Surrogate Models of Gravitational Waves from Numerical Relativity Simulations of Binary Black Hole Mergers

Thesis by

Jonathan Blackman

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2017
Defended May 11, 2017

To my family, for all their love and support

Acknowledgments

I must first thank my parents for their unwavering support, and for encouraging me to pursue my passions regardless of whether they might be music, science, or driving a garbage truck. I thank my brother Anthony for helping me develop my intellectual curiosity, and my grandfather John for sparking my interest in physics.

I thank my advisor Christian Ott and my numerous unofficial mentors, Chad Galley, Scott Field, Mark Scheel, Béla Szilágyi, and Manuel Tiglio, for all their advice and guidance. In addition to Christian Ott and Mark Scheel, I thank Yanbei Chen, Alan Weinstein, and Mark Wise for serving on my thesis committee and for their advice. I have benefited from conversations with and the advice of countless others during my graduate studies. Some of them are Kevin Barkett, Alessandra Buonanno, Matt Giesler, Sarah Gossan, Daniel Hemberger, Geoffrey Lovelace, Maria Okounkova, Richard O'Shaughnessy, Harald Pfeiffer, Michael Pürrer, Patricia Schmidt, Rory Smith, Saul Teukolsky, and Vijay Varma. I also thank all my fellow TAPIR and physics graduate students for many enlightening conversations. I thank Kip Thorne for inspiration, and for hosting wonderful parties. I thank JoAnn Boyd for all her help. I would also like to thank Richard Anstee for introducing me to the world of academic research, and for very helpful advice all the way from undergrad linear algebra to the start of my graduate studies at Caltech.

Beyond academics, countless people contributed to a wonderfully stimulating and enjoyable experience at Caltech - my fellow TAPIR and physics graduate students, the Öcsi Bácsi lunch crew, the Ath+tabletop crew, Caltech Aftermath during my first year, the Caltech-Occidental Orchestra members and conductor Allen Gross, the Caltech Band members and conductor Bill Bing, the Boldly Go reading cast, the nameless trivia team, the BaFF team, camping buddies, and of course the GodfAther adventure team.

Finally, to the TAPIR coffee machine and the Cahill air conditioning units, thanks for everything. I couldn't have done it without you.

Abstract

On September 14 2015, the two Laser Interferometer Gravitational Observatory (LIGO) detectors observed gravitational waves from the coalescence of two black holes [250]. Known as GW150914, this first direct measurement of gravitational waves was the start of the new era of gravitational wave astronomy. The observed signals agree with the waveform predicted by general relativity (GR) for the merger of a pair of black holes with masses $36M_{\odot}$ and $29M_{\odot}$ [20, 252, 263]. As there is no analytic solution to the two-body problem in GR, the predicted waveforms during the merger must be obtained through expensive numerical simulations on supercomputers. The field of numerical relativity (NR) has made significant progress since the first successful simulation of a black hole merger in 2005 [19], and NR codes are robust and accurate enough to simulate binary black hole (BBH) mergers for most black hole parameters that might result in gravitational waves detected in LIGO.

In order to determine the black hole parameters leading to the best match with the observed signal, as well as determining the parameter uncertainties, the signal must be compared to predicted waveforms from many different BBH systems. Due to the high dimensionality of the parameter space, current parameter estimation methods [20] may require the waveforms from millions of physically different systems. A single BBH simulation can take days to months running on dozens of cores, making it prohibitively expensive to do millions of NR simulations. The analysis of GW150914 therefore resorted to using approximate waveform models [66, 70, 152, 265] which are calibrated to some NR simulations. Waveform modeling errors can lead to systematic biases in the measured black hole and extrinsic parameters, as well as false positives or negatives when testing GR. For low signal-to-noise ratio (SNR) detections and in regions of the parameter space where the models are well calibrated, this systematic error can be small compared to the statistical error due to detector

noise [268]. But as gravitational wave detectors improve, SNRs will increase and the accuracy requirements of waveform models will become increasingly stringent.

NR surrogate models [14, 15, 76, 269] make no underlying assumptions about the structure of the waveform, and directly interpolate or fit NR data across the parameter space. With sufficiently many NR simulations, NR surrogate models become as accurate as the underlying NR waveforms. Additionally, improving the surrogate model by including additional NR waveforms is a simple task, requiring very little human effort. The $7d$ intrinsic parameter space of the black hole mass ratio and two black hole spin vectors presents a tough challenge; naive waveform interpolation methods would require $O(10^7)$ NR simulations. We overcome this challenge by interpolating not the waveform itself, but various physically-motivated components which depend very weakly on most parameter dimensions, allowing accurate models even with a sparse covering of the parameter space. This method was built up in stages. We first built a surrogate model for the $1d$ parameter subspace of non-spinning black holes [269]. Next we tackled a $4d$ parameter subspace including precession [14]. Finally, we built a NR surrogate model including all 7 parameter dimensions [15] using fewer than 1000 NR simulations.

