $q = 4$ is comparable to the relative error in the modes depicted in Figs. 4.2 and 4.3.

4.4.3 Intermediate-time memory model

Analytical models for the GW strain of BBH mergers during the late-inspiral and merger phases often have some phenomenological component, because PN theory and black-hole perturbation theory alone do not have sufficient accuracy to be a robust model of these phases. For this reason, we will take a purely phenomenological modeling approach to model the memory during this interval.

We model the intermediate memory as a sum of exponential functions (and a constant)

$$h_{20}^{\mathrm{int}}(t) = \frac{M}{r}\sum_{j=0}^{6} c_j e^{p_j t}. \tag{4.45}$$

We denote the amplitudes of the exponential functions by c_j and the exponents by p_j, with the convention that $p_0 = 0$ (so that the first term is a constant). We require that the intermediate memory is continuous with inspiral and ringdown memory at times $t_{\mathrm{int}} = -2000M$ and $t_{\mathrm{rd}} = 2M$, respectively. Similarly, we enforce continuity of the first and second time derivatives of the memory signal at these times. The choice of starting the ringdown model at $t_{\mathrm{rd}} = 2M$ was made empirically by observing that it produced a better matching of the time-derivatives of the memory signal than the start time of the ringdown memory signal ($t = t_0 = 0$).

To construct a memory signal that can be evaluated for any mass ratio in the range $q \in [1, 8]$, we expand the parameters p_j (for $j \neq 0$) as linear functions of the symmetric mass ratio η:

$$p_j = p_{j0} + p_{j1}\eta\,. \tag{4.46}$$

The intermediate model nominally has nineteen free parameters: seven amplitude coefficients c_j and twelve exponent coefficients p_{ji}. By matching the value of the memory model and its first and second derivatives at the times $t_{\rm int}$ and $t_{\rm rd}$, we fix six of the free parameters, leaving thirteen in our model. We choose to solve linear equations for the coefficients c_j for $j = 0, \ldots, 5$ in terms of the other parameters, which leaves c_6 and p_{10}, $p_{11}, \ldots, p_{61}$ as the thirteen free parameters.

The system of linear equations for the six coefficients $c_0, \ldots, c_5$ that we solve can be written in matrix form as

$$\mathbb{M} \cdot \mathbf{c} = \mathbf{h}\,. \tag{4.47}$$

Here $\mathbb{M}$ is the model in Eq. (4.45) evaluated at the initial and final times, $\mathbf{c}$ is a vector of the coefficients $c_0, \ldots, c_5$, and $\mathbf{h}$ is the inspiral and ringdown memory models evaluated at the initial and final times, respectively. The components of $\mathbb{M}$ are constructed by evaluating the exponential terms in the memory model (and its first and second derivatives) in Eq. (4.45) at times $t_{\rm int}$ and $t_{\rm rd}$

$$\mathbb{M} = \begin{pmatrix} 1 & e^{p_1 t_{\rm int}} & \dots & e^{p_5 t_{\rm int}} \\ 1 & e^{p_1 t_{\rm rd}} & \dots & e^{p_5 t_{\rm rd}} \\ 0 & p_1 e^{p_1 t_{\rm int}} & \dots & p_5 e^{p_5 t_{\rm int}} \\ 0 & p_1 e^{p_1 t_{\rm rd}} & \dots & p_5 e^{p_5 t_{\rm rd}} \\ 0 & p_1^2 e^{p_1 t_{\rm int}} & \dots & p_5^2 e^{p_5 t_{\rm int}} \\ 0 & p_1^2 e^{p_1 t_{\rm rd}} & \dots & p_5^2 e^{p_5 t_{\rm rd}} \end{pmatrix} . \tag{4.48}$$

The vector **c** is

$$\boldsymbol{c} = \begin{pmatrix} c_0 \\ \vdots \\ c_5 \end{pmatrix} . \tag{4.49}$$

Finally, the vector **h** is constructed from the values of the inspiral and ringdown models (and their first and second derivatives) evaluated at the times $t_{\rm int}$ and $t_{\rm rd}$, respectively

$$\mathbf{h} = \begin{pmatrix} h_{20}^{\rm insp}(t_{\rm int}) \\ h_{20}^{\rm rd}(t_{\rm rd}) \\ \dot{h}_{20}^{\rm insp}(t_{\rm int}) \\ \dot{h}_{20}^{\rm rd}(t_{\rm rd}) \\ \ddot{h}_{20}^{\rm insp}(t_{\rm int}) \\ \ddot{h}_{20}^{\rm rd}(t_{\rm rd}) \end{pmatrix} . \tag{4.50}$$

Solving the system of linear equations in Eq. (4.47) gives the values of the coefficients $(c_0, \dots, c_5)$ that matches the intermediate memory model (and its first and

Coefficient	Numerical Value
p_{10}	8.49306×10^{-1}
p_{11}	1.02024×10^{1}
p_{20}	4.29098×10^{-2}
p_{21}	2.36601×10^{-1}
p_{30}	8.37061×10^{-3}
p_{31}	1.59611×10^{-2}
p_{40}	5.97250×10^{-4}
p_{41}	1.10423×10^{-3}
p_{50}	8.23572×10^{-1}
p_{51}	1.01875×10^{0}
p_{60}	9.40991×10^{-4}
p_{61}	6.22363×10^{-3}
c_6	1.19119×10^{-3}

Table 4.4: The coefficients of the intermediate memory model in Eq. (4.45). The remaining coefficients $c_0, \ldots, c_5$ are obtained from solving the linear system of equations in (4.47).

second time-derivatives) to the inspiral and ringdown models, for a given choice of the remaining free parameters (c_6 and p_{10}, $p_{11}, \ldots, p_{61}$).

We determine values of the remaining thirteen parameters by performing a non-linear least-squares fit by minimizing the cost function $C[h_{20}^{\rm surr}, h_{20}^{\rm int}]$:

$$C[h_{20}^{\rm surr}, h_{20}^{\rm int}] = \sum_{q=1}^{8} C_q[h_{20}^{\rm surr}, h_{20}^{\rm int}]\,. \tag{4.51}$$

The cost function is defined analogously to that in Eq. (4.18): namely, it is the sum over eight mass ratios $q = [1, \ldots, 8]$, but now it is evaluated using the intermediate model over the time interval $[t_{\rm int}, t_{\rm rd}] = [-2000M, 2M]$. The cost function is minimized using the `SciPy` *minimize* function. The resulting coefficients that minimize the cost function are listed in Table 4.4.

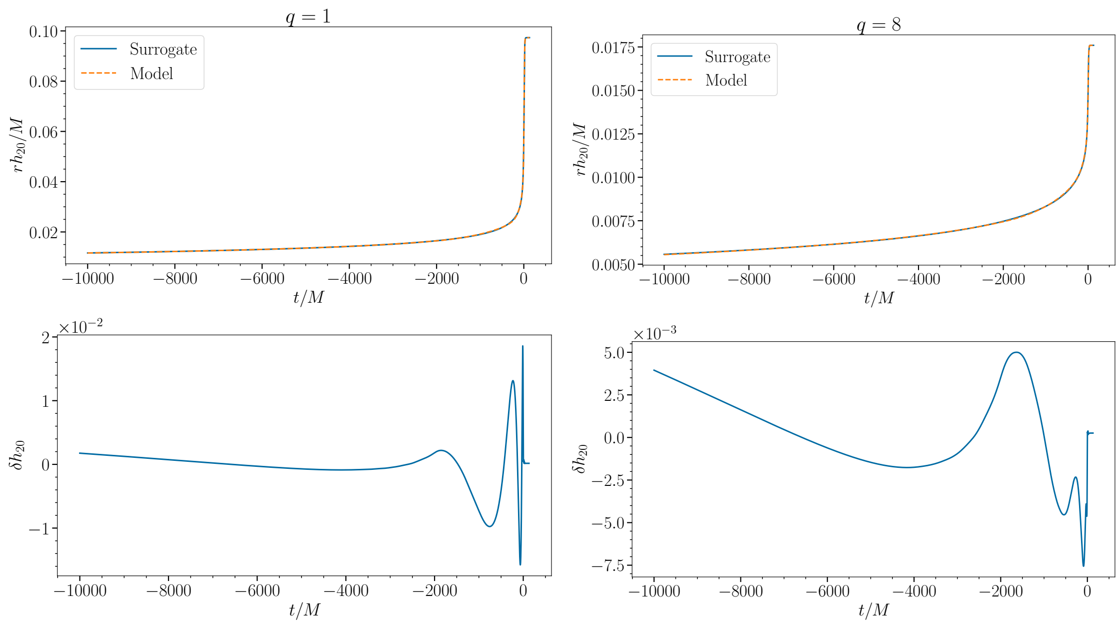

Figure 4.5: **Memory model and relative error versus time**: *Top*: The GW memory signal computed directly from the hybridized surrogate (solid blue) and from the time-domain model (dashed orange). The left panel is for an equal mass binary ($q = 1$), whereas the right panel is for the largest mass ratio of $q = 8$ that we model. *Bottom*: The relative error between the surrogate and the time-domain model for the two mass ratios.

4.4.4 Full time-domain memory signal model

Now that we have described how we construct all three parts of the model, we can now show the complete time-domain model. We do so in Fig. 4.5, which displays the scaled memory signal $(r/M)h_{20}$ as a function of time. In both the left and right panels on top, the solid, blue curve is the memory signal computed directly from the NR hybrid surrogate model. The dashed, orange curve in these panels is the memory signal computed using our model. The left panel is the memory signal from an equal mass ($q = 1$) BBH system, whereas the right panel is that for a mass ratio of $q = 8$. The relative error between the two signals in the top panels (as defined in Eq. (4.19)) is depicted in the bottom row of Fig. 4.5 for the corresponding mass ratios illustrated above them.

For the equal mass case, the relative error during the inspiral and ringdown portions of the waveform are smaller than that during the times when the phenomenological intermediate model is used. Nevertheless, there is good agreement between the surrogate (namely, relative errors of at most a few percent). At the larger mass ratio of $q = 8$, the magnitude of the relative error is smaller than in the $q = 1$ case (specifically, it is less than one percent). This makes the relative error during the PN inspiral comparable to the error during the intermediate stage (the error during the ringdown is smaller, because the mass ratio $q = 8$ was a value used to fit the coefficients in the memory model, as discussed in Sec. 4.4.2.3).

To obtain more intuition about how significant a relative error of a few percent is for analyzing GW memory signals, it will be useful to compute the mismatch between the NR surrogate and our GW memory signal model. To do so, however, it is advantageous to analyze the signals in the frequency domain, in which the mismatch is typically computed. We thus turn in the next section to transforming our time-domain memory model to the frequency domain.

4.5 Frequency-domain memory signal

In this section, we compute and discuss the GW memory signal in the frequency domain. In the first subsection, we derive an analytic expressions for the frequency-domain model by taking the (continuous) Fourier transform (FT) of the time-domain model. Next, we discuss how we compute the (discrete) fast Fourier transform (FFT) of the time-domain memory model. We compare both results with the FFT of the time-domain signal computed from the NR hybrid surrogate model. In certain parts of the discussion of the frequency-domain memory signal below, we will drop the delta function term at zero-frequency, because most GW detectors cannot measure the zero-frequency component of the signal. We will note the points at which we make this approximation.

4.5.1 Analytic frequency-domain model

Our convention for the (forward) Fourier transform is

$$\tilde{H}(f) = \int_{-\infty}^{\infty} dt e^{-2\pi i f t} H(t). \tag{4.52}$$

Given the piecewise nature of our time-domain GW memory model in Eq. (4.16), its continuous Fourier transform can be written as

$$\begin{aligned}\tilde{h}_{20}(f) = &\int_{-\infty}^{t_{\rm int}} dt\, e^{-2\pi i f t} h_{20}^{\rm insp}(t) + \int_{t_{\rm int}}^{t_{\rm rd}} dt\, e^{-2\pi i f t} h_{20}^{\rm int}(t) \\ &+ \int_{t_{\rm rd}}^{\infty} dt\, e^{-2\pi i f t} \big(\Delta h_{20} - h_{20}^{\rm rd}(t)\big)\,. \end{aligned} \tag{4.53}$$

We discuss how we evaluate the three integrals on the right-hand side of Eq. (4.53) in the next three parts of this subsection.

4.5.1.1 Fourier transform of the inspiral memory signal

We write the inspiral memory model (3.5PN memory in Eq. (4.13)) as

$$h_{20}^{\text{insp}}(t) \equiv \frac{4M}{7r}\sqrt{\frac{5\pi}{6}}\eta \sum_n C_n^{\text{PN}} x^{1+n}\,, \tag{4.54}$$

where n runs over the PN orders ($n = [0, 1, 2, 2.5, 3, 3.5]$). For conciseness, we denote the coefficients that appear in the PN expansion in Eq. (4.13) as C_n^{PN}. Using the PN parameter $x(t)$ in Eq. (4.11), we find that the Fourier transform of the inspiral memory signal requires evaluating the following integrals:

$$\tilde{h}_{20}^{\text{insp}}(f) = \frac{4M}{7r}\sqrt{\frac{5\pi}{6}}\eta \sum_n C_n^{\text{PN}} (C_x)^{1+n} \int_{-\infty}^{t_{\text{int}}} dt\, e^{-2\pi i f t} (t_c - t)^{-(1+n)/4}\,. \tag{4.55}$$

We defined $C_x \equiv (1/4)(\eta/5)^{-1/4}$ to be the coefficient multiplying $(t_c - t)^{-1/4}$ in the definition of the x. After a change of variables to $y = (t_{\text{int}} - t)/(t_c - t_{\text{int}})$, we can write the integral as

$$\begin{aligned}\tilde{h}_{20}^{\text{insp}}(f) = {}& \frac{4M}{7r}\sqrt{\frac{5\pi}{6}}\eta \sum_n C_n^{\text{PN}} C_x^{(1+n)} (t_c - t_{\text{int}})^{(3-n)/4} \\ & \times e^{-2\pi i f t_{\text{int}}} \int_0^{\infty} dy\, e^{2\pi i f (t_c - t_{\text{int}}) y} (1+y)^{-(1+n)/4}\,. \end{aligned} \tag{4.56}$$

As was noted by Favata [30] for the Newtonian term, the result can be written in terms of Kummer's confluent hypergeometric function of the second kind, $U(a, b, z)$, which has the following integral representation (see, e.g., [148]):

$$\Gamma(a) U(a, b, z) = \int_0^{\infty} dt\, e^{-zt} t^{a-1} (1+t)^{b-a-1}. \tag{4.57}$$

Specifically, by choosing $a = 1$, $b = (7-n)/4$, and $z \equiv -2\pi i f(t_c - t_{\rm int})$, the inspiral memory can be written in terms of Kummer's function as

$$\tilde{h}^{\rm insp}_{20}(f) = \frac{4M}{7r}\sqrt{\frac{5\pi}{6}}\eta \sum_n C^{\rm PN}_n C^{(1+n)}_x (t_c - t_{\rm int})^{(3-n)/4} \\ \times e^{-2\pi i f t_{\rm int}} U\big(1, (7-n)/4, -2\pi i f(t_c - t_{\rm int})\big) \,. \tag{4.58}$$

To compute the value numerically, we use the implementation of the hypergeometric function ${}_2F_0$ in the `mpmath` Python package [149], which is related to the Kummer confluent hypergeometric function by

$$U(a, b, z) = z^{-a}\, {}_2F_0(a, 1 + a - b; -1/z). \tag{4.59}$$

4.5.1.2 Fourier transform of the intermediate memory signal

Taking the Fourier transform of the intermediate part of the memory model in Eq. (4.45) is reasonably straightforward. The relevant integrals that must be evaluated are

$$\tilde{h}^{\rm int}_{20}(f) = \frac{M}{r}\sum_{j=0}^{6}\int_{t_{\rm int}}^{t_{\rm rd}} dt\, e^{-2\pi i f t} c_j\, e^{p_j t} \,. \tag{4.60}$$

A short calculation shows that the frequency-domain intermediate memory model is

$$\tilde{h}^{\rm int}_{20}(f) = \frac{M}{r}\sum_{j=0}^{6}\frac{c_j}{p_j - 2\pi i f}\left[e^{(p_j - 2\pi i f)t_{\rm rd}} - e^{(p_j - 2\pi i f)t_{\rm int}}\right] \tag{4.61}$$

(recall that $p_0 = 0$). We dropped the zero-frequency Dirac delta-function contribution in Eq. (4.61).