Chapter 1 presents a general introduction to gravitational waves, their sources, and detectors. Chapter 2 gives a brief technical overview of topics relevant to the work in this thesis. Chapters 3-6 contain papers previously published in academic journals, with introductory sections including the details of my own contributions appearing before the chapter author list and abstract. Chapter 3 demonstrates how reduced basis methods can be used to generate accurate representations of gravitational waves in the full $7d$ parameter space using a very sparse catalog of gravitational waveforms. Chapter 4 presents a 176 orbit NR simulation, which begins early enough in the inspiral that it agrees with analytic approximations. The resulting waveform is used to test the accuracy of approximate waveform models. Chapter 5 presents the NR surrogate model for the $1d$ parameter subspace of non-spinning black holes. Chapter 6 presents the NR surrogate model for the $4d$ parameter subspace including precession. Chapter 7 contains a paper on the full $7d$ parameter space NR surrogate model that has been submitted for publication. Chapter 8 details my contributions to publications in which I did not play a major role. Appendix A provides a thorough technical documentation of the method of building fast tensor spline surrogates used in chapter 6.

Published Content and Contributions

1. J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, C. D. Ott, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and B. Szilágyi. A Numerical Relativity Waveform Surrogate Model for Generically Precessing Binary Black Hole Mergers (2017). Submitted to Physical Review D. arXiv: `1705.07089 [gr-qc]`.
   J.B. helped in selecting parameters for the SpEC simulations, performed some of the SpEC simulations, built the surrogate, performed the error studies, and was the principal author of the manuscript.
2. J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith. A surrogate model of gravitational waveforms from numerical relativity simulations of precessing binary black hole mergers. *Phys. Rev. D* 95 (10 2017), p. 104023. DOI: `10.1103/PhysRevD.95.104023`.
   J.B. participated in the conception and planning of the project, determined the greedy parameters, performed some of the SpEC simulations, determined and implemented the new surrogate methods, built the surrogate models, performed the error analysis, and was the principal author of the manuscript.
3. J. Blackman, S. E. Field, C. R. Galley, B. Szilágyi, M. A. Scheel, M. Tiglio, and D. A. Hemberger. Fast and Accurate Prediction of Numerical Relativity Waveforms from Binary Black Hole Coalescences Using Surrogate Models. *Phys. Rev. Lett.* 115 (12) (2015), p. 121102. DOI: `10.1103/PhysRevLett.115.121102`.
   J.B. performed most of the SpEC simulations, contributed to the surrogate methods and implementation, performed the error analysis, and contributed to the writing of the manuscript.
4. J. Blackman, B. Szilagyi, C. R. Galley, and M. Tiglio. Sparse representations of gravitational waves from precessing compact binaries. *Phys. Rev. Lett.* 113 (2 2014), p. 021101. DOI: `10.1103/PhysRevLett.113.021101`.
   J.B. contributed to developing the method of using the coprecessing frame, implemented and performed the studies, and contributed to the writing of the manuscript.

5. B. Szilágyi, J. Blackman, A. Buonanno, A. Taracchini, H. P. Pfeiffer, M. A. Scheel, T. Chu, L. E. Kidder, and Y. Pan. Approaching the post-newtonian regime with numerical relativity: a compact-object binary simulation spanning 350 gravitational-wave cycles. *Phys. Rev. Lett.* 115 (3 2015), p. 031102. DOI: `10.1103/PhysRevLett.115.031102`.
   J.B. performed the long SpEC simulation and follow-up simulations with reduced outer boundary radii, computed radiated energies and momenta and compared them to analytic results to investigate the center of mass acceleration, computed the unfaithfulness against analytical models, and assisted in editing the manuscript.

6. J. M. Fedrow, C. D. Ott, U. Sperhake, J. Blackman, R. Haas, C. Reisswig, and A. De Felice. Gravitational Waves from Binary Black Hole Mergers Inside of Stars (2017). arXiv: `1704.07383 [astro-ph.HE]`.
   J.B. helped compute the mismatches between vacuum and gas waveforms, and used numerical relativity surrogate models to minimize the mismatches over vacuum waveforms from all possible binary black hole mergers.

7. R. O'Shaughnessy, J. Blackman, and S. E. Field. An architecture for efficient multimodal gravitational wave parameter estimation with linear surrogate models (2017). arXiv: `1701.01137 [gr-qc]`.
   J.B. contributed to the gwsurrogate code which evaluates surrogate models in a way that can be interfaced with the implementation of the ILE method, and assisted in editing the manuscript.