4.5.1.3 Fourier transform of the ringdown memory signal

The FT of the ringdown memory in Eq. (4.44) is a similar to that of the intermediate signal, because it is a superposition of complex (rather than real) exponential functions. Given the somewhat involved form of the coefficients and the arguments of the complex exponentials, the result is rather lengthy:

$$\begin{aligned}\tilde{h}_{20}^{\rm rd}(f) &= \frac{-M^2}{2\sqrt{6}r}\sum_{l',l''}\sum_{m'=1}^{l'}(-1)^{m'}C_l(-2,l',m';2,l'',-m')\\ &\times\sum_{\bar{l},\bar{\bar{l}},\bar{n},\bar{\bar{n}},j,j'}\eta^{j+j'}\Bigg[C_{\bar{l}m'\bar{n}j}\bar{C}_{\bar{\bar{l}}m'\bar{\bar{n}}j'}A_{l'\bar{l}m'}\bar{A}_{l''\bar{\bar{l}}m'}\left(\frac{\omega_{\bar{l}m'\bar{n}}\bar{\omega}_{\bar{\bar{l}}m'\bar{\bar{n}}}}{\omega_{\bar{l}m'\bar{n}}-\bar{\omega}_{\bar{\bar{l}}m'\bar{\bar{n}}}}\right)\frac{e^{-i((\omega_{\bar{l}m'\bar{n}}-\bar{\omega}_{\bar{\bar{l}}m'\bar{\bar{n}}})+2\pi f)t_{\rm rd}}}{(\omega_{\bar{l}m'\bar{n}}-\bar{\omega}_{\bar{\bar{l}}m'\bar{\bar{n}}})+2\pi f}\\ &-\bar{C}_{\bar{l}m'\bar{n}j}C_{\bar{\bar{l}}m'\bar{\bar{n}}j'}\bar{A}_{l'\bar{l}m'}A_{l''\bar{\bar{l}}m'}\left(\frac{\bar{\omega}_{\bar{l}m'\bar{n}}\omega_{\bar{\bar{l}}m'\bar{\bar{n}}}}{\bar{\omega}_{\bar{l}m'\bar{n}}-\omega_{\bar{\bar{l}}m'\bar{\bar{n}}}}\right)\frac{e^{-i(-(\bar{\omega}_{\bar{l}m'\bar{n}}-\omega_{\bar{\bar{l}}m'\bar{\bar{n}}})+2\pi f)t_{\rm rd}}}{(-(\bar{\omega}_{\bar{l}m'\bar{n}}-\omega_{\bar{\bar{l}}m'\bar{\bar{n}}})+2\pi f)}\Bigg]. \end{aligned} \tag{4.62}$$

We have used the fact that $\Im[\omega_{lmn}-\bar{\omega}_{l'm'n'}]<0$ to derive the above result. The FT of the Δh_{20} term in Eq. (4.53) is simply Δh_{20} times the FT of the step function $\Theta(t-t_{\rm rd})$. We denote this contribution by

$$\tilde{h}_{20}^{\Delta}(f) = \frac{\Delta h_{20}}{2}\left[\delta(f)+\frac{e^{-2\pi i f t_{\rm rd}}}{\pi i f}\right], \tag{4.63}$$

where $\delta(f)$ is the Dirac delta function.

4.5.1.4 Fourier transform of the full memory signal

Although the time-domain memory signal is a piecewise function, each piece in the time domain has contributions at all frequencies. The, Fourier transform is linear, so the total frequency-domain memory signal is given by the sum of the different time domain parts:

$$\tilde{h}_{20}(f) = \tilde{h}_{20}^{\rm insp}(f) + \tilde{h}_{20}^{\rm int}(f) + \tilde{h}_{20}^{\Delta}(f) - \tilde{h}_{20}^{\rm rd}(f). \tag{4.64}$$

The inspiral, intermediate, offset, and ringdown expressions are given in Eqs. (4.58), (4.61), (4.63) and (4.62), respectively.

In some of the results below, we find it useful to compare our memory model with a step-function approximation of the memory, where the memory signal starts from zero and jumps to the final value computed from the final memory fit in Eq. (4.15), at the peak time of the $l = 2$, $m = 2$ mode of the waveform $t/M = 0$:

$$h_{20}^{\rm step} = \Delta h_{20} \Theta(t). \tag{4.65}$$

The FT of a step function was computed above, so we immediately write down the Fourier transform:

$$\tilde{h}_{20}^{\rm step}(f) = \frac{\Delta h_{20}}{2} \left[\delta(f) + \frac{1}{2\pi i f} \right]. \tag{4.66}$$

Although we include the Dirac delta-function term in our analytical expression, when we compute the signal for any nonzero frequency range, it does not contribute to the expression (and we will often ignore it henceforth). Thus, the step-function approximation to the memory is proportional to $1/f$ at all nonzero frequencies.

4.5.2 Continuous and discrete Fourier transforms of the memory signal

In this section, we compare our analytical frequency-domain model with the numerical FFT of the surrogate model, the FFT of the time-domain memory model, and the step-function approximation. We first discuss some subtleties related to the computation of the FFT of the memory signal, which were also recently discussed in [110,111]. Our resolution to these subtleties was not discussed in either paper [110, 111].

The main subtleties arise from the fact that the memory signal asymptotes to a nonzero value and is computed for a finite stretch of time. When taking the discrete Fourier transform (or FFT) of this time series, the series is represented as a periodic function. Thus, the offset between the earliest time and the final time in the time series behaves like a discontinuity when decomposed into the periodic basis of the FFT. This discontinuity produces an artificial $1/f$ fall off at high frequencies associated with the difference between the initial and final values of the memory signal.

This data artifact in the FFT can be mitigated by windowing (or tapering) the signal. However, as was shown recently in [110, 111], the choice of the window can produce other artifacts in the memory signal. Methods were introduced to improve these effects, specifically the linear subtraction method of [110] and the symbolic sigmoid subtraction of [111]. We discuss a different method for mitigating the effects of the final offset when computing the FFT of the memory signal.

As we show in Appendix G, one can compute the Fourier transform of the memory effect from the time derivative of the memory signal (namely, the integrand, which is proportional to the product of two $\dot{h}_{lm}$ modes). Using the Fourier integral theorem, we find that the time derivative of the memory signal determines the Fourier transform, up to a term involving a delta function at zero frequency—see, Eq. (G.7). Because we compute the memory signal for frequencies $f > 0$, then we will use Eq. (G.7) to compute the FFT, and we will neglect the zero-frequency contribution. We will compare the result with our analytical expression for the frequency-domain waveform to show the efficacy of this procedure.

Specifically, Fig. 4.6 shows the amplitude of the FFT of the surrogate memory signal (solid blue), the analytical FT (solid orange), the FFT of our time-domain model (dashed maroon), and the step-function approximation (dotted-dashed gray). We show the two mass ratios $q = 1$ (top panel) and $q = 8$ (bottom panel), which are the largest and smallest mass ratios in our model.

At low frequencies, all four curves have the same $1/f$ behavior, which is related to the fact that the signal has the same final offset in all cases. For $q = 1$, the surrogate and the model converges more rapidly to the step function approximation at low frequencies than they do for $q = 8$. We suspect this occurs for a similar reason to why the inspiral memory signal in the time domain fits less well at $q = 8$ than at $q = 1$. For a fixed frequency, the higher mass ratio system is more relativistic at this frequency, thereby making the PN inspiral contribution more important. Thus, a larger fraction of the memory accumulates from frequencies lower than those depicted in the figure, thereby making the signal converge more slowly to the step-function approximation.

As the frequency increases, the amplitude of all the signals (aside from the step-function approximation) fall off faster than $1/f$. This occurs because the surrogate

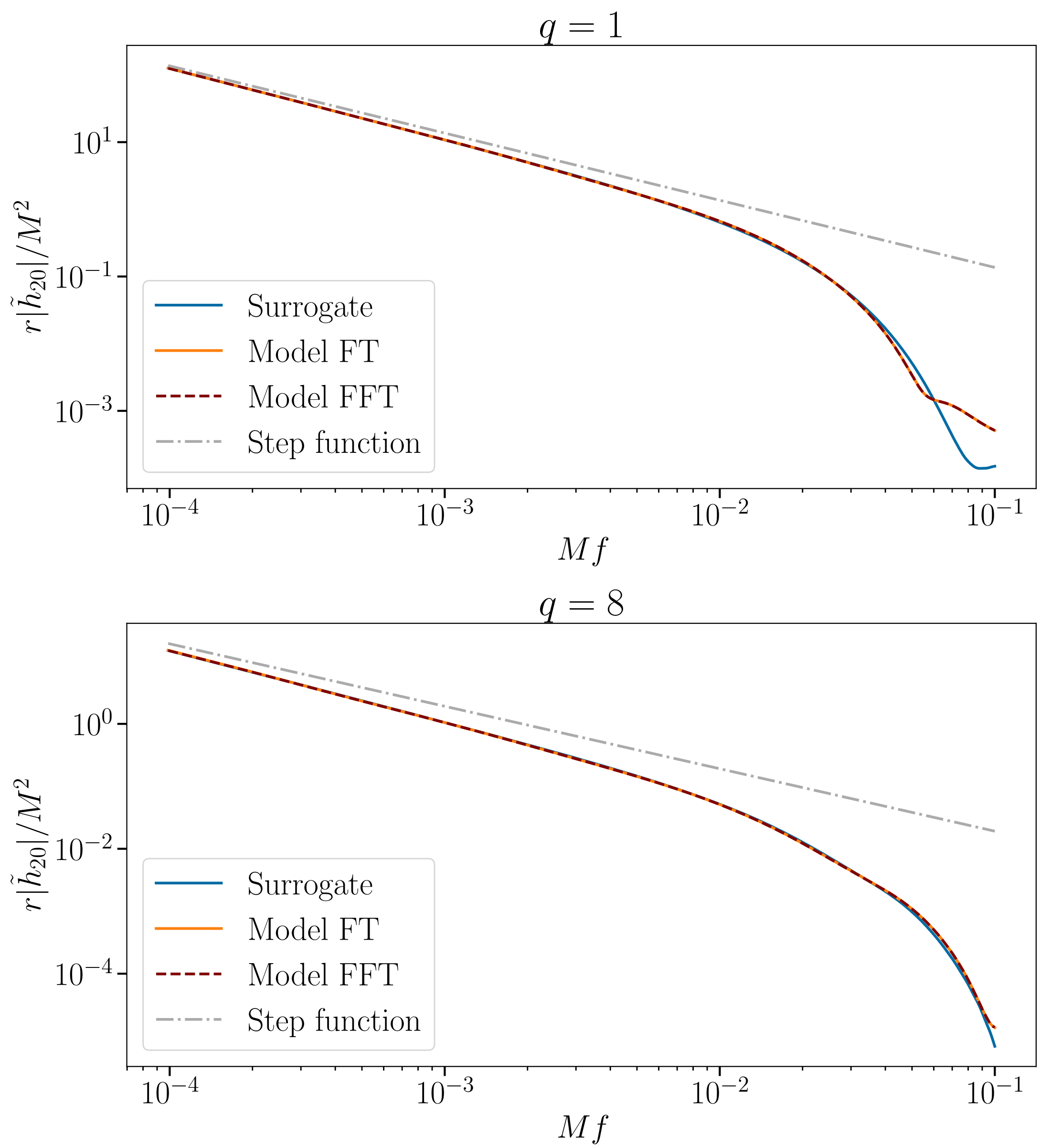

Figure 4.6: **Amplitude of the frequency-domain GW memory signal versus frequency**: The FFT of the surrogate (solid blue), the analytical FT of the time domain signal (solid orange), the FFT of the time-domain model (dashed maroon), and the step-function model (dashed-dotted gray) are shown for a mass ratio $q = 1$ (top panel) and $q = 8$ (bottom panel). We discuss the qualitative features of the memory signals in the text of Sec. 4.5.2.

and the memory models are smoother functions, which causes the frequency-domain representation of the functions to fall off more rapidly. At frequencies above $Mf \sim 10^{-1}$, both the surrogate model and our memory model have high frequency artifacts. For the memory model, the artifacts are related to the finite order of continuity of the derivatives of the time-domain signal at the times $t_{\rm int}$ and $t_{\rm rd}$. We do not know the origin of the artifacts in the surrogate, but we suspect that they are related to the finite accuracy of the interpolant which forms the basis of the surrogate model. Thus, we do not use the model above the dimensionless frequency of 0.1.

4.5.3 Mismatch results for advanced LIGO

With the analytical frequency-domain waveform or the FFT of the time-domain waveform, we can compute the mismatch, so as to assess the performance of our memory waveform model. The mismatch is defined to be

$$\mathcal{M} = 1 - \frac{\langle h_{\rm surr}, h_{\rm model}\rangle}{\sqrt{\langle h_{\rm surr}, h_{\rm surr}\rangle\langle h_{\rm model}, h_{\rm model}\rangle}}, \tag{4.67}$$

where the noise-weighted inner product between two signals h_1 and h_2 is given by

$$\langle h_1, h_2\rangle = 4\mathcal{R}\Big[\int_{f_{\rm min}}^{f_{\rm max}} df \frac{\tilde{h}_1(f)\bar{\tilde{h}}_2(f)}{S_n(f)}\Big]. \tag{4.68}$$

For the noise power-spectral density, $S_n(f)$, we use the "low" advanced LIGO design sensitivity curve for the fourth observing run, which was used in [36] and can be downloaded from [150].

We compute the mismatch between the memory signal computed from the FFT

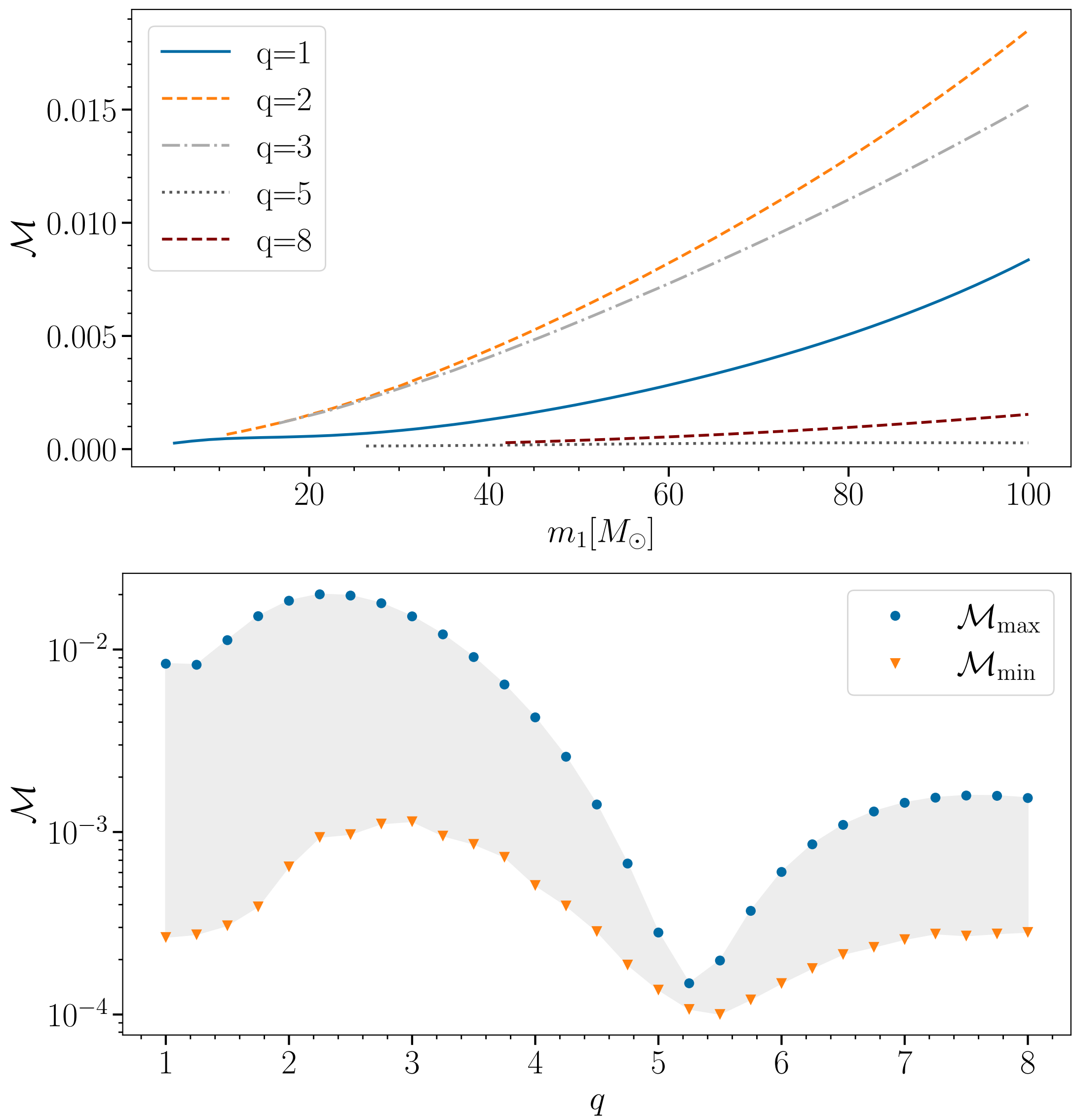

Figure 4.7: **Mismatch for different binary masses and mass ratios**: *Top*: The mismatch $\mathcal{M}$ between our GW memory model and memory computed directly from the surrogate model versus the primary mass m_1 for BBH systems with different mass ratios $1 \leq q \leq 8$. The specific mass ratios considered are indicated in the figure legend. We require that both the primary and secondary masses be greater than $5\,M_\odot$ but less than $100\,M_\odot$. *Bottom*: The maximum and minimum mismatch versus mass ratio shown in the blue circles and orange triangles, respectively. The gray shaded region shows the range of mismatch values between the minimum and maximum.

of our time-domain model and the memory signal computed from the FFT of the surrogate, for different BBH systems with mass ratios $1 \leq q \leq 8$. The resulting mismatch for a selection of BBH systems with mass ratios ($q \in \{1, 2, 3, 5, 8\}$) is shown in the top panel of Fig. 4.7 as a function of the primary mass m_1. We choose this mass to be in a range typical of the LIGO BBH detections, namely $m_1 \in [5, 100]M_\odot$. We also require that m_2 is in this same range, so the least massive primary mass increases as a function of increasing mass ratio.

The mismatch increases with the primary mass m_1 for all mass ratios. This occurs because lower mass systems have more of the low-frequency $1/f$ behavior of the memory signal in the LIGO band, which is the range of frequencies at which our model and the surrogate model most closely agree. At higher frequencies, the part of the signal that is in the LIGO band for more massive binaries, there is a larger disagreement between the two. The bottom panel shows the maximum and minimum mismatch versus the mass ratio q. The gray shaded region spans the same range of primary masses shown in the top panel; however the plot is displayed against the mass ratio on the horizontal axis. The typical mismatch is of order 10^{-3}, with a range that spans an order of magnitude. This should be a sufficient accuracy for most analyses with LVK data.

Because the mismatch computes a normalized inner product between two GW signals, it is a measure of the alignment of the two signals, which depends primarily on the phase. As we show in Appendix D, however, the phase of the complex frequency-domain memory signal does not evolve significantly over the frequency range of our model. It would also be useful to determine how well the amplitudes of two waveforms agree. For this purpose, we introduce a "signal-to-noise ratio (SNR) mismatch" which we define to be the following normalized difference of the optimal SNRs:

$$\mathcal{M}_\rho = \frac{\rho_{\rm surr} - \rho_{\rm model}}{\rho_{\rm surr}}. \tag{4.69}$$

We use the notation $\rho = \sqrt{\langle h, h \rangle}$ to denote the SNR.

Figure 4.8 shows the SNR mismatch for BBH systems with different mass ratios. The top panel depicts the same mass ratios as in Fig 4.7, though now for the SNR mismatch is plotted against m_1 on the vertical axis. The bottom panel shows the range of SNR mismatches against the mass ratio. Note that the SNR mismatch is no longer a monotonic function of the primary mass, and can be either positive or negative (namely, there are mass ratios for which the model can either underestimate or overestimate the SNR). However, there is a trend that at smaller mass ratios, the SNR mismatch increases with increasing primary mass. This likely occurs for reasons similar to those described above for the mismatch.

4.6 Conclusions

In this paper, we continued our development of waveform models of the gravitational-wave memory effect, which was initiated in Paper I [143]. We produced a time-domain waveform model for the memory effect, which covered the inspiral, merger and ring-down stages of the waveform. During the inspiral, we used an existing 3.5PN mem-

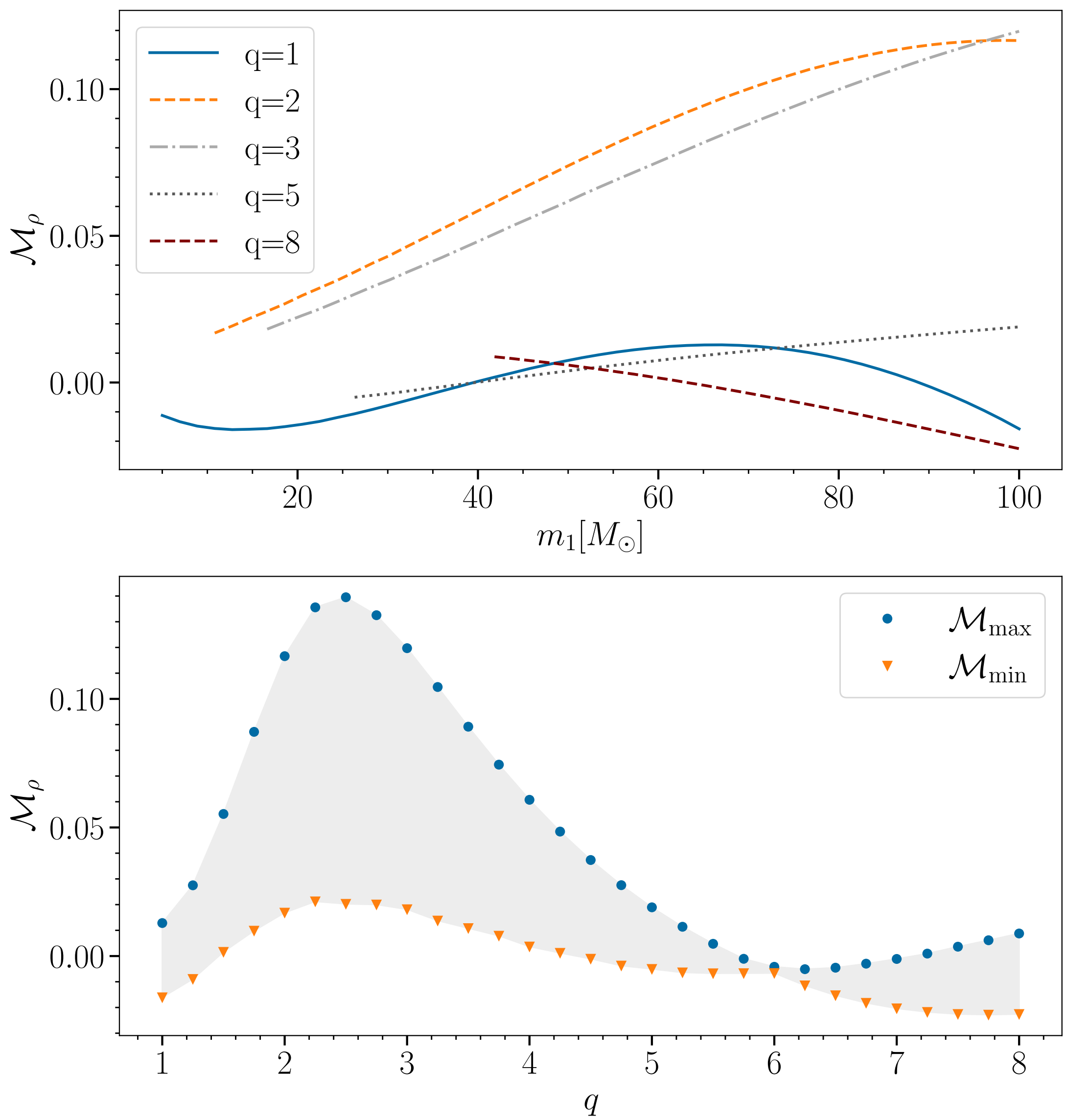

Figure 4.8: **SNR mismatch for different binary masses and mass ratios**: This figure is similar to Fig. 4.7 but it shows the SNR mismatch. *Top*: The SNR mismatch $\mathcal{M}_\rho$ versus the primary mass m_1, which is computed for different BBH systems with different mass ratios $1 \leq q \leq 8$ given in the legend of the figure. *Bottom*: The maximum and minimum SNR mismatch versus mass ratio. The orange triangles are again the minimum and the blue circles are the maximum. The gray shaded region indicates the range of values between the extremes.

ory signal, which we calibrated by hybridizing the result to the NR hybrid surrogate model. During the ringdown, we performed multimode QNM fitting for three oscillatory modes, and used the multimode fits to compute the corresponding memory signal. Finally, we used a phenomenological ansatz for an intermediate temporal region of the memory waveform between the inspiral and merger-ringdown phases. The memory signal over the three regions is continuous and had continuous first and second derivatives. The model was calibrated to the NR surrogate, and it spanned a parameter space of nonspinning binary black holes with mass ratios from one to eight.

The Fourier transform of the full time-domain signal (inspiral, intermediate and ringdown stages) was computed analytically. We also computed the FFT of the time-domain signal, which agreed with the analytical model. We assessed the performance of the model by computing the mismatch between the memory computed from the surrogate and from our time-domain model. The mismatch, at its largest, was of order 10^{-2}, but the typical value was of order 10^{-3}. We also introduced an SNR mismatch to better determine how well the amplitudes of the two signals agreed or disagreed. The SNR mismatch was an order of magnitude larger than the usual mismatch.

In future work, we would like to generalize this model to include the effects of black-hole spins. We would start with spins aligned or anti-aligned with the orbital angular momentum. Precession could also be added later. Given the relatively simple form of the frequency-domain amplitude and phase of the memory signal, we also plan to investigate purely phenomenological frequency-domain waveforms that directly model the signal in the Fourier domain. Covering a larger region of the BBH parameter space is important for being able to use the model to analyze GW data from the LVK collaboration. This will be the primary application of this and future iterations of our GW memory signal model.

Acknowledgments

A.E. and D.A.N. acknowledge support from the NSF grants No. PHY-2011784 and No. PHY-2309021. D.A.N. also acknowledges support from the NSF CAREER Award PHY-2439893.

Appendix

Appendix A

Conversion between STF tensors and spherical harmonics

In this section, we compare our expressions for the difference in the intrinsic and center-of-mass angular momentum from the Wald-Zoupas values in Eqs. (2.52) and (2.55) to a related result obtained by Compère *et al.* in [69]. We start with the intrinsic angular momentum terms, and we make this comparison by converting the u integral of the expression in Eq. (4.16) of [69] for the intrinsic angular momentum in terms of STF l-index tensors $U_L \equiv U_{\langle i_1 \dots i_l \rangle}$ and $V_L \equiv V_{\langle i_1 \dots i_l \rangle}$ to the multipole moments U_{lm} and V_{lm} used in this paper (the angle brackets around indices mean that the symmetric, trace-free part of the tensor should be taken). We focus on the second term in Eq. (4.16) of [69] which represents the difference from the Wald-Zoupas value of the angular momentum. We denote this correction term by $\delta J_i^{(\alpha=\beta=1)}$, where the index i means the angular momentum was computed with respect to a vector on the 2-sphere $Y_i^A = \epsilon^{AB} D_B n_i$. The quantity n_i is a unit vector in quasi-Cartesian coordinates that is constructed from spherical polar coordinates (θ, ϕ) as follows

$$n_i = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)\,. \tag{A.1}$$

The expression for $\delta J_i^{(\alpha=\beta=1)}$ from [69] is given by

$$\delta J_i^{(\alpha=\beta=1)} = -\sum_{l\geq 2}(l+1)^2\mu_{l+1}\left(b_l U_{iL}V_L - b_{l+1}U_L V_{iL}\right)\,. \tag{A.2}$$

The coefficients b_l (not to be confused with $b_{lm}^{(\pm)}$ defined in the main text) and μ_l were defined in [69] to be

$$b_l = \frac{2l}{l+1}\,, \tag{A.3a}$$

$$\mu_l = \frac{(l+1)(l+2)}{(l-1)l l!(2l+1)!!}\,. \tag{A.3b}$$

To rewrite Eq. (A.2) in terms of U_{lm} and V_{lm} modes, we relate the spherical harmonics Y^{lm} to the symmetric trace-free tensors of rank-l (STF-l tensors) $N_L = n_{\langle i_1}\dots n_{i_l\rangle}$ using the result in [61]

$$Y^{lm} = \mathcal{Y}_L^{lm} N_L\,. \tag{A.4}$$

The tensors $\mathcal{Y}_L^{lm}$ with $-l \leq m \leq$ are a basis for the vector space of l-index STF tensors and are defined in [61] (we do not need their explicit form here). They transform under complex conjugation in the same way as the scalar spherical harmonics:

$$\bar{\mathcal{Y}}_L^{lm} = (-1)^m \mathcal{Y}_L^{l,-m}\,. \tag{A.5}$$

The STF mass and current moments U_L and V_L are related to U_{lm}, V_{lm}, and $\mathcal{Y}_L^{lm}$

by

$$U_L = \frac{l!}{4}\sqrt{\frac{2l(l-1)}{(l+1)(l+2)}}\sum_{m=-l}^{l} U^{lm}\mathcal{Y}_L^{lm}\,, \tag{A.6a}$$

$$V_L = -\frac{(l+1)!}{8l}\sqrt{\frac{2l(l-1)}{(l+1)(l+2)}}\sum_{m=-l}^{l} V^{lm}\mathcal{Y}_L^{lm}\,; \tag{A.6b}$$

see, e.g., Eq. (2.10) of Ref. [87]. It is useful to make the definitions

$$s_l \equiv \frac{l!}{4}\sqrt{\frac{2l(l-1)}{(l+1)(l+2)}}\,, \tag{A.7a}$$

$$g_l \equiv -\frac{(l+1)!}{8l}\sqrt{\frac{2l(l-1)}{(l+1)(l+2)}}\,, \tag{A.7b}$$

though note that s_l and g_l should not be confused with $s^{l,(\pm)}_{l';l''}$ or $g^{l}_{l',m';l'',m''}$ defined in the main text. By substituting the STF moments into Eq. (A.2), we can write $\delta J_i^{(\alpha=\beta=1)}$ as

$$\delta J_i^{(\alpha=\beta=1)} = \sum_{l\geq 2}(l+1)^2\mu_{l+1}\sum_{m,m'}\big(b_l s_{l+1} g_l U_{l+1,m'}\bar{V}_{lm} \\ -b_{l+1}s_l g_{l+1}\bar{U}_{lm}V_{l+1,m'}\big)\,\bar{\mathcal{Y}}_L^{lm}\mathcal{Y}_{iL}^{l+1,m'}\,. \tag{A.8}$$

We used the properties in Eqs. (2.38) and (A.5) to simplify the result. The quantity $\bar{\mathcal{Y}}_L^{lm}\mathcal{Y}_{iL}^{l+1,m'}$ can be written in terms of Clebsch-Gordan coefficients using Eq. (2.26b) of [61], and it is only non-zero only when m' satisfies $m'=m$ or $m'=m\pm 1$ (though note that we need to multiply the result in [61] by a factor of 4π to account for the different normalization of the spherical harmonics used in [69]). Evaluating the relevant Clebsch-Gordon coefficients gives

$$
\begin{aligned}
\delta J_i^{(\alpha=\beta=1)} &= \sum_{l\geq 2,m} \mu_{l+1} \frac{(l+1)(2l-1)!!}{l!} \sqrt{(2l+3)(2l+1)} \\
&\times \Big[\left(b_l s_{l+1} g_l U_{l+1,m} \bar{V}_{lm} - b_{l+1} s_l g_{l+1} \bar{U}_{lm} V_{l+1,m} \right) c_{lm} \xi_i^0 \\
&+ \left(b_l s_{l+1} g_l U_{l+1,m+1} \bar{V}_{lm} - b_{l+1} s_l g_{l+1} \bar{U}_{lm} V_{l+1,m+1} \right) b_{lm}^{(+)} \xi_i^1 \\
&+ \left(b_l s_{l+1} g_l U_{l+1,m-1} \bar{V}_{lm} - b_{l+1} s_l g_{l+1} \bar{U}_{lm} V_{l+1,m-1} \right) b_{lm}^{(-)} \xi_i^{-1} \Big] ,
\end{aligned} \tag{A.9}
$$

where the basis vectors ξ_i^0 and $\xi_i^{\pm 1}$ are defined in Eq. (2.15) of [61]:

$$
\xi_i^0 = \delta_i^z \,, \qquad \xi_i^{\pm 1} = \frac{1}{\sqrt{2}} (\mp \delta_i^x - i \delta_i^y) \,. \tag{A.10}
$$

To relate the multipole moments of the angular momentum to the components of the angular momentum in inertial Minkowski coordinates, we follow a procedure similar to that described in [24, 25]. First we note that one can write the magnetic-parity vector harmonics as

$$
\bar{T}^A_{(b),1m} = \omega^i_{1m} \epsilon^{AB} D_B n_i \,, \tag{A.11}
$$

where the ω^i_{1m} are then given by

$$
\omega^x_{10} = 0 \,, \qquad \omega^y_{10} = 0 \,, \qquad \omega^z_{10} = \frac{1}{2}\sqrt{\frac{3}{2\pi}} \,, \tag{A.12a}
$$

$$
\omega^x_{1\pm 1} = \mp \frac{1}{4}\sqrt{\frac{3}{\pi}} \,, \qquad \omega^{0y}_{1\pm 1} = \frac{i}{4}\sqrt{\frac{3}{\pi}} \,, \qquad \omega^{0z}_{1\pm 1} = 0 \,. \tag{A.12b}
$$

Because the angular momentum is a linear functional of the vector field Y^A, then the relationship between $\delta J_{1m}^{(\alpha=\beta=1)}$ and $\delta J_i^{(\alpha=\beta=1)}$ is given by

$$\delta J_{1m}^{(\alpha=\beta=1)} = \omega_{1m}^i \delta J_i^{(\alpha=\beta=1)} \,. \tag{A.13}$$

After substituting Eq. (A.9) into Eq. (A.13), we find that

$$\delta J_{10}^{(\alpha=\beta=1)} = \frac{1}{16}\sqrt{\frac{3}{2\pi}} \sum_{l\geq 2,m} a_l c_{lm} (\bar{U}_{lm} V_{l+1,m} - \bar{V}_{lm} U_{l+1,m}), , \tag{A.14a}$$

$$\delta J_{1\pm1}^{(\alpha=\beta=1)} = \frac{1}{32}\sqrt{\frac{3}{\pi}} \sum_{l\geq 2,m} a_l b_{lm}^{(\pm)} (\bar{U}_{lm} V_{l+1,m\pm1} - \bar{V}_{lm} U_{l+1,m\pm1}) \,, \tag{A.14b}$$

where each term in the sum is a factor of $l+1$ larger than in Eq. (2.52) as noted in the text after that equation.