8. H. Yang, K. Yagi, J. Blackman, L. Lehner, V. Paschalidis, F. Pretorius, and N. Yunes. Black hole spectroscopy with coherent mode stacking. *Phys. Rev. Lett.* 118 (16 2017), p. 161101. DOI: `10.1103/PhysRevLett.118.161101`.
   J.B. produced Fig. 2 including computing the curves shown, and assisted in editing the manuscript.

9. P. D. Lasky, E. Thrane, Y. Levin, J. Blackman, and Y. Chen. Detecting gravitational-wave memory with ligo: implications of gw150914. *Phys. Rev. Lett.* 117 (6 2016), p. 061102. DOI: `10.1103/PhysRevLett.117.061102`.
   J.B. helped find and resolve the waveform degeneracy leading to the difficulty in determining the sign of the memory signal, performed Monte-Carlo studies on the effects of parameter uncertainty and ensuring the memory sign is recovered 95% of the time, and assisted in editing the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1 Introduction

1.1 Overview

Einstein's general theory of relativity (GR) predicts the existence of gravitational waves (GWs), which are small fluctuations in the curvature of spacetime propagating at the speed of light [21]. As electromagnetic waves are sourced by accelerating charges, GWs are sourced by changing mass-energy quadrupole moments and higher order multipoles, such as in a binary neutron star or black hole (BH) system. In a binary black hole (BBH) system, the two BHs orbit their center of mass while slowly radiating energy and angular momentum in the form of GWs. As the system radiates energy, the BH separation decreases and the orbital frequency increases, which leads to GWs with higher amplitudes being emitted. This is known as the *inspiral* phase. Eventually, the two BHs get close enough to merge into a single BH that is initially highly deformed (*merger* phase). Once the deformities are small, it can be treated as a deformed Kerr BH that radiates away its deformations as quasi-normal modes (QNMs) [22] during the *ringdown* phase.

Before September 2015, we had only indirect evidence for the existence of GWs from the gradual increase in the orbital frequencies of binary pulsars [23–25]. Over the last decades, a network of ground-based interferometric GW detectors have been built [26–30]. No confident detections of GWs were made with the initial generation, and an advanced generation of ground-based detectors with increased sensitivity is being built [31–35]. The two Laser Interferometer Gravitational-Wave Observatory (LIGO) detectors [1, 36] were the first to be upgraded to the advanced generation, and began recording data in September 2015. On September 14 2015, the LIGO detectors made the first direct observation of GWs from the merger of a BBH system [250] known as GW150914. A second confident GW detection [251] called GW151226 was made on December 26 2015, which was also from a BBH merger.

As more advanced generation GW detectors come online, and as further improvements are made bringing the detectors closer to their design sensitivities, the sensitivity of the GW detector network will increase. There should be many more detections of GWs from BBH mergers and other systems in the near future. These detected GW signals can be compared to the predictions of GR, offering a test of GR in the strong field, highly dynamical limit. Other theories of gravity that agree with GR in the weak field limit and agree with experiments to date can produce different GW signals from BBHs and other strongly gravitating systems [37–40]. Many tests of GR were performed with GW150914 [252], and no inconsistencies were found.

In addition to tests of GR, GW detectors open up a new form of astronomy; we may determine certain GW source parameters through the structure of the detected GWs. For example, in BBH systems, the orbital frequency at merger decreases as the masses of the BHs increase, so by measuring the frequencies of the detected GW signal, we can infer information about the masses of the progenitor BBH system. This is known as GW parameter estimation [41–44]. For example, the masses of the two black holes of GW150914 were found in this way to be about $36M_\odot$ and $29M_\odot$ [20, 263], where $M_\odot$ is the mass of our sun. As we detect more and more GW signals, we can learn about the populations of GW sources. The population of BBHs is fairly uncertain [45–49], but using GW150914 [2] and all current gravitational wave detections [3] the BBH merger rate has been constrained. Compact binaries consisting of binary neutron stars (BNS) or one black hole and one neutron star (BHNS) are also strong sources of GWs. No detections of GWs from these systems have been made so far, placing upper limits on the merger rate of these systems [4].