We next perform a similar check for the center-of-mass angular momentum. Since only the β-dependent term was computed in [69], we convert their expression in terms of STF tensors and compare it to the β-dependent term in Eq. (2.54). We start from Eq. (4.17) of [69], and we denote the second term by $\delta k_i^{(\beta=1)}$, which is given by

$$\delta k_i^{(\beta=1)} = \sum_{l\geq 2} \left[(l+1)\mu_{l+1} \left(U_{iL} U_L + b_l b_{l+1} V_{iL} V_L \right) + \frac{1}{2}\sigma_l \epsilon_{ijk} U_{jL-1} V_{kL-1} \right] . \tag{A.15}$$

The coefficient σ_l is defined in [69] by

$$\sigma_l = \frac{8(l+2)}{(l-1)(l+1)!(2l+1)!!} \,. \tag{A.16}$$

We perform the same procedure of converting the l-index STF mass and current moments into the U_{lm} and V_{lm}. The β-dependent difference term in the CM can then be written as follows:

$$\delta k_i^{(\beta=1)} = \sum_{l\geq 2,m} \frac{(2l+1)!!}{l!}\Big\{\mu_{l+1}s_l s_{l+1}\sqrt{\frac{(2l+3)}{(2l+1)}}\Big[(\bar{U}_{lm}U_{lm}+\bar{V}_{lm}V_{lm})c_{lm}\xi_i^0$$
$$+(\bar{U}_{lm}U_{l,m+1}+\bar{V}_{lm}V_{l,m+1})\frac{b_{lm}^{(+)}}{\sqrt{2}}\xi_i^1+(\bar{U}_{lm}U_{l,m-1}+\bar{V}_{lm}V_{l,m-1})\frac{b_{lm}^{(-)}}{\sqrt{2}}\xi_i^{-1}\Big]$$
$$+\frac{im}{2l}\sigma_l s_l g_l \bar{U}_{lm}V_{lm}\xi_i^0-\frac{d_{lm}^{(+)}}{\sqrt{2}}\bar{U}_{lm}V_{l,m+1}\xi_i^1+\frac{d_{lm}^{(-)}}{\sqrt{2}}\bar{U}_{lm}V_{l,m-1}\xi_i^{-1}\Big\} \quad \text{(A.17)}$$

To relate the multipole moments of the CM angular momentum to its components in inertial Minkowski coordinates, we follow the same procedure as with the intrinsic angular momentum. We first write the electric-type vector harmonics as

$$\bar{T}^A_{(e),1m} = \omega^i_{1m} D^A n_i, \quad \text{(A.18)}$$

where the coefficients ω^i_{1m} are given in Eq. (A.12). We can then solve for the multipole moments of the CM angular momentum given the relation

$$\delta k_{1m}^{(\beta=1)} = \omega^i_{1m}\delta k_i^{(\beta=1)}. \quad \text{(A.19)}$$

Using Eqs. (A.12) and (A.19) with Eq. (A.17), we find that the multipole moments of the CM angular momentum are

$$\delta k_{1,0}^{(\beta=1)} = -\frac{1}{16}\sqrt{\frac{3}{2\pi}}\sum_{l\geq 2,m}\frac{1}{l+1}\Big[a_l c_{lm}\left(\bar{U}_{lm}U_{l+1,m}+\bar{V}_{lm}V_{l+1,m}\right)-\frac{2im}{l}\bar{U}_{lm}V_{lm}\Big], \quad \text{(A.20a)}$$

$$\delta k_{1,\pm 1}^{(\beta=1)} = -\frac{1}{32}\sqrt{\frac{3}{\pi}}\sum_{l\geq 2,m}\frac{1}{l+1}\Big[a_l b_{lm}^{(\pm)}\left(\bar{U}_{lm}U_{l+1,m\pm 1}+\bar{V}_{lm}V_{l+1,m\pm 1}\right)$$
$$\pm\frac{2i}{l}d_{lm}^{(\pm)}\bar{U}_{lm}V_{l,m\pm 1}\Big]. \quad \text{(A.20b)}$$

This is identical to the result in Eq. (2.54).

Appendix B

Properties of the memory integrand

In this appendix, we give a few supplementary results on the memory integrands dh_{20}/dt and dh_{20}/dv, which highlight other features of the contributions of higher multipole and higher PN terms to the memory signal.

B.1 Memory integrand at different multipole orders

In Fig. B.1, the integrand for the memory (for an integral over time), $(r/\eta^2)dh_{20}^{(\bar{\ell},22)}/dt$, is shown on the top left, and the integrand for an integral over velocity, $(r/\mu)dh_{20}^{(\bar{\ell},22)}/dv$ is shown in the top-right panel, for different values of $\bar{\ell}$. The velocity values and the line styles for the curves are identical to those in the left panels of Fig. 3.4. The bottom panels show the relative contributions of higher multipoles to the full memory computed with all multipoles up to $\bar{\ell} = 25$ at 22PN order. The notation used is analogous to that defined in Eq. (3.39), with the memory integrand replacing the memory signal.

The main feature to highlight in Fig. B.1 is that while dh_{20}/dt is a monotonically increasing function with v, the velocity integrand dh_{20}/dv has a peak near $v \approx 0.3$ for

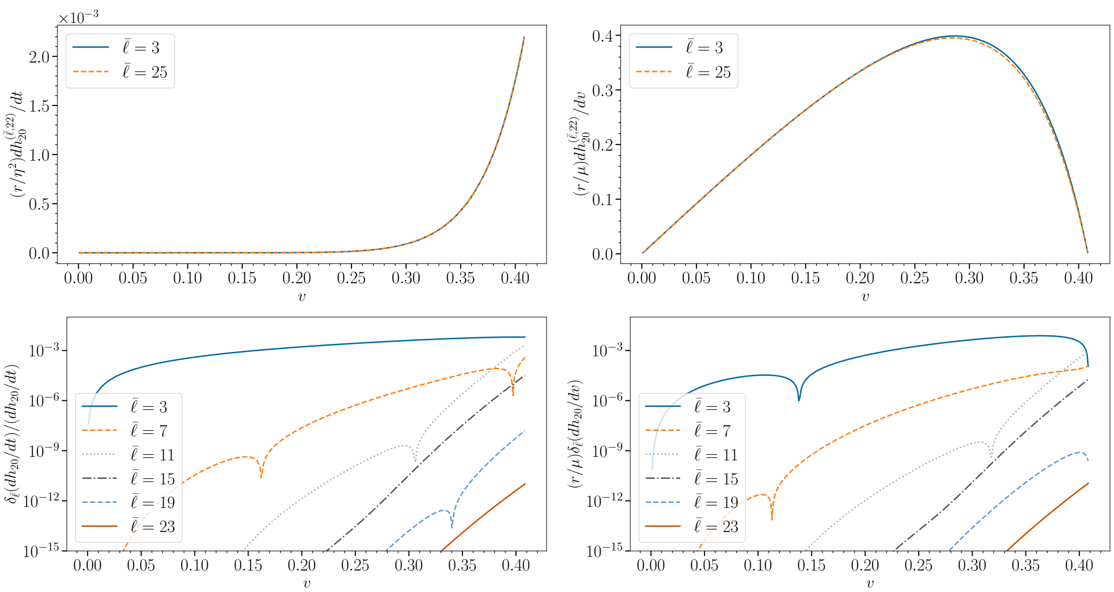

Figure B.1: **Memory integrand versus velocity for different multipole orders**: This figure is the analogue of the left column of Fig. 3.4, but here the left column shows $(r/\eta^2)dh_{20}^{(\bar{\ell},22)}/dt$ in the top panel and the relative contribution from higher modes, $\delta_{\bar{\ell}}(dh_{20}/dt)/(dh_{20}/dt)$, in the bottom panel, where the $\delta_{\bar{\ell}}$ notation is defined analogously to the expression for the memory signal in Eq. (3.39). The right column shows the integrand of the memory $(r/\mu)dh_{20}^{(\bar{\ell},22)}/dv$ in the top panel when the integral is performed over velocity rather than time (as in the left column). The fractional relative contribution from higher multipoles is in the bottom panel of the right column. The coloring and line styles of the curves correspond to the same values of $\bar{\ell}$ as in Fig. 3.4, and the values of the velocity depicted on the horizontal axis is also the same.

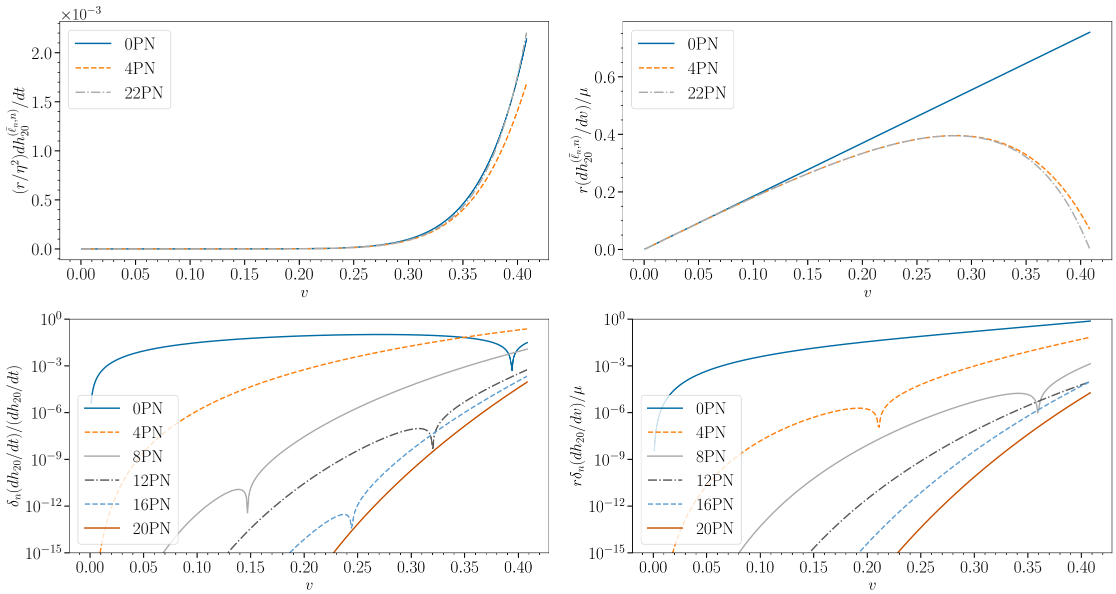

Figure B.2: **Memory integrand versus velocity for different PN orders**: This figure is the analogue of Fig. B.1 for PN order rather than multipole order. It uses the same interval of velocities and has the same line styles and colors for the different curves as is used in Fig. 3.6. Additional discussion of this figure is given in the text of Appendix B.2.

all multipole orders. This arises from the terms $(dE/dv)/(dE_{\rm GW}/dt)$ used to convert the integral with respect to time to one with respect to velocity. Otherwise, there are not any substantive differences between the relative importance of the higher spherical-harmonic modes from the memory integrand versus the memory signal.

B.2 Memory integrand at different post-Newtonian orders

Figure B.2 has several similarities with both Fig. B.1 and the left column of Fig. 3.6. Like Fig. B.1, it focuses on the memory integrand with respect to time in the left column and velocity in the right column. Like Fig. 3.6, it focuses on the contributions and results from different PN orders, and it uses the same interval of velocities and the same line styles and colors there. However, the memory integrand $(r/\eta^2)dh_{20}^{(\bar{\ell}_n,n)}/dt$ does have some more unusual properties. In the top-left panel, the Newtonian (0PN) curve agrees better with the 22PN result than the 4PN result does at higher velocities above $v \approx 0.35$, though not below. However, comparing this with $(r/\mu)dh_{20}^{(\bar{\ell}_n,n)}/dv$, the 4PN result is closer to the 22PN result for all values of v shown. Otherwise, there is a relatively clear trend of higher PN orders contributing less to the total value of the memory. Thus, we suspect that the improved performance of the Newtonian signal at larger velocities is coincidental.

Another feature worth noting about the top-right panel is that at Newtonian order dh_{20}/dv is proportional to v, which is why the 0PN curve is a straight line in the upper right panel. Higher PN orders do not have this property, and instead have a peak around $v \approx 0.3$. This was noted in Sec. 3.5.2.4 as the reason for why the Newtonian memory is notably larger than all the other higher PN cases shown.

Appendix C

Further analysis of the final memory fit

Figure C.1 is similar to Fig. 3.8 in the main text, but the fit used is $\Delta h_{20}^{\mathrm{comp}}$, which does not use any information from the EMRI calculation, whereas the fit $\Delta h_{20}^{\mathrm{EMRI}}$ in Fig. 3.8 does. We include this figure primarily for completeness.

A useful check of the EMRI fit would be to compare it with an accurate calculation of the memory effect for values of η between 0 and 0.1 (i.e., mass ratios greater than $q = 8$) to determine how the two fits perform. Specifically, we would like to determine if performing extrapolation by using $\Delta h_{20}^{\mathrm{comp}}$ is more or less accurate in this range of η than performing interpolation from the EMRI data point ($\eta \to 0$) to the next largest value of $\eta \approx 0.1$ ($q = 8$) from the hybridized surrogate. Using the EMRISur1dq1e4 surrogate model cannot address this, because it is not hybridized and spans a time interval of about $10^4 M$, whereas the EMRI calculation showed that long time spans during the inspiral are required to obtain an accurate calculation of the final memory offset accumulated from a widely separated EMRI. We cannot simply hybridize with

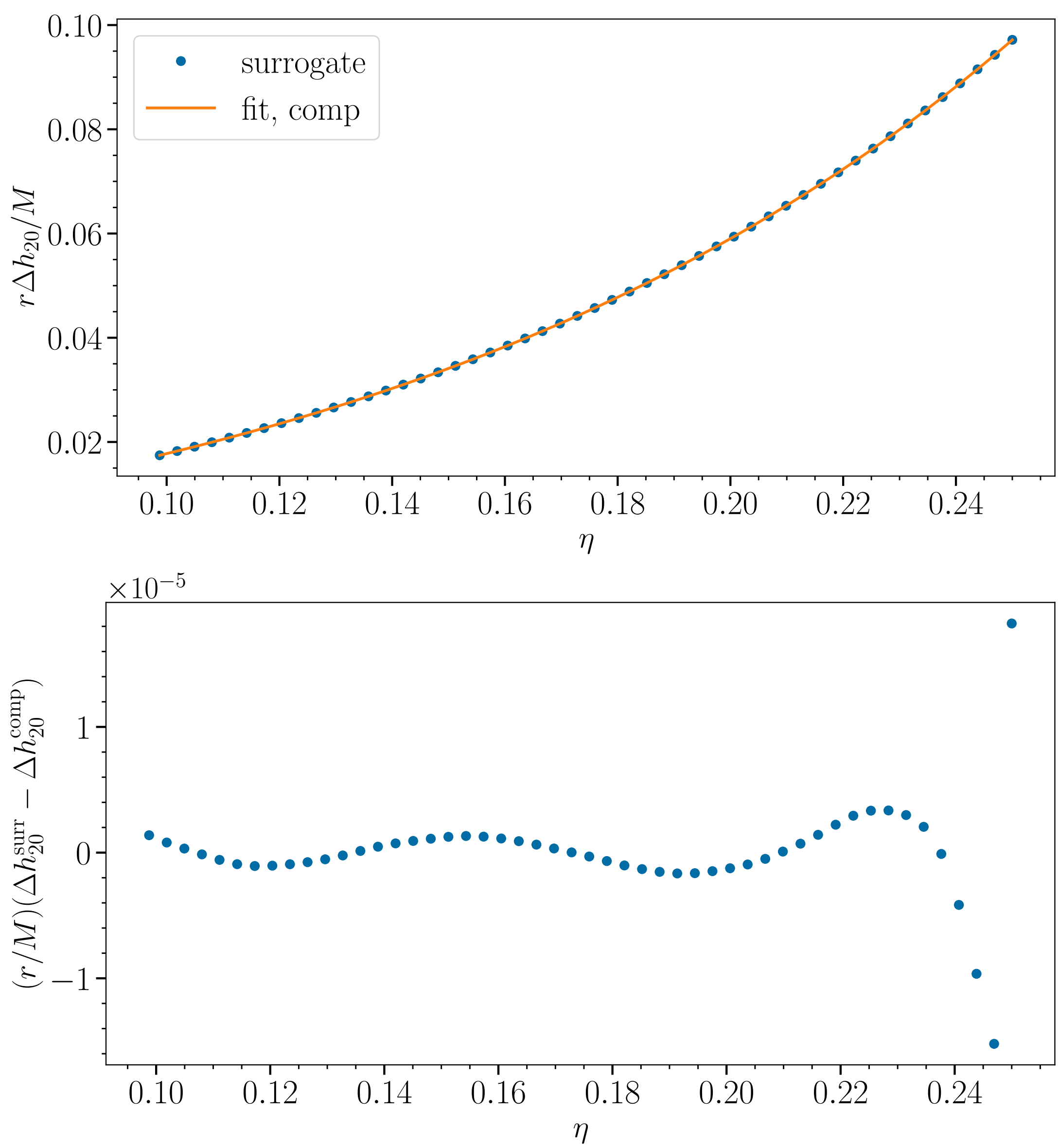

Figure C.1: **Final memory offset fit and residuals**: This figure is identical to Fig. 3.8, except we plot the final memory computed from the comparable-mass-ratio fit, $\Delta h_{20}^{\rm comp}$ as the solid orange line in the top panel, and the residual between $\Delta h_{20}^{\rm comp}$ and $\Delta h_{20}^{\rm surr}$ is shown in the bottom panel.

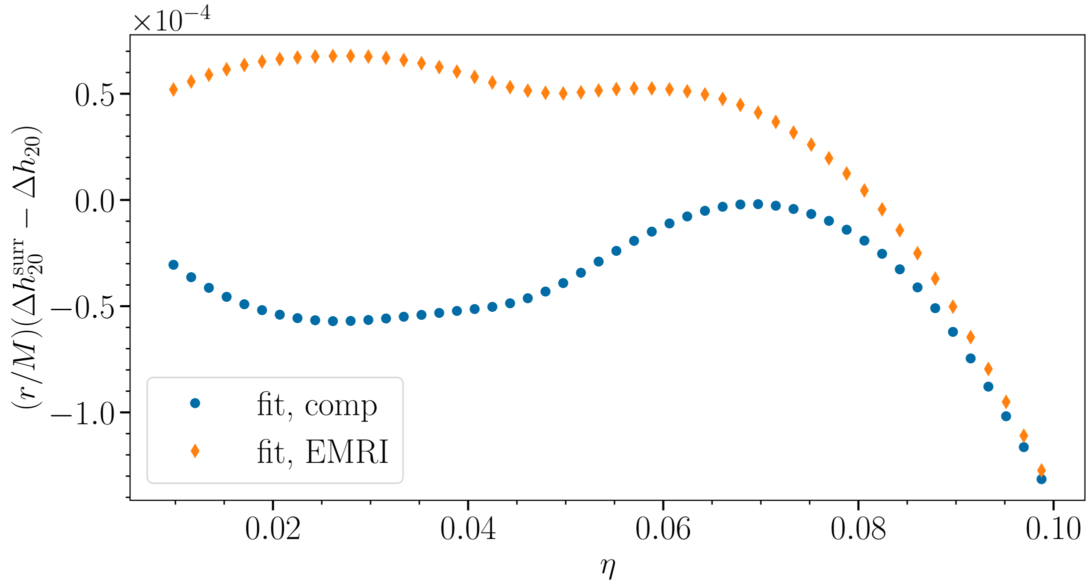

Figure C.2: **Residuals between two fitting functions and an estimate for the memory from IMRIs**: We discuss how we compute the memory from hybridizing the 3PN inspiral memory with EMRISur1dq1e4 surrogate model to obtain the $\Delta h_{20}^{\mathrm{surr}}$ in the text of Appendix C. We use this value to compute the absolute error between it and the comparable mass fit $\Delta h_{20}^{\mathrm{comp}}$ (the blue filled circles) or the EMRI fit $\Delta h_{20}^{\mathrm{EMRI}}$ (the orange filled diamonds). At larger η values, both fits tend to predict a larger final memory offset than our IMRI estimate, but at smaller η values, the EMRI surrogate is smaller, whereas the comparable fit is larger.

the PN-expanded EMRI memory signal, because it neglects order η^2 corrections, which become important as η is in the 0.01 to 0.1 range.

Instead, we opt to use the 3PN memory, which has nonlinear corrections in η, to hybridize with EMRISur1dq1e4 surrogate model, so as to obtain an estimate of the final memory offset for IMRI mass ratios. Specifically, we compute the memory for 50 BBH systems with mass ratios in the interval $8 \le q \le 100$, uniformly spaced in η, using the EMRISur1dq1e4 surrogate model. We then hybridize the surrogate memory with a 3PN waveform, over a length of time of $10^3 M$, between $-10^4 M \le t \le -9 \times 10^3 M$, as the EMRI surrogate does not extend before $-10^4 M$. This allows us to compute data to compare against both of our fitting functions $\Delta h_{20}^{\mathrm{comp}}$ and $\Delta h_{20}^{\mathrm{EMRI}}$. These residuals are shown in Fig. C.2.

At the largest values of η shown ($q = 8$) both fits give larger values of the memory than our estimate with the hybridized EMRISur1dq1e4 data, which fit the hybridized NRHybSur3dq8 data to a higher precision. The EMRISur1dq1e4 surrogate was trained on EMRI simulations and rescaled to match the numerical-relativity results at less extreme mass ratios; however, based on this memory calculation, it does not seem to be accurate to more that one percent. At the smallest η value of roughly 0.01, both fits have a comparable magnitude of the error, but the EMRI is consistently lower than the EMRI surrogate estimate, whereas the comparable mass fit is larger. Given the limitations of the data and waveform models for computing the memory in this parameter space of η, such a consistency check of the fits will likely have to wait for more numerical relativity or second-order self-force calculations to make a more definitive assessment of the accuracy of these two fitting functions in this regime.

Appendix D

Phase of the frequency-domain GW memory signals

In this appendix, we discuss the phase of the complex frequency-domain GW memory signal. To help interpret the phase, it is first helpful to consider the phase of the analytical FT of the step-function approximation to the memory signal in Eq. (4.66). The phase is the constant value of $-\pi/2$ for all frequencies $f > 0$, because the $1/f$ part of the solution has a negative, purely imaginary constant multiplying the $1/f$ term. Thus, we expect our model to approach a phase of $-\pi/2$ at low frequencies.

In Fig. D.1, we show the phase of the FFT of the surrogate memory signal, the FFT and analytical FT of our time-domain model, and the step-function approximation, for mass ratios $q = 1$ and $q = 8$ in the top and bottom panels, respectively. The solid, blue curve shows the phase of the FFT of the surrogate memory signal, while the solid orange curve shows the phase of the analytical FT of the memory signal computed from our time-domain memory model in Eq. (4.64). The figure also displays the FFT of our time-domain memory model (the dashed, maroon curve), which agrees

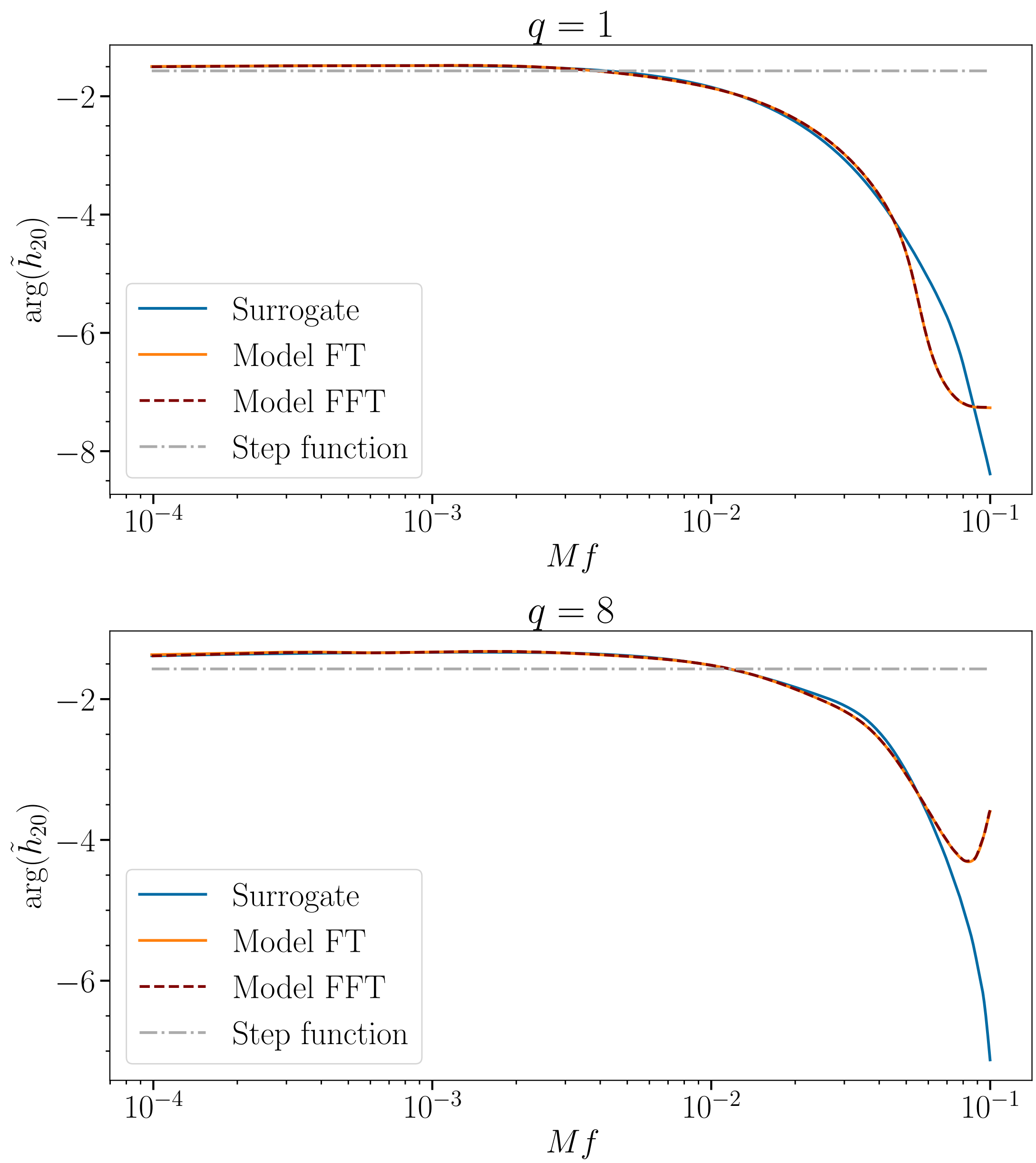

Figure D.1: **Phase of the frequency-domain memory signal versus frequency**: The FFT of the surrogate memory signal (solid blue), the analytical FT of the time-domain model (solid orange), the FFT of the time-domain model (dashed maroon), and the step-function model (dashed-dotted gray) are shown for a mass ratio $q = 1$ (top panel) and $q = 8$ (bottom panel).

well with the phase of the analytic FT of the memory model. Finally, the phase of the step-function approximation in Eq. (4.66) is shown as the dashed-dotted, gray line. The phases computed through the FFT were computed from signals that were time shifted to have the peak time reside at the end of the time series; otherwise the discrete Fouier time-shift theorem would modify the phase from that of the analytical expression.

As anticipated, at low frequencies, the phase of the FFT of the surrogate memory, the FFT of our model, and the analytic FT all approach the phase of the step-function model. The convergence of the phase for the mass ratio of $q = 8$ is slower than that of the lower mass ratio $q = 1$, as it was for the amplitude (as discussed in Sec. 4.5.2. The phase evolves by at most roughly one cycle over the three decades in frequency shown in Fig. D.1. At higher frequencies, the phase of the memory signal that was computed directly from the NR surrogate begins to differ from that of our model. It occurs at a similar frequency to that were the amplitude becomes less reliable, as well.

Appendix E

Ringdown memory model for generic QNMs

In this section, we derive the ringdown memory model for a more generic superposition of QNMs that includes both prograde and retrograde modes. We denote these QNM frequencies by ω^{+}_{lmn} and ω^{-}_{lmn}, respectively. The ringdown strain model can be written as a superposition of the two modes

$$h^{\rm rd}_{\rm model}(t) = \frac{M}{r}\sum_{l,m,n}\Big[C^{+}_{lmn}e^{-i\omega^{+}_{lmn}(t-t_0)}\,{}_{-2}S_{lm}(\theta,\phi;a\omega^{+}_{lmn}) \\ + C^{-}_{lmn}e^{-i\omega^{-}_{lmn}(t-t_0)}\,{}_{-2}S_{lm}(\theta,\phi;a\omega^{-}_{lmn})\Big]\,. \tag{E.1}$$

We can rewrite this expression in terms of only the prograde frequencies ω^{+}_{lmn} by using the following relation between the prograde and retrograde frequencies:

$$\omega^{+}_{lmn} = -\bar{\omega}^{-}_{l-mn}\,. \tag{E.2}$$

When written in terms of the prograde frequencies, Eq. (E.1) becomes

$$h^{\rm rd}_{\rm model}(t) = \frac{M}{r}\sum_{l,m,n}\Big[C^{+}_{lmn}e^{-i\omega^{+}_{lmn}(t-t_0)}\,{}_{-2}S_{lm}(\theta,\phi;a\omega^{+}_{lmn})$$

$$+ C^{-}_{lmn} e^{i\bar{\omega}^{+}_{l-mn}(t-t_0)} {}_{-2}S_{lm}(\theta, \phi; -a\bar{\omega}^{+}_{l-mn})\Big] . \tag{E.3}$$

We now make use of a few properties of the spin-weighted spheroidal harmonics ${}_sS_{lm}(\theta, \phi; c)$ under several discrete transformations that change the sign of the harmonic indices, as well as changes in the sign of the spheroidal parameter c, complex conjugation of c, and reflections about the equatorial plane. Specifically, these transformations are

$${}_s\bar{S}_{lm}(\theta, \phi; c) = (-1)^{s+m} {}_sS_{l-m}(\theta, \phi; -\bar{c}), \tag{E.4a}$$

$${}_sS_{l-m}(\theta, \phi; c) = (-1)^{s+l} {}_s\bar{S}_{lm}(\pi - \theta, \phi; -\bar{c}), \tag{E.4b}$$

$${}_{-s}S_{lm}(\theta, \phi; c) = (-1)^{l+m} {}_sS_{lm}(\pi - \theta, \phi; c) . \tag{E.4c}$$

Relabeling the index m as $-m$ in the sum involving the retrograde amplitudes, and using the property in Eq. (E.4b), we can write the final line in Eq. (E.3) as

$$h^{\rm rd}_{\rm model}(t) = \frac{M}{r} \sum_{l,m,n} \Big[C^{+}_{lmn} e^{-i\omega^{+}_{lmn}(t-t_0)} {}_{-2}S_{lm}(\theta, \phi; a\omega^{+}_{lmn})$$
$$+ (-1)^l C^{-}_{l-mn} e^{i\bar{\omega}^{+}_{lmn}(t-t_0)} {}_{-2}\bar{S}_{lm}(\pi - \theta, \phi; a\omega^{+}_{lmn})\Big] . \tag{E.5}$$