In order to accurately estimate source parameters from GWs and test GR, we must have accurate knowledge of the GW signal predicted by GR given a set of source parameters. There is currently no analytic solution for BBH systems, so we must resort to approximations or numerical methods. In 2005, after decades of work, the field of numerical relativity (NR) enabled us to integrate Einstein's equations for BBH systems on computers [19, 50–53]. While we can now obtain accurate NR gravitational waveforms for most BH parameters [54–58], the computations are quite expensive, requiring months of supercomputer time. Since parameter estimation typically uses matched-filtering [59, 60] where the data stream containing both a signal and noise is compared with many GW templates to maximize the match, it may require determining the GW signal for millions of systems. NR is many orders of magnitude too slow and expensive for this purpose. There are several fast approximate waveform models:

- Post-Newtonian (PN) expansions [61, 62] add GR corrections to Newtonian orbital dynamics to obtain waveforms that are accurate for low orbital velocities.
- Phenomenological (“Phenom") models [63–66] assume some model for the structure of the gravitational waveforms and calibrate numerical coefficients in the model to a handful of NR waveforms.
- The effective one-body (EOB) model [67–70, 152] sets up a Hamiltonian for a point particle in a background metric such that the GW solution agrees with PN at low velocities (large separations) but also gives phenomenologically reasonable solutions for high velocities (small separations) and when the BHs merge. Additional coefficients can be added to the model and calibrated to NR simulations to improve the model accuracy.

For GW detections from sources with low to moderate spins and mass ratios, and with small to moderate signal-to-noise ratios (SNRs), Phenom and EOB models are sufficiently accurate to give small or undetectable parameter biases for advanced-generation ground-based GW detectors [16, 71, 72, 268]. As detector sensitivities improve, however, we will encounter signals with large SNRs where model errors can become noticeable. If the waveform from a model evaluated at the BBH source parameters is observationally distinct from the observed signal, there are two possibilities. The first is that the waveform model can produce a waveform that *does* agree with the observed signal using different source parameters. This would lead to a biased measurement of the source parameters. The other case is that the waveform model differs from the observed signal for all source parameters, which, if we neglect the model error, could be falsely interpreted as a violation of GR. We therefore need waveform models that are accurate enough to produce waveforms that are indistinguishable from GR for all expected SNRs.

Improving Phenom and EOB models can be very difficult and would require more and more calibration to NR waveforms as waveform accuracy requirements become more strict. A better method is to use a set of NR waveforms and build a model that can quickly interpolate them over the BBH parameter space using reduced-order modeling (ROM) techniques [73–76]. With sufficient NR waveforms, the model can predict waveforms for new parameters with roughly the same accuracy as the input waveforms, making the model a “surrogate model” of the code that produced the NR waveforms. Improving this model is straightforward: if the interpolation is not sufficiently accurate, perform additional NR simulations. If any of the input waveforms are below the desired output accuracy, redo them with greater resolution. We have created several such surrogate models, which are presented in this thesis.

1.2 Sources of Gravitational Waves

To leading order in the velocities of matter compared to the speed of light, GWs are sourced by the second time derivative of a mass quadrupole moment. In particular, spherically symmetric systems do not emit GWs in GR, no matter how much they oscillate in the radial direction. Even spinning systems such as a spinning star, BH, or even a galaxy will not emit any GWs if they are axially symmetric and stationary. The strongest sources will have some large mass asymmetry which changes rapidly. Binary systems, where two objects orbit a common center of mass, are one of the most well-understood sources of GWs [21, 77].

Since the emitted waves are strongest at and just before merger, a binary consisting of compact objects such as neutron stars or BHs is a good candidate for emitting detectable GWs. Since neutron stars emit beams of radiation which may be detected on Earth as a pulsar (or a binary pulsar when the neutron star is in a binary system with a companion), we have reasonably good estimates of populations of neutron stars [78]. We have even observed a double pulsar [79], where both neutron stars in a binary system are pulsars. Neutron star masses lie between about $1M_\odot$ and $2M_\odot$ [80]. They also spin quite rapidly compared to other stars, with periods as small as a millisecond [81, 82], although neutron stars in observed binaries have smaller spins. The small range of masses of neutron stars means binary neutron star systems will consistently merge at orbital frequencies on the order of $1\,\mathrm{kHz}$. BHs, on the other hand, are much harder to detect through electromagnetic astronomy, as we must resort to indirect signals such as observing x-rays emitted from matter falling into the BH [45] or gravitational lensing [83]. Their masses fall into two categories: stellar-mass BHs with masses between $\sim 5M_\odot$ and $\sim 40M_\odot$ [47], and supermassive black holes (SMBHs) which lie at the centers of galaxies, which have masses in the $10^6 M_\odot$ to $10^{10} M_\odot$ range [84]. While we have observed SMBHs in binaries [85], there has been no observation of a BBH system consisting of stellar-mass BHs other than from the GWs they produce. Intermediate mass black holes (IMBHs) may exist, with masses between these two ranges, although there is so far no definitive observational evidence [86]. Evidence for IMBHs and determining their populations would have a great impact on our understanding of galaxy formation and SMBH formation. The lack of IMBH merger signals in the first observing run of Advanced LIGO has placed upper limits on the IMBH merger rate [5].