We next use the expansion of the spin-weighted spheroidal harmonics in Eq. (4.20) to write ${}_{-2}\bar{S}_{lm}(\pi - \theta, \phi; a\omega^{+}_{lmn})$ as

$${}_{-2}\bar{S}_{lm}(\pi - \theta, \phi; a\omega^{+}_{lmn}) = \sum_{\bar{l}} \bar{A}_{\bar{l}lm}(a\omega^{+}_{lmn}) {}_{-2}\bar{Y}_{\bar{l}m}(\pi - \theta, \phi) . \tag{E.6}$$

The spin-weighted spherical harmonics satisfy the properties

$$ {}_{-s}\bar{Y}_{lm}(\theta,\phi) = (-1)^{s+m}{}_{s}Y_{l-m}(\theta,\phi)\,, \tag{E.7a}$$

$$ {}_{s}\bar{Y}_{l-m}(\pi-\theta,\phi) = (-1)^{l+m}{}_{-s}Y_{l-m}(\theta,\phi)\,. \tag{E.7b}$$

With these results, we can recast Eq. (E.6) as

$$ {}_{-2}\bar{S}_{lm}(\pi-\theta,\phi;a\omega^{+}_{lmn}) = \sum_{\bar{l}}(-1)^{\bar{l}}\bar{A}_{\bar{l}lm}(a\omega^{+}_{lmn}){}_{-2}\bar{Y}_{\bar{l}-m}(\theta,\phi)\,. \tag{E.8}$$

In terms of spin-weighted spherical harmonics, the ringdown memory strain model is

$$\begin{aligned} h^{\rm rd}_{\rm model}(t) = \frac{M}{r}\sum_{\bar{l},m}\Bigg\{\sum_{l,n}\Bigg[&C^{+}_{lmn}A_{\bar{l}lm}(a\omega^{+}_{lmn})e^{-i\omega^{+}_{lmn}(t-t_0)} \\ &+(-1)^{l+\bar{l}}C^{-}_{lmn}\bar{A}_{\bar{l}l-m}(a\omega^{+}_{l-mn})e^{i\bar{\omega}^{+}_{l-mn}(t-t_0)}\Bigg]\Bigg\}{}_{-2}Y_{\bar{l}m}(\theta,\phi)\,. \end{aligned} \tag{E.9}$$

We made the substitution of $m\to -m$ in the second term. The memory model strain modes $h^{mem}_{lm}(t)$ can be read-off from this expression as

$$\begin{aligned} h^{\rm model}_{lm}(t) = \frac{M}{r}\sum_{l',n}\Bigg[&C^{+}_{l'mn}A_{ll'm}(a\omega^{+}_{l'mn})e^{-i\omega^{+}_{l'mn}(t-t_0)} \\ &+(-1)^{l+l'}C^{-}_{l'mn}\bar{A}_{ll'-m}(a\omega^{+}_{l'-mn})e^{i\bar{\omega}^{+}_{l'-mn}(t-t_0)}\Bigg]\,. \end{aligned} \tag{E.10}$$

Because we express the memory only in terms of the prograde frequencies, we drop the $+$ label on the frequency (i.e., $\omega^{+}_{lmn}\equiv\omega_{lmn}$) and write $A_{\bar{l}lm}(a\omega_{lmn})\equiv A_{\bar{l}lm}$. We compute the memory from Eq. (4.7) by substituting the ringdown model strain modes $h^{\rm model}_{lm}(t)$ from Eq. (E.10) into the expression. We integrate the product of two quasinormal modes, which for a single mode gives

$$\int_t^\infty dt' e^{-i(\omega_{\bar{l}m'\bar{n}} - \bar{\omega}_{\bar{\bar{l}}-m''\bar{\bar{n}}})(t'-t_0)} = \frac{e^{-i(\omega_{\bar{l}m'\bar{n}} - \bar{\omega}_{\bar{\bar{l}}-m''\bar{\bar{n}}})(t-t_0)}}{-i(\omega_{\bar{l}m'\bar{n}} - \bar{\omega}_{\bar{\bar{l}}-m''\bar{\bar{n}}})}. \tag{E.11}$$

We used the fact that the imaginary parts of ω_{lmn} and ω_{l-mn} are negative, so the integration result vanishes at infinity. Summing over multiple QNM modes, the ringdown memory (l, m) modes are

$$\begin{aligned}
h^{\rm mem}_{lm}(t) = i\frac{M^2}{r}\sqrt{\frac{(l-2)!}{(l+2)!}} &\sum_{l',l'',m',m''} (-1)^{m''} C_l(-2, l', m'; 2, l'', m'') \\
\times \sum_{\bar{l},\bar{\bar{l}},\bar{n},\bar{\bar{n}}} \Bigg[&\left(\frac{\omega_{\bar{l}m'\bar{n}}\,\bar{\omega}_{\bar{\bar{l}}-m''\bar{\bar{n}}}}{\omega_{\bar{l}m'\bar{n}} - \bar{\omega}_{\bar{\bar{l}}-m''\bar{\bar{n}}}}\right) C^+_{\bar{l}m'\bar{n}} \bar{C}^+_{\bar{\bar{l}}-m''\bar{\bar{n}}} A_{l'\bar{l}m'} \bar{A}_{l''\bar{\bar{l}}-m''} e^{-i(\omega_{\bar{l}m'\bar{n}} - \bar{\omega}_{\bar{\bar{l}}-m''\bar{\bar{n}}})(t-t_0)} \\
+(-1)^{l'+l''+\bar{l}+\bar{\bar{l}}} &\left(\frac{\bar{\omega}_{\bar{l}-m'\bar{n}}\,\omega_{\bar{\bar{l}}m''\bar{\bar{n}}}}{\bar{\omega}_{\bar{l}-m'\bar{n}} - \omega_{\bar{\bar{l}}m''\bar{\bar{n}}}}\right) C^-_{\bar{l}m'\bar{n}} \bar{C}^-_{\bar{\bar{l}}-m''\bar{\bar{n}}} \bar{A}_{l'\bar{l}-m'} A_{l''\bar{\bar{l}}m''} e^{i(\bar{\omega}_{\bar{l}-m'\bar{n}} - \omega_{\bar{\bar{l}}m''\bar{\bar{n}}})(t-t_0)} \\
-(-1)^{l''+\bar{\bar{l}}} &\left(\frac{\omega_{\bar{l}m'\bar{n}}\,\omega_{\bar{\bar{l}}m''\bar{\bar{n}}}}{\omega_{\bar{l}m'\bar{n}} + \omega_{\bar{\bar{l}}m''\bar{\bar{n}}}}\right) C^+_{\bar{l}m'\bar{n}} \bar{C}^-_{\bar{\bar{l}}-m''\bar{\bar{n}}} \bar{A}_{l'\bar{l}-m'} A_{l''\bar{\bar{l}}m''} e^{-i(\omega_{\bar{l}m'\bar{n}} + \omega_{\bar{\bar{l}}m''\bar{\bar{n}}})(t-t_0)} \\
+(-1)^{l'+\bar{l}} &\left(\frac{\bar{\omega}_{\bar{l}-m'\bar{n}}\,\bar{\omega}_{\bar{\bar{l}}-m''\bar{\bar{n}}}}{\bar{\omega}_{\bar{l}-m'\bar{n}} + \bar{\omega}_{\bar{\bar{l}}-m''\bar{\bar{n}}}}\right) C^-_{\bar{l}m'\bar{n}} \bar{C}^+_{\bar{\bar{l}}-m''\bar{\bar{n}}} \bar{A}_{l'\bar{l}-m'} \bar{A}_{l''\bar{\bar{l}}-m''} e^{i(\bar{\omega}_{\bar{l}-m'\bar{n}} - \bar{\omega}_{\bar{\bar{l}}-m''\bar{\bar{n}}})(t-t_0)} \Bigg].
\end{aligned} \tag{E.12}$$

Setting the amplitudes of the retrograde QNMs equal to zero reproduces the result in the main text.

Appendix F

QNM fit coefficients

We give all of the QNM fit coefficients C_{lmnj} for the $l = 2$, $m = 1$ mode and the $l = 2, 3$, $m = 2$ modes in Tables F.1 and F.2, respectively. They are given in terms of their modulus and phase.

Mode	Amplitude	Phase
C_{2100}	4.82281×10^{-2}	-1.41047
C_{2110}	5.67762×10^{-1}	2.13952
C_{2120}	3.80294×10^{0}	-1.10247
C_{2130}	1.59829×10^{1}	1.85066
C_{2140}	3.91383×10^{1}	-1.42840
C_{2150}	5.31916×10^{1}	1.619741
C_{2160}	3.71577×10^{1}	-1.58920
C_{2170}	1.03971×10^{1}	1.50051
C_{2101}	2.38039×10^{0}	2.31646
C_{2111}	1.48580×10^{1}	-0.545106
C_{2121}	7.00706×10^{1}	2.24528
C_{2131}	2.75410×10^{2}	-1.27624
C_{2141}	6.79526×10^{2}	1.64367
C_{2151}	9.34160×10^{2}	-1.62912
C_{2161}	6.58423×10^{2}	1.42574
C_{2171}	1.85623×10^{2}	-1.77847
C_{2102}	6.64711×10^{0}	-1.55498
C_{2112}	5.68596×10^{1}	1.93398
C_{2122}	3.00496×10^{2}	-1.38291
C_{2132}	1.16339×10^{3}	1.46882
C_{2142}	2.80241×10^{3}	-1.86753
C_{2152}	3.78281×10^{3}	1.15090
C_{2162}	2.62757×10^{3}	-2.07400
C_{2172}	7.31855×10^{2}	1.00835

Table F.1: The coefficients of the QNM fit for the $(2, 1)$ mode.

$l=2,\, m=2$			$l=3,\, m=2$		
Coefficient	Amplitude	Phase	Coefficient	Amplitude	Phase
C_{2200}	3.19296×10^{-1}	-2.77026	C_{3200}	3.74080×10^{-2}	0.289391
C_{2210}	1.71980×10^{0}	1.01148	C_{3210}	2.16618×10^{-1}	-2.05806
C_{2220}	5.92015×10^{0}	-1.59268	C_{3220}	5.24707×10^{-1}	2.64654
C_{2230}	1.58611×10^{1}	1.84418	C_{3230}	3.57704×10^{0}	0.622898
C_{2240}	2.88209×10^{1}	-1.19031	C_{3240}	1.06986×10^{1}	-2.26776
C_{2250}	3.24851×10^{1}	1.99983	C_{3250}	1.52928×10^{1}	0.964857
C_{2260}	2.05656×10^{1}	-1.11638	C_{3260}	1.07697×10^{1}	-2.13948
C_{2270}	5.55581×10^{0}	2.02049	C_{3270}	3.00686×10^{0}	1.01691
C_{2201}	9.27953×10^{0}	0.499780	C_{3201}	7.96203×10^{-1}	2.47208
C_{2211}	4.13466×10^{1}	-2.15539	C_{3211}	4.46748×10^{0}	0.131493
C_{2221}	1.28053×10^{2}	1.31012	C_{3221}	2.12174×10^{0}	2.38641
C_{2231}	2.91826×10^{2}	-1.67756	C_{3231}	5.97475×10^{1}	-2.14853
C_{2241}	4.54998×10^{2}	1.48876	C_{3241}	1.97603×10^{2}	1.04384
C_{2251}	4.42701×10^{2}	-1.65677	C_{3251}	2.90392×10^{2}	-2.07309
C_{2261}	2.44140×10^{2}	1.49805	C_{3261}	2.07484×10^{2}	1.07278
C_{2271}	5.95055×10^{1}	-1.62020	C_{3271}	5.85327×10^{1}	-2.07471
C_{2202}	3.31764×10^{1}	-3.10437	C_{3202}	3.27197×10^{0}	-0.783394
C_{2212}	1.68339×10^{2}	0.635284	C_{3212}	1.60781×10^{1}	-3.06586
C_{2222}	5.05646×10^{2}	-2.11001	C_{3222}	1.50722×10^{1}	-2.65323
C_{2232}	1.08443×10^{3}	1.20664	C_{3232}	2.94877×10^{2}	0.736590
C_{2242}	1.59435×10^{3}	-1.91641	C_{3242}	8.95912×10^{2}	-2.30299
C_{2252}	1.45739×10^{3}	1.19785	C_{3252}	1.27318×10^{3}	0.879673
C_{2262}	7.48094×10^{2}	-1.95375	C_{3262}	8.92803×10^{2}	-2.25037
C_{2272}	1.69904×10^{2}	1.205426	C_{3272}	2.48810×10^{2}	0.888774

Table F.2: The coefficients of the QNM fit for the $l = 2, 3$, $m = 2$ modes.

Appendix G

Alternate expression for the frequency-domain memory signal

Because the memory signal involves computing an integral of derivatives of $h(t)$, if we compute the Fourier transform of $\dot{h}(t) = dh/dt$ instead, then the memory can be evaluated from just the Fourier transform of an "instantaneous" quantities rather than a "hereditary" quantity (i.e., the time integral of the instantaneous integrand). For a signal that goes to zero as t approaches $\pm\infty$, this can be done with the Fourier integral theorem Computing the Fourier transform of quantities with a nonzero late-time value, such as the memory signal, has some subtleties. We will avoid these by splitting the the memory signal into a part that does approach zero at early and late times, and a step-function contribution as follows:

$$h_{\rm mem}(t) = \Delta h^{\rm fit}_{\rm mem}\Theta(t) + h^{\rm diff}_{\rm mem}(t)\,. \tag{G.1}$$

The quantity $h^{\rm diff}_{\rm mem}(t)$ has the same time dependence as the memory signal for $t < 0$; for $t > 0$, it has the same time dependence as the memory, but it is offset by $-\Delta h^{\rm fit}$ from the memory signal. Thus, it has a discontinuity at $t = 0$, but it smoothly goes to zero as $t \to \pm\infty$. Taking the derivative of the expression gives

$$\dot{h}_{\mathrm{mem}}(t) = \Delta h^{\mathrm{fit}}_{\mathrm{mem}} \delta(t) + \dot{h}^{\mathrm{diff}}_{\mathrm{mem}}(t)\,. \tag{G.2}$$

The fact that the derivative (in a distributional sense) of a step function is a Dirac delta function was used above. Note that the derivatives of $h_{\mathrm{mem}}(t)$ and $h^{\mathrm{diff}}_{\mathrm{mem}}(t)$ agree everywhere except $t = 0$, where $\dot{h}_{\mathrm{mem}}(t)$ is well defined, but $\dot{h}^{\mathrm{diff}}_{\mathrm{mem}}(t)$ is singular (with a delta-function singularity that cancels the delta function at $t = 0$).