Core-collapse supernovae are another potential source of GWs, especially when the core is rapidly rotating. These occur when the iron core of a large star with mass $\gtrsim 8M_{\odot}$ becomes unstable and collapses to form a proto-neutron star. At nuclear densities, the nuclear equation of state is stiff, causing the core to "bounce," which sends out a shock wave towards the surface of the star. The GWs emitted from such an event occurring in the Milky Way or a nearby galaxy may be detectable on Earth with the advanced-generation of GW detectors [87].

1.3 Detecting Gravitational Waves

Laser interferometry is used in LIGO and other ground-based detectors. The method makes use of small differences in the laser beam path length when a GW passes through the detector. A simplified description consists of a laser beam that is sent through a beam splitter and travels along one of two different directions. The beams are then reflected back with mirrors, and combined into a single beam. The amplitude of the final beam depends on the phase difference of the two beams. A GW passing through the detector will alter the two path lengths in different ways, altering the final phases of the beams and changing the interference pattern of the signal. While the GW strain may be minuscule, the beam paths may be made very long such that non-negligible phase differences are accumulated.

Ground-based interferometric GW detectors have already been built [26–30], and an advanced generation of detectors with greater sensitivity is currently being built, upgraded, and in the case of advanced LIGO (aLIGO), is already taking data [31–35]. While the initial generation of detectors made no detections, their sensitivities were such that a reasonable chance of making a detection would require extremely optimistic rates of compact binary coalescences. Ground-based detectors are sensitive to GW frequencies above $\sim 10\,\mathrm{Hz}$, while lower frequencies are inaccessible due to the vibrations of the Earth. They are typically sensitive up to frequencies of a few kHz. This sensitivity band is ideal for coalescences of neutron stars and/or stellar-mass BHs with total masses between $\sim 3M_{\odot}$ and $\sim 300M_{\odot}$, which merge at orbital frequencies ranging from $\sim 30\,\mathrm{Hz}$ to $\sim 3\,\mathrm{kHz}$. They are also sensitive to GWs emitted from core-collapse supernovae, which have frequencies around $100\,\mathrm{Hz} - 2\,\mathrm{kHz}$. There are also proposals for third generation ground-based gravitational wave detectors [88, 89]. The additional sensitivity at low frequencies would improve the detection rate for compact binary merger events, and greater sensitivities at high frequencies would allow the investigation of the merger phase of binary neutron stars. This could provide valuable information about the neutron star equation of state [90]. Improved sensitivities at high frequencies would also improve detection and parameter estimation for GWs from core-collapse supernovae.

Space-based interferometric GW detectors have also been proposed [91, 92]. A detector in space would not be limited to frequencies above 10 Hz, and could be sensitive from a fraction of a mHz up to ~ 0.1Hz. The beam lengths can also be made very long, increasing the sensitivity and allowing smaller laser powers. This sensitivity band corresponds to the mergers of smaller SMBHs with masses around $10^6\, M_\odot$, as well as the merger of a SMBH with stellar mass compact objects, known as extreme mass-ratio inspirals (EMRIs). It also includes portions of the early inspiral from galactic white dwarf binaries. The success of the LISA pathfinder mission [6] makes a space-based GW detector a likely possibility in the coming decades.

The second method which may detect GWs in the near future uses pulsar timing arrays (PTAs) [93]. By monitoring many stable millisecond pulsar signals and looking for correlations in their timing residuals, we may detect the presence of GWs emitted from the merger of two galaxies containing SMBHs. PTAs are most sensitive to periods on the order of a few years [94], meaning we require many years or decades of data to build up a strong signal.

1.4 Tests of General Relativity

Gravitational wave observations give us the opportunity to test GR in the strong-field, highly dynamical regime. Solar system, binary pulsar, and cosmological tests of GR have placed tight constraints on the weak-field, quasi-stationary regime [95, 96], but there are alternative theories of gravity that pass these tests while giving different predictions during, for example, the merger of two black holes. Even binary pulsar constraints only test velocities up to $v/c = \mathcal{O}(10^{-3})$ [24, 25] where c is the speed of light, whereas the merger of two compact objects has $v/c = \mathcal{O}(1)$. For overviews of alternative theories of gravity, experimental constraints, and proposed tests using gravitational wave observations, see Refs. [95–98].

Scalar tensor theories [99, 100] include a scalar field that couples to the metric. While a scalar tensor no-hair theorem [101] means black holes radiate away any scalar "hair", neutron stars can have non-trivial scalar profiles. These systems emit dipolar gravitational radiation which dominates the GR quadrupolar emission at low velocities [102].

Some alternatives to GR have a massive graviton [103]. This would lead to a dispersion of the GWs, with high-frequency GWs traveling faster than low-frequency GWs [104]. Graviton dispersion can also be obtained through Lorentz violations, which can arise in theories of quantum gravity [105]. The dispersion can be constrained very well by LIGO observations of GWs from BBH mergers, since the ditance between the source and Earth is very large. If the mass of the graviton and the distance are large enough, the merger signal could even arrive at the Earth before the inspiral signal. GW150914 was used to constrain the compton wavelength $\lambda_g = h/(m_g c)$ better than solar system tests [252]. Here, h is Planck's constant and m_g is the mass of the graviton.

Dynamical Chern-Simons modified gravity [106] and Einstein-Dilaton-Gauss-Bonnet (EDGB) gravity [107] arise from string theory, and their actions include terms that are quadratic in the curvature. These additional terms become small as the masses of the objects become large, so the merger of stellar mass BBH, BNS, and BHNS are better candidates than EMRIs or the merger of SMBHs. EDGB has already been significantly constrained [108–110], while dynamical Chern-Simons is only loosely constrained [111].

Testing particular theories of gravity using gravitational wave observations can be difficult, as precise theoretical predictions of gravitational waves from merging compact objects are not generally available. For some theories such as dynamical Chern-Simons, it is not even clear if it can be formulated as a stable initial value problem for numerical simulations, although perturbative methods can be used to modify GR simulations. Additionally, it could be that neither GR or any of the known alternatives are correct and that a complete description of gravity has yet to be formulated. It is therefore desireable to use theory-independent tests of GR that don't rely on the predictions of one particular theory. One such test is to measure potential modifications to PN coefficients [112, 113]. In the analysis of GW150914 [252], the post-Newtonian coefficients were constrained and were found to be consistent with GR. The inspiral and ringdown signals were also found to be mutually consistent with GR, and the ringdown is consistent with the quasi-normal modes predicted by GR.

2 Technical Background

2.1 Gravitational Waves

For a more comprehensive description of GWs, see e.g. [21].

We will use units where $G = c = 1$, such that masses, lengths and times all have the same dimensionality. If we look for perturbations to the flat-space metric which solve the Einstein field equations in vacuum to linear order, i.e. take

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{2.1}$$

and enforce $G_{\mu\nu} = \mathcal{O}(|h|^2)$, we obtain linear wave solutions for $h_{\mu\nu}$. For a point source at the origin, after eliminating gauge degrees of freedom and choosing spherical coordinates such that

$$\eta_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2\sin^2\theta \end{bmatrix}, \tag{2.2}$$

we can write the general solution as

$$h_{\mu\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & h_+(t,r,\theta,\phi) & h_\times(t,r,\theta,\phi) \\ 0 & 0 & h_\times(t,r,\theta,\phi) & -h_+(t,r,\theta,\phi) \end{bmatrix}, \tag{2.3}$$

where

$$h_+(t,r,\theta,\phi) = \frac{1}{r}A_+(t-r,\theta,\phi) \tag{2.4}$$

and

$$h_\times(t,r,\theta,\phi) = \frac{1}{r}A_\times(t-r,\theta,\phi). \tag{2.5}$$

This linear approximation is good when $|h| \ll 1$, which will always be true in the limit $r \to \infty$. Since we are interested in detecting GWs from sources very far away ($r \gg GM/c^2$ for a source of mass M), non-linear corrections to the GWs are negligible. We can compute the linear limit of the GWs by setting $t = \tau + r$ and taking $r \to \infty$, which corresponds to computing the GWs at future-null infinity, denoted by $\mathcal{I}^+$. Of course, since h vanishes at $\mathcal{I}^+$, we instead compute rh. The two polarizations of the GW, h_+ and $h_\times$, have the effect of gently squeezing one direction and stretching another, both of which are orthogonal to the direction of propagation of the GW.

Suppose we have a GW source consisting of some matter moving slowly compared to the speed of light. For an observer far away from the source in the $\hat{z}$ direction at some distance r, it can be shown [21, 62] that (to leading order in the velocities and $1/r$)

$$h_+(t) = \frac{1}{r}\left(\ddot{I}_{xx}(t-r) - \ddot{I}_{yy}(t-r)\right) , \tag{2.6}$$

$$h_\times(t) = \frac{2}{r}\ddot{I}_{xy}(t-r) , \tag{2.7}$$

where $I_{ij}(t)$ is the mass quadrupole moment of the source

$$I_{ij}(t) = \int d^3x' \rho(t, \vec{x}') x'_i x'_j , \tag{2.8}$$

and ρ is the density of the source. This is known as the quadrupole formula.