Now we can take the Fourier transform of Eq. (G.1) to find that

$$\tilde{h}_{\mathrm{mem}}(f) = \Delta h^{\mathrm{fit}}_{\mathrm{mem}} \left[\frac{1}{2}\delta(f) + \frac{1}{2\pi i f}\right] + \mathcal{F}[h^{\mathrm{diff}}_{\mathrm{mem}}]\,. \tag{G.3}$$

Thus, we would like to evaluate the Fourier transform of $h^{\mathrm{diff}}_{\mathrm{mem}}$ in terms of other smooth functions and elementary functions. Since $h^{\mathrm{diff}}_{\mathrm{mem}}(t)$ has a discontinuity at $t = 0$, we write the Fourier transform out explicitly as

$$\mathcal{F}[h^{\mathrm{diff}}_{\mathrm{mem}}] = \lim_{\epsilon\to 0} \int_{-\infty}^{-\epsilon} h^{\mathrm{diff}}_{\mathrm{mem}} e^{-i2\pi f t} dt + \lim_{\epsilon\to 0} \int_{\epsilon}^{\infty} h^{\mathrm{diff}}_{\mathrm{mem}} e^{-i2\pi f t} dt\,. \tag{G.4}$$

Next we integrate by parts and use the fact that $h^{\mathrm{diff}}_{\mathrm{mem}}(t)$ vanishes as $t \to \pm\infty$ (which removes one set of boundary terms) but has a discontinuity at $t = 0$. This means that there are boundary terms that are nonzero and depend on whether one approaches from $t < 0$ or from $t > 0$. This gives that

$$\begin{aligned}\mathcal{F}[h^{\mathrm{diff}}_{\mathrm{mem}}] = \lim_{\epsilon\to 0} \frac{h^{\mathrm{diff}}_{\mathrm{mem}}(\epsilon) - h^{\mathrm{diff}}_{\mathrm{mem}}(-\epsilon)}{2\pi i f} + \frac{1}{2\pi i f} \lim_{\epsilon\to 0} \int_{-\infty}^{-\epsilon} \dot{h}^{\mathrm{diff}}_{\mathrm{mem}} e^{-i2\pi f t} dt \\ + \frac{1}{2\pi i f} \lim_{\epsilon\to 0} \int_{\epsilon}^{\infty} \dot{h}^{\mathrm{diff}}_{\mathrm{mem}} e^{-i2\pi f t} dt\,.\end{aligned} \tag{G.5}$$

The limit of $h^{\rm diff}_{\rm mem}(\epsilon) - h^{\rm diff}_{\rm mem}(-\epsilon)$ is just $-\Delta h^{\rm fit}_{\rm mem}$. Note that $\dot{h}^{\rm diff}_{\rm mem}$ agrees with $\dot{h}_{\rm mem}$ everywhere except $t = 0$, but $\dot{h}_{\rm mem}$ is finite at $t = 0$. This allows us to replace the two integrals of $\dot{h}^{\rm diff}_{\rm mem}$ over the positive and negative real numbers with a single integral of $\dot{h}_{\rm mem}$ over the entire reals. This latter integral is just the Fourier transform of $\dot{h}_{\rm mem}$, so we determine that

$$\mathcal{F}[h^{\rm diff}_{\rm mem}] = \frac{1}{2\pi i f}\mathcal{F}[\dot{h}_{\rm mem}] - \frac{1}{2\pi i f}\Delta h^{\rm fit}_{\rm mem}\,. \tag{G.6}$$

The memory signal's derivative $\dot{h}_{\rm mem}$ has no discontinuities and approaches zero as $t \to \pm\infty$. Substituting this expression for $\mathcal{F}[h^{\rm diff}_{\rm mem}]$ into Eq. (G.3) gives

$$\tilde{h}_{\rm mem}(f) = \frac{1}{2}\Delta h^{\rm fit}_{\rm mem}\delta(f) + \frac{1}{2\pi i f}\mathcal{F}[\dot{h}_{\rm mem}]\,. \tag{G.7}$$

Thus, aside from a delta function at zero frequency, we can compute the Fourier transform $\tilde{h}_{\rm mem}(f)$ from the Fourier transform for $\dot{h}_{\rm mem}$. Aside from the delta-function term, the result has the same form as the Fourier integral theorem.

Bibliography

[1] B. P. Abbott *et al.*, "Observation of Gravitational Waves from a Binary Black Hole Merger," *Phys. Rev. Lett.*, vol. 116, no. 6, p. 061102, 2016, 1602.03837.

[2] A. Strominger, *Lectures on the Infrared Structure of Gravity and Gauge Theory*, 3 2017, 1703.05448.

[3] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, "Gravitational waves in general relativity. 7. Waves from axisymmetric isolated systems," *Proc. R. Soc. Lond. A*, vol. 269, pp. 21–52, 1962.

[4] R. K. Sachs, "Gravitational waves in general relativity. 8. Waves in asymptotically flat space-times," *Proc. R. Soc. Lond. A*, vol. 270, pp. 103–126, 1962.

[5] R. Sachs, "Asymptotic symmetries in gravitational theory," *Phys. Rev.*, vol. 128, pp. 2851–2864, 1962.

[6] R. Penrose, "Asymptotic properties of fields and space-times," *Phys. Rev. Lett.*, vol. 10, pp. 66–68, 1963.

[7] R. Penrose, “Zero rest mass fields including gravitation: Asymptotic behavior,” *Proc. Roy. Soc. Lond. A*, vol. 284, p. 159, 1965.

[8] G. Barnich and C. Troessaert, “Symmetries of asymptotically flat 4 dimensional spacetimes at null infinity revisited,” *Phys. Rev. Lett.*, vol. 105, p. 111103, 2010, 0909.2617.

[9] G. Barnich and C. Troessaert, “Aspects of the BMS/CFT correspondence,” *J. High Energy Phys.*, vol. 05, p. 062, 2010, 1001.1541.

[10] G. Barnich and C. Troessaert, “BMS charge algebra,” *J. High Energy Phys.*, vol. 12, p. 105, 2011, 1106.0213.

[11] M. Campiglia and A. Laddha, “Asymptotic symmetries and subleading soft graviton theorem,” *Phys. Rev. D*, vol. 90, no. 12, p. 124028, 2014, 1408.2228.

[12] M. Campiglia and A. Laddha, “New symmetries for the Gravitational S-matrix,” *J. High Energy Phys.*, vol. 04, p. 076, 2015, 1502.02318.

[13] E. Noether, “Invariante variationsprobleme,” *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, vol. 1918, pp. 235–257, 1918. [Online]. Available: http://eudml.org/doc/59024

[14] R. M. Wald and A. Zoupas, “A General definition of ’conserved quantities’ in general relativity and other theories of gravity,” *Phys. Rev. D*, vol. 61, p. 084027, 2000, gr-qc/9911095.

[15] L. Blanchet, "Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries," *Living Rev. Relativ.*, vol. 17, p. 2, 2014, 1310.1528.

[16] S. A. Teukolsky, "Perturbations of a rotating black hole. 1. Fundamental equations for gravitational electromagnetic and neutrino field perturbations," *Astrophys. J.*, vol. 185, pp. 635–647, 1973.

[17] G. B. Cook and M. Zalutskiy, "Gravitational perturbations of the Kerr geometry: High-accuracy study," *Phys. Rev. D*, vol. 90, no. 12, p. 124021, 2014, 1410.7698.

[18] E. W. Leaver, "An Analytic representation for the quasi normal modes of Kerr black holes," *Proc. Roy. Soc. Lond. A*, vol. 402, pp. 285–298, 1985.

[19] Y. B. Zel'dovich and A. G. Polnarev, "Radiation of gravitational waves by a cluster of superdense stars," *Sov. Astron.*, vol. 18, p. 17, 1974.

[20] M. S. Turner, "Gravitational Radiation from Supernova Neutrino Bursts," *Nature*, vol. 274, pp. 565–566, 1978.

[21] D. Christodoulou, "Nonlinear nature of gravitation and gravitational wave experiments," *Phys. Rev. Lett.*, vol. 67, pp. 1486–1489, 1991.

[22] A. G. Wiseman and C. M. Will, "Christodoulou's nonlinear gravitational wave memory: Evaluation in the quadrupole approximation," *Phys. Rev. D*, vol. 44, no. 10, pp. 2945–2949, 1991.

[23] L. Blanchet and T. Damour, "Hereditary effects in gravitational radiation," *Phys. Rev. D*, vol. 46, pp. 4304–4319, 1992.

[24] E. E. Flanagan and D. A. Nichols, "Conserved charges of the extended Bondi-Metzner-Sachs algebra," *Phys. Rev. D*, vol. 95, no. 4, p. 044002, 2017, 1510.03386, [Erratum: Phys.Rev.D 108, 069902 (2023)].

[25] D. A. Nichols, "Spin memory effect for compact binaries in the post-Newtonian approximation," *Phys. Rev. D*, vol. 95, no. 8, p. 084048, 2017, 1702.03300.

[26] D. A. Nichols, "Center-of-mass angular momentum and memory effect in asymptotically flat spacetimes," *Phys. Rev. D*, vol. 98, no. 6, p. 064032, 2018, 1807.08767.

[27] V. B. Braginsky and L. P. Grishchuk, "Kinematic Resonance and Memory Effect in Free Mass Gravitational Antennas," *Sov. Phys. JETP*, vol. 62, pp. 427–430, 1985.

[28] V. B. Braginsky and K. S. Thorne, "Gravitational-wave bursts with memory and experimental prospects," *Nature*, vol. 327, pp. 123–125, 1987.

[29] D. Kennefick, "Prospects for detecting the Christodoulou memory of gravitational waves from a coalescing compact binary and using it to measure neutron star radii," *Phys. Rev. D*, vol. 50, pp. 3587–3595, 1994.

[30] M. Favata, "Nonlinear gravitational-wave memory from binary black hole mergers," *Astrophys. J. Lett.*, vol. 696, pp. L159–L162, 2009, 0902.3660.

[31] D. Pollney and C. Reisswig, "Gravitational memory in binary black hole mergers," *Astrophys. J. Lett.*, vol. 732, p. L13, 2011, 1004.4209.

[32] P. D. Lasky, E. Thrane, Y. Levin, J. Blackman, and Y. Chen, "Detecting gravitational-wave memory with LIGO: implications of GW150914," *Phys. Rev. Lett.*, vol. 117, no. 6, p. 061102, 2016, 1605.01415.

[33] M. Hübner, C. Talbot, P. D. Lasky, and E. Thrane, "Measuring gravitational-wave memory in the first LIGO/Virgo gravitational-wave transient catalog," *Phys. Rev. D*, vol. 101, no. 2, p. 023011, 2020, 1911.12496.

[34] O. M. Boersma, D. A. Nichols, and P. Schmidt, "Forecasts for detecting the gravitational-wave memory effect with Advanced LIGO and Virgo," *Phys. Rev. D*, vol. 101, no. 8, p. 083026, 2020, 2002.01821.

[35] A. M. Grant and D. A. Nichols, "Outlook for detecting the gravitational-wave displacement and spin memory effects with current and future gravitational-wave detectors," *Phys. Rev. D*, vol. 107, no. 6, p. 064056, 2023, 2210.16266, [Erratum: Phys.Rev.D 108, 029901 (2023)].

[36] B. P. Abbott *et al.*, "Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA," *Living Rev. Rel.*, vol. 19, p. 1, 2016, 1304.0670.

[37] B. Goncharov, L. Donnay, and J. Harms, "Inferring Fundamental Spacetime Symmetries with Gravitational-Wave Memory: From LISA to the Einstein Telescope," *Phys. Rev. Lett.*, vol. 132, no. 24, p. 241401, 2024, 2310.10718.

[38] K. Islo, J. Simon, S. Burke-Spolaor, and X. Siemens, "Prospects for Memory Detection with Low-Frequency Gravitational Wave Detectors," , 6 2019, 1906.11936.

[39] S. Gasparotto, R. Vicente, D. Blas, A. C. Jenkins, and E. Barausse, "Can gravitational-wave memory help constrain binary black-hole parameters? A LISA case study," *Phys. Rev. D*, vol. 107, no. 12, p. 124033, 2023, 2301.13228.

[40] H. Inchauspé, S. Gasparotto, D. Blas, L. Heisenberg, J. Zosso, and S. Tiwari, "Measuring gravitational wave memory with LISA," , 6 2024, 2406.09228.

[41] G. Agazie *et al.*, "The NANOGrav 12.5 yr Data Set: Search for Gravitational Wave Memory," *Astrophys. J.*, vol. 963, no. 1, p. 61, 2024, 2307.13797.

[42] J. B. Wang *et al.*, "Searching for gravitational wave memory bursts with the Parkes Pulsar Timing Array," *Mon. Not. Roy. Astron. Soc.*, vol. 446, pp. 1657–1671, 2015, 1410.3323.

[43] F. Pretorius, "Evolution of binary black hole spacetimes," *Phys. Rev. Lett.*, vol. 95, p. 121101, 2005, gr-qc/0507014.

[44] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, “Surrogate models for precessing binary black hole simulations with unequal masses,” *Phys. Rev. Research.*, vol. 1, p. 033015, 2019, 1905.09300.

[45] N. E. M. Rifat, S. E. Field, G. Khanna, and V. Varma, “Surrogate model for gravitational wave signals from comparable and large-mass-ratio black hole binaries,” *Phys. Rev. D*, vol. 101, no. 8, p. 081502, 2020, 1910.10473.

[46] J. Yoo *et al.*, “Numerical relativity surrogate model with memory effects and post-Newtonian hybridization,” *Phys. Rev. D*, vol. 108, no. 6, p. 064027, 2023, 2306.03148.

[47] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, L. E. Kidder, and H. P. Pfeiffer, “Surrogate model of hybridized numerical relativity binary black hole waveforms,” *Phys. Rev. D*, vol. 99, no. 6, p. 064045, 2019, 1812.07865.

[48] N. T. Bishop, R. Gomez, L. Lehner, and J. Winicour, “Cauchy-characteristic extraction in numerical relativity,” *Phys. Rev. D*, vol. 54, pp. 6153–6165, 1996, gr-qc/9705033.

[49] N. T. Bishop, R. Gomez, L. Lehner, M. Maharaj, and J. Winicour, “High powered gravitational news,” *Phys. Rev. D*, vol. 56, pp. 6298–6309, 1997, gr-qc/9708065.

[50] K. Mitman, J. Moxon, M. A. Scheel, S. A. Teukolsky, M. Boyle, N. Deppe, L. E. Kidder, and W. Throwe, "Computation of displacement and spin gravitational memory in numerical relativity," *Phys. Rev. D*, vol. 102, no. 10, p. 104007, 2020, 2007.11562.

[51] B. P. Abbott *et al.*, "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*, vol. 9, no. 3, p. 031040, 2019, 1811.12907.

[52] R. Abbott *et al.*, "GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run," *Phys. Rev. X*, vol. 11, p. 021053, 2021, 2010.14527.

[53] V. Varma, D. Gerosa, L. C. Stein, F. Hébert, and H. Zhang, "High-accuracy mass, spin, and recoil predictions of generic black-hole merger remnants," *Phys. Rev. Lett.*, vol. 122, no. 1, p. 011101, 2019, 1809.09125.

[54] M. G. J. van der Burg, "Gravitational Waves in General Relativity. IX. Conserved Quantities," *Proc. R. Soc. Lond. A*, vol. 294, pp. 112–122, Sep. 1966.

[55] E. T. Newman and R. Penrose, "Note on the Bondi-Metzner-Sachs group," *J. Math. Phys.*, vol. 7, pp. 863–870, 1966.

[56] T. Adamo, C. Kozameh, and E. Newman, "Null Geodesic Congruences, Asymptotically Flat Space-Times and Their Physical Interpretation," *Living Rev. Relativ.*, vol. 12, p. 6, 2009, 0906.2155.

[57] R. Owen, "Topics in numerical relativity : the periodic standing-wave approximation, the stability of constraints in free evolution, and the spin of dynamical black holes," Ph.D. dissertation, Caltech, 2007.

[58] G. B. Cook and B. F. Whiting, "Approximate Killing Vectors on S^2," *Phys. Rev. D*, vol. 76, p. 041501, 2007, 0706.0199.

[59] G. Lovelace, R. Owen, H. P. Pfeiffer, and T. Chu, "Binary-black-hole initial data with nearly-extremal spins," *Phys. Rev. D*, vol. 78, p. 084017, 2008, 0805.4192.

[60] C. J. Handmer, B. Szilágyi, and J. Winicour, "Spectral Cauchy Characteristic Extraction of strain, news and gravitational radiation flux," *Classical Quantum Gravity*, vol. 33, no. 22, p. 225007, 2016, 1605.04332.

[61] K. S. Thorne, "Multipole expansions of gravitational radiation," *Rev. Mod. Phys.*, vol. 52, pp. 299–339, Apr 1980. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.52.299

[62] R. P. Geroch and J. Winicour, "Linkages in general relativity," *J. Math. Phys.*, vol. 22, pp. 803–812, 1981.

[63] R. Penrose, "Quasilocal mass and angular momentum in general relativity," *Proc. Roy. Soc. Lond. A*, vol. 381, pp. 53–63, 1982.

[64] T. Dray and M. Streubel, "Angular momentum at null infinity," *Classical Quantum Gravity*, vol. 1, no. 1, pp. 15–26, 1984.

[65] A. Ashtekar and M. Streubel, "Symplectic Geometry of Radiative Modes and Conserved Quantities at Null Infinity," *Proc. Roy. Soc. Lond. A*, vol. 376, pp. 585–607, 1981.

[66] B. Bonga and E. Poisson, "Coulombic contribution to angular momentum flux in general relativity," *Phys. Rev. D*, vol. 99, no. 6, p. 064024, 2019, 1808.01288.

[67] S. Pasterski, A. Strominger, and A. Zhiboedov, "New Gravitational Memories," *JHEP*, vol. 12, p. 053, 2016, 1502.06120.

[68] S. Weinberg, "Infrared photons and gravitons," *Phys. Rev.*, vol. 140, pp. B516–B524, 1965.

[69] G. Compère, R. Oliveri, and A. Seraj, "The Poincaré and BMS flux-balance laws with application to binary systems," *J. High Energy Phys.*, vol. 10, p. 116, 2020, 1912.03164.

[70] G. Compère, A. Fiorucci, and R. Ruzziconi, "The Λ-BMS$_4$ charge algebra," *J. High Energy Phys.*, vol. 10, p. 205, 2020, 2004.10769.

[71] A. Ashtekar and J. Winicour, "Linkages and hamiltonians at null infinity," *J. Math. Phys.*, vol. 23, no. 12, pp. 2410–2417, 1982.

[72] A. Komar, "Covariant conservation laws in general relativity," *Phys. Rev.*, vol. 113, pp. 934–936, 1959.

[73] A. Ashtekar, T. De Lorenzo, and N. Khera, "Compact binary coalescences: The subtle issue of angular momentum," *Phys. Rev. D*, vol. 101, no. 4, p. 044005, 2020, 1910.02907.

[74] G. Compère, A. Fiorucci, and R. Ruzziconi, "Superboost transitions, refraction memory and super-Lorentz charge algebra," *J. High Energy Phys.*, vol. 11, p. 200, 2018, 1810.00377, [Erratum: J. High Energy Phys. 04, 172 (2020)].

[75] G. Compère and J. Long, "Classical static final state of collapse with supertranslation memory," *Classical Quantum Gravity*, vol. 33, no. 19, p. 195001, 2016, 1602.05197.

[76] G. Compère and D. A. Nichols, "Classical and Quantized General-Relativistic Angular Momentum," , 3 2021, 2103.17103.

[77] R. M. Wald, *General Relativity.* Chicago, USA: Chicago Univ. Pr., 1984.

[78] T. Mädler and J. Winicour, "Bondi-Sachs Formalism," *Scholarpedia*, vol. 11, no. 12, p. 33528, 2016, revision #194095.

[79] T. Banks, "A Critique of pure string theory: Heterodox opinions of diverse dimensions," , 6 2003, hep-th/0306074.

[80] É. É. Flanagan, K. Prabhu, and I. Shehzad, "Extensions of the asymptotic symmetry algebra of general relativity," *J. High Energy Phys.*, vol. 01, p. 002, 2020, 1910.04557.

[81] S. W. Hawking, M. J. Perry, and A. Strominger, "Superrotation Charge and Supertranslation Hair on Black Holes," *J. High Energy Phys.*, vol. 05, p. 161, 2017, 1611.09175.

[82] L. Blanchet and G. Faye, "Flux-balance equations for linear momentum and center-of-mass position of self-gravitating post-Newtonian systems," *Classical Quantum Gravity*, vol. 36, no. 8, p. 085003, 2019, 1811.08966.

[83] G. Compère and J. Long, "Vacua of the gravitational field," *J. High Energy Phys.*, vol. 07, p. 137, 2016, 1601.04958.

[84] R. P. Geroch, "Asymptotic Structure of Space-Time," in *Asymptotic Structure of Space-Time*, Esposito, F. P. and Witten, L., Ed. New York: Plenum Press, 1977, pp. 1–105.

[85] M. Boyle *et al.*, "The SXS Collaboration catalog of binary black hole simulations," *Class. Quant. Grav.*, vol. 36, no. 19, p. 195006, 2019, 1904.04831.

[86] K. Mitman *et al.*, “Adding gravitational memory to waveform catalogs using BMS balance laws,” *Phys. Rev. D*, vol. 103, no. 2, p. 024031, 2021, 2011.01309.

[87] M. Favata, “Post-Newtonian corrections to the gravitational-wave memory for quasi-circular, inspiralling compact binaries,” *Phys. Rev. D*, vol. 80, p. 024002, 2009, 0812.0069.

[88] J. A. Gonzalez, U. Sperhake, B. Bruegmann, M. Hannam, and S. Husa, “Total recoil: The Maximum kick from nonspinning black-hole binary inspiral,” *Phys. Rev. Lett.*, vol. 98, p. 091101, 2007, gr-qc/0610154.

[89] R. Abbott *et al.*, “GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run,” *Phys. Rev. X*, vol. 13, no. 4, p. 041039, 2023, 2111.03606.

[90] The LIGO Scientific Collaboration, “GraceDB | The Gravitational-Wave Candidate Event Database,” https://gracedb.ligo.org/ (Accessed: April 15, 2025).

[91] B. P. Abbott *et al.*, “Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo,” *Astrophys. J. Lett.*, vol. 882, no. 2, p. L24, 2019, 1811.12940.

[92] R. Abbott *et al.*, “Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog,” *Astrophys. J. Lett.*, vol. 913, no. 1, p. L7, 2021, 2010.14533.

[93] R. Abbott *et al.*, "Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3," *Phys. Rev. X*, vol. 13, no. 1, p. 011048, 2023, 2111.03634.

[94] B. P. Abbott *et al.*, "Tests of General Relativity with the Binary Black Hole Signals from the LIGO-Virgo Catalog GWTC-1," *Phys. Rev. D*, vol. 100, no. 10, p. 104036, 2019, 1903.04467.

[95] R. Abbott *et al.*, "Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog," *Phys. Rev. D*, vol. 103, no. 12, p. 122002, 2021, 2010.14529.

[96] R. Abbott *et al.*, "Tests of General Relativity with GWTC-3," , 12 2021, 2112.06861.

[97] R. Epstein, "The Generation of Gravitational Radiation by Escaping Supernova Neutrinos," *Astrophys. J.*, vol. 223, pp. 1037–1045, 1978.

[98] K. S. Thorne, "Gravitational-wave bursts with memory: The Christodoulou effect," *Phys. Rev. D*, vol. 45, no. 2, pp. 520–524, 1992.

[99] A. Strominger, "On BMS Invariance of Gravitational Scattering," *JHEP*, vol. 07, p. 152, 2014, 1312.2229.

[100] A. Strominger and A. Zhiboedov, "Gravitational Memory, BMS Supertranslations and Soft Theorems," *J. High Energy Phys.*, vol. 01, p. 086, 2016, 1411.5745.

[101] M. Hübner, P. Lasky, and E. Thrane, "Memory remains undetected: Updates from the second LIGO/Virgo gravitational-wave transient catalog," *Phys. Rev. D*, vol. 104, no. 2, p. 023004, 2021, 2105.02879.

[102] S. Y. Cheung, P. D. Lasky, and E. Thrane, "Does spacetime have memories? Searching for gravitational-wave memory in the third LIGO-Virgo-KAGRA gravitational-wave transient catalogue," *Class. Quant. Grav.*, vol. 41, no. 11, p. 115010, 2024, 2404.11919.

[103] M. Ebersold and S. Tiwari, "Search for nonlinear memory from subsolar mass compact binary mergers," *Phys. Rev. D*, vol. 101, no. 10, p. 104041, 2020, 2005.03306.

[104] A. M. Grant and D. A. Nichols, "Persistent gravitational wave observables: Curve deviation in asymptotically flat spacetimes," *Phys. Rev. D*, vol. 105, no. 2, p. 024056, 2022, 2109.03832, [Erratum: Phys.Rev.D 107, 109902 (2023)].

[105] S. Siddhant, A. M. Grant, and D. A. Nichols, "Higher memory effects and the post-Newtonian calculation of their gravitational-wave signals," *Class. Quant. Grav.*, vol. 41, no. 20, p. 205014, 2024, 2403.13907.

[106] K. Mitman *et al.*, “A review of gravitational memory and BMS frame fixing in numerical relativity,” *Class. Quant. Grav.*, vol. 41, no. 22, p. 223001, 2024, 2405.08868.

[107] C. Talbot, E. Thrane, P. D. Lasky, and F. Lin, “Gravitational-wave memory: waveforms and phenomenology,” *Phys. Rev. D*, vol. 98, no. 6, p. 064031, 2018, 1807.00990.

[108] R. Geroch, “Asymptotic Structure of Space-Time,” in *Symposium on Asymptotic Structure of Space-Time*, 1977.

[109] A. Ashtekar, T. De Lorenzo, and N. Khera, “Compact binary coalescences: Constraints on waveforms,” *Gen. Rel. Grav.*, vol. 52, no. 11, p. 107, 2020, 1906.00913.

[110] Y. Chen *et al.*, “Improved frequency spectra of gravitational waves with memory in a binary-black-hole simulation,” *Phys. Rev. D*, vol. 110, no. 6, p. 064049, 2024, 2405.06197.

[111] J. Valencia, R. Tenorio, M. Rosselló-Sastre, and S. Husa, “Frequency-domain analysis of gravitational-wave memory waveforms,” *Phys. Rev. D*, vol. 110, no. 12, p. 124026, 2024, 2406.16636.

[112] P. Amaro-Seoane *et al.*, “Laser Interferometer Space Antenna,” , Feb. 2017, 1702.00786.

[113] P. A. Seoane *et al.*, "Astrophysics with the Laser Interferometer Space Antenna," *Living Rev. Rel.*, vol. 26, no. 1, p. 2, 2023, 2203.06016.

[114] F. D. Ryan, "Gravitational waves from the inspiral of a compact object into a massive, axisymmetric body with arbitrary multipole moments," *Phys. Rev. D*, vol. 52, pp. 5707–5718, 1995.

[115] L. M. Burko and G. Khanna, "Climbing up the memory staircase: Equatorial zoom-whirl orbits," *Phys. Rev. D*, vol. 102, no. 8, p. 084035, 2020, 2007.12545.

[116] G. Agazie *et al.*, "The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background," *Astrophys. J. Lett.*, vol. 951, no. 1, p. L8, 2023, 2306.16213.

[117] D. J. Reardon *et al.*, "Search for an Isotropic Gravitational-wave Background with the Parkes Pulsar Timing Array," *Astrophys. J. Lett.*, vol. 951, no. 1, p. L6, 2023, 2306.16215.

[118] J. Antoniadis *et al.*, "The second data release from the European Pulsar Timing Array - III. Search for gravitational wave signals," *Astron. Astrophys.*, vol. 678, p. A50, 2023, 2306.16214.

[119] H. Xu *et al.*, "Searching for the Nano-Hertz Stochastic Gravitational Wave Background with the Chinese Pulsar Timing Array Data Release I," *Res. Astron. Astrophys.*, vol. 23, no. 7, p. 075024, 2023, 2306.16216.

[120] R. Fujita, "Gravitational Waves from a Particle in Circular Orbits around a Schwarzschild Black Hole to the 22nd Post-Newtonian Order," *Prog. Theor. Phys.*, vol. 128, pp. 971–992, 2012, 1211.5535.

[121] F. Beyer, B. Daszuta, J. Frauendiener, and B. Whale, "Numerical evolutions of fields on the 2-sphere using a spectral method based on spin-weighted spherical harmonics," *Class. Quant. Grav.*, vol. 31, p. 075019, 2014, 1308.4729.

[122] L. Barack and A. Pound, "Self-force and radiation reaction in general relativity," *Rept. Prog. Phys.*, vol. 82, no. 1, p. 016904, 2019, 1805.10385.

[123] A. Ori and K. S. Thorne, "The Transition from inspiral to plunge for a compact body in a circular equatorial orbit around a massive, spinning black hole," *Phys. Rev. D*, vol. 62, p. 124022, 2000, gr-qc/0003032.

[124] A. Buonanno and T. Damour, "Transition from inspiral to plunge in binary black hole coalescences," *Phys. Rev. D*, vol. 62, p. 064015, 2000, gr-qc/0001013.

[125] G. Compère and L. Küchler, "Self-consistent adiabatic inspiral and transition motion," *Phys. Rev. Lett.*, vol. 126, no. 24, p. 241106, 2021, 2102.12747.

[126] G. Compère and L. Küchler, "Asymptotically matched quasi-circular inspiral and transition-to-plunge in the small mass ratio expansion," *SciPost Phys.*, vol. 13, no. 2, p. 043, 2022, 2112.02114.

[127] L. Küchler, G. Compère, L. Durkan, and A. Pound, "Self-force framework for transition-to-plunge waveforms," *SciPost Phys.*, vol. 17, no. 2, p. 056, 2024, 2405.00170.

[128] S. Mano, H. Suzuki, and E. Takasugi, "Analytic solutions of the Teukolsky equation and their low frequency expansions," *Prog. Theor. Phys.*, vol. 95, pp. 1079–1096, 1996, gr-qc/9603020.

[129] T. Damour, B. R. Iyer, and A. Nagar, "Improved resummation of post-Newtonian multipolar waveforms from circularized compact binaries," *Phys. Rev. D*, vol. 79, p. 064004, 2009, 0811.2069.

[130] R. Fujita and B. R. Iyer, "Spherical harmonic modes of 5.5 post-Newtonian gravitational wave polarisations and associated factorised resummed waveforms for a particle in circular orbit around a Schwarzschild black hole," *Phys. Rev. D*, vol. 82, p. 044051, 2010, 1005.2266.

[131] R. Fujita, "Black Hole Perturbation Club - Data," https://sites.google.com/view/bhpc1996/data (Accessed: July 23, 2024).

[132] R. Fujita, "Gravitational Waves from a Particle in Circular Orbits around a Rotating Black Hole to the 11th Post-Newtonian Order," *PTEP*, vol. 2015, no. 3, p. 033E01, 2015, 1412.5689.

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

[134] A. Elhashash and D. A. Nichols, "Series data for the 22PN nonlinear memory effect in the l=2, m=0 mode." Jul. 2024. [Online]. Available: https://doi.org/10.5281/zenodo.13132039

[135] B. Wardell, A. Pound, N. Warburton, J. Miller, L. Durkan, and A. Le Tiec, "Gravitational Waveforms for Compact Binaries from Second-Order Self-Force Theory," *Phys. Rev. Lett.*, vol. 130, no. 24, p. 241402, 2023, 2112.12265.

[136] A. Albertini, A. Nagar, A. Pound, N. Warburton, B. Wardell, L. Durkan, and J. Miller, "Comparing second-order gravitational self-force, numerical relativity, and effective one body waveforms from inspiralling, quasicircular, and nonspinning black hole binaries," *Phys. Rev. D*, vol. 106, no. 8, p. 084061, 2022, 2208.01049.

[137] A. Albertini, A. Nagar, A. Pound, N. Warburton, B. Wardell, L. Durkan, and J. Miller, "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*, vol. 106, no. 8, p. 084062, 2022, 2208.02055.

[138] M. van de Meent, A. Buonanno, D. P. Mihaylov, S. Ossokine, L. Pompili, N. Warburton, A. Pound, B. Wardell, L. Durkan, and J. Miller, "Enhancing the SEOBNRv5 effective-one-body waveform model with second-order gravitational self-force fluxes," *Phys. Rev. D*, vol. 108, no. 12, p. 124038, 2023, 2303.18026.

[139] T. Islam, S. E. Field, G. Khanna, and N. Warburton, "Survey of gravitational wave memory in intermediate mass ratio binaries," *Phys. Rev. D*, vol. 108, no. 2, p. 024046, 2023, 2109.00754.

[140] K. Cunningham, C. Kavanagh, A. Pound, D. Trestini, N. Warburton, and J. Neef, "Gravitational memory: new results from post-Newtonian and self-force theory," , 10 2024, 2410.23950.

[141] R. Abbott *et al.*, "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run," , 11 2021, 2111.03606.

[142] R. Abbott *et al.*, "GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object," *Astrophys. J. Lett.*, vol. 896, no. 2, p. L44, 2020, 2006.12611.

[143] A. Elhashash and D. A. Nichols, "Waveform models for the gravitational-wave memory effect: Extreme mass-ratio limit and final memory offset," *Phys. Rev. D*, vol. 111, no. 4, p. 044052, 2025, 2407.19017.

[144] M. Giesler, M. Isi, M. A. Scheel, and S. Teukolsky, "Black Hole Ringdown: The Importance of Overtones," *Phys. Rev. X*, vol. 9, no. 4, p. 041060, 2019, 1903.08284.

[145] G. B. Cook, "Aspects of multimode Kerr ringdown fitting," *Phys. Rev. D*, vol. 102, no. 2, p. 024027, 2020, 2004.08347.

[146] L. Magaña Zertuche *et al.*, "High precision ringdown modeling: Multimode fits and BMS frames," *Phys. Rev. D*, vol. 105, no. 10, p. 104015, 2022, 2110.15922.

[147] L. C. Stein, "qnm: A Python package for calculating Kerr quasinormal modes, separation constants, and spherical-spheroidal mixing coefficients," *J. Open Source Softw.*, vol. 4, no. 42, p. 1683, 2019, 1908.10377.

[148] "*NIST Digital Library of Mathematical Functions*," https://dlmf.nist.gov/, Release 1.2.3 of 2024-12-15, f. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds. [Online]. Available: https://dlmf.nist.gov/

[149] The mpmath development team, *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 1.3.0)*, 2023, https://mpmath.org/.

[150] LIGO Collaboration, "Noise curves used for simulations in the update of the Observing Scenarios Paper," https://dcc.ligo.org/ligo-t2000012/public (Last Accessed: April 24, 2025).