2.2 Binary Black Hole Systems

This work is primarily concerned with GW signals from BBH systems. However, when one (or both) of the BHs is (are) replaced with a neutron star, the orbital dynamics and GWs emitted are nearly unchanged during the inspiral phase. This allows BBH waveforms to be used for the detection of GWs emitted from binary neutron star inspirals, and for the estimation of the source parameters. During the late inspiral, however, the neutron star(s) undergo tidal deformations which modify the dynamics and GW emission [114]. The merger phase is then completely different from that of a BBH system [115]. While two BHs merge smoothly into a single BH, two neutron stars will form a single hypermassive neutron star which oscillates at a characteristic frequency determined by the mass and neutron star equation of state. The hypermassive neutron star may eventually collapse into a BH. Of course, if we wish to gain valuable information about the neutron star equation of state through GW signals, we cannot use BBH waveforms to do the analysis.

Consider an isolated BBH system. When the two BHs are very far apart from each other, they can be approximated as Kerr BHs which have a mass m and spin vector $\vec{S}$. The dimensionless spin, $\vec{\chi} = \vec{S}/m^2$, is restricted to values $|\vec{\chi}| \leq 1$, and there is observational evidence for BHs with nearly extremal spins [116–118]. While BHs may also be charged, astrophysically we expect their charge to be negligible. Large electromagnetic forces would attract oppositely-charged particles, and spontaneous pair-production near the horizon can discharge black holes [119]. We thus have 4 parameters for each BH in addition to the orbital period, eccentricity, and orientation of the periastron in the orbital plane. While the separation is still large, the orbital motion is well approximated by Newtonian gravity. Calculating post-Newtonian corrections to the orbital dynamics shows that the eccentricity is radiated away as GWs faster than the orbital separation [120]. This process is sufficiently fast that all astrophysical stellar mass BBH orbits will circularize before the BHs merge, except in globular clusters [121] where three body interactions can sometimes create eccentric binaries in a state too close to merger to circularize.

Once we restrict to non-eccentric orbits, we may choose a fixed radius r_0 and investigate the BH parameters when their separation is r_0. The unit normal to the instantaneous orbital plane defines one direction, and the separation vector defines another. Here, we are still assuming the BHs are far enough away that a coordinate separation “vector" makes sense and isn’t obscured by the curvature of spacetime. This fixes a coordinate system in which we may measure the spins of the BHs, so for any non-eccentric BBH coalescence, we have 8 parameters. Since GR in vacuum is scale-invariant, if we know the GW strain $rh(t,\theta,\phi)$ at $\mathcal{I}^+$ from a BBH system with some parameters $(m_A, m_B, \vec{\chi}_A, \vec{\chi}_B)$, the corresponding GW strain for a system with scaled parameters $(\alpha m_A, \alpha m_B, \vec{\chi}_A, \vec{\chi}_B)$ is $\alpha rh(\alpha t,\theta,\phi)$. We may therefore restrict to systems with a fixed total mass $M = m_A + m_B$ of 1 and determine the dimensionless strain $rh(t/M,\theta,\phi)/M$. This leaves us with seven parameters: the mass ratio $q = m_A/m_B$ and the two spin vectors. Including eccentricity would require two additional parameters: the eccentricity and orientation of the periastron in the orbital plane, both measured at a particular value of the orbital angular momentum.

If the BH spins are aligned with the orbital angular momentum $\vec{L}$, they will remain as such and the orbital plane will be fixed. It is then useful to decompose the GWs emitted in all directions into spin-weighted spherical harmonics (SWSHs) of spin weight −2. For a full description of SWSHs and why they are needed for describing GWs, see [122]. If the BH spins are misaligned with the orbital angular momentum, the spins and orbital plane will undergo precession. To lowest PN order, they precess about the total angular momentum $\vec{J} = \vec{L} + \vec{S}_1 + \vec{S}_2$, which is conserved on the precession timescale.

2.3 Spectral Einstein Code

The Spectral Einstein Code (SpEC) [123] is a numerical relativity code that can perform BBH simulations. It presently can handle mass ratios up to 10, dimensionless spin magnitudes of almost 1 [55], and is typically used for simulations lasting dozens of orbits, although it is capable of performing hundreds of orbits given enough computing time [16]. The singularities inside the BHs are excised from the computational domain. The excised regions are chosen to lie within the BH apparent horizons, which are gauge-dependent null surfaces guaranteed to lie inside the event horizons. This ensures the outside physics are not affected. The remaining domain, which extends out to some outer boundary, is decomposed into many subdomains. Each subdomain is treated spectrally [124, 125] – all quantities are decomposed into basis functions, and the basis function expansions are truncated at some finite number of terms which determines the resolution. The advantage of spectral methods is that for problems with smooth solutions, the numerical truncation error decays exponentially with the number of basis functions. Subdomains are split and merged, and their resolution is controlled automatically using adaptive mesh refinement. The computational grid is fixed but mapped onto the physical domain using rotations, translations, scalings and other maps. These maps are controlled to ensure the excised regions remain inside the BH apparent horizons [126]. SpEC uses a generalized harmonic evolution system [127] which chooses coordinates x^μ to satisfy an inhomogeneous wave equation,

$$g_{\mu\nu}\nabla^\rho\nabla_\rho x^\nu = -g^{\alpha\beta}\Gamma_{\mu\alpha\beta} = H_\mu(x,g)\,, \tag{2.9}$$

where $H_\mu(x,g)$ are arbitrary fixed algebraic functions of the coordinates x^μ and the metric $g_{\mu\nu}$ described in [125]. The evolved fields are the metric components $g_{\mu\nu}$, the spatial derivatives of the metric components $\Phi_{i\mu\nu} = \partial_i g_{\mu\nu}$, and fields related to the time derivatives of the metric components $N\Pi_{\mu\nu} = -\partial_t g_{\mu\nu} + \gamma_1 N^i \Phi_{i\mu\nu}$, where N is the lapse, N^i is the shift, and γ_1 is a constraint damping term. This puts the system in first order differential form but introduces additional constraints. The generalized harmonic system is symmetric hyberbolic, which ensures that numerical errors do not grow faster than exponentially (the problem is well-posed). The constraints may grow exponentially in time, and without constraint damping they do. To remedy this, various constraint terms are added to the evolution equations [128, 129]. For exact solutions, the constraints are zero and the solutions are unaffected.

Since the total mass scales out of vacuum solutions of Einstein's equations, the total mass $M_A + M_B$ is typically set to 1 in SpEC simulations. As initial parameters, we then must choose the mass ratio q, the dimensionless spin vectors $\vec{\chi}_A$ and $\vec{\chi}_B$, and the initial BH positions and velocities. By choosing an initial orbital frequency Ω_0 we can use PN to determine the corresponding separation D_0 for a non-eccentric BBH inspiral. By choosing the initial separation vector to be in the x-direction and setting the center of mass to the origin, we fix the initial BH positions. By choosing the initial BH velocities to be in the y-direction and achieve our desired Ω_0 and setting the initial linear momentum of the system to 0, we fix the initial velocities. Initial data is constructed to solve the constraints by specifying a conformal metric in one of two ways, creating an elliptic equation for the conformal factor. Conformally flat data [130] uses a flat conformal metric, and superposed Kerr-Schild (SKS) [131] uses a smoothly blended linear combination of a flat metric and two Kerr metrics centered on either BH. The system is then evolved for a few orbits. Since the choice of D_0 was determined by PN and the BH velocities are chosen to be orthogonal to the separation vector, the system will have some small eccentricity. The orbital eccentricity is measured, and updates to the initial parameters are computed in order to reduce the eccentricity [132, 133]. This now includes an initial rate of change of the orbital separation, as this will be slightly negative for a non-eccentric inspiral. While the eccentricity is above a threshold (typically a few 10^{-4}) this process is iterated. Once the eccentricity is sufficiently small, the evolution is continued through merger and ringdown.

The computational domain is finite, and we are interested in the GWs much further away. One way of obtaining this information is to extract the GWs at several finite radii, and extrapolate this data to an infinite radius by fitting the data to inverse powers of radius [134]. While this is computationally inexpensive, it is not gauge invariant. Choosing spheres of fixed coordinate radius depends on the choice of coordinates, and the extraction itself is not a fully gauge-invariant procedure [135]. A better but more computationally expensive method is to perform Cauchy characteristic extraction (CCE) [136–138]. All the necessary metric data is read from some worldtube, which is simply taken to be a surface of constant coordinate radius.

The data is then evolved outward along null rays out to $\mathcal{I}^+$, where the GWs can be extracted after determining the appropriate (Bondi) coordinates [139]. This method has the advantage of being gauge-invariant up to the symmetry group of future null infinity, which contains super translations in addition to boosts, rotations, and translations [140].

2.4 Gravitational Wave Data Analysis

For a more comprehensive treatment of GW data analysis, see [141]. An interferometric GW detector, such as advanced LIGO, outputs a data stream $d(t)$ consisting of noise from several sources $n_i(t)$, in addition to any strain $s(t)$ due to GWs. The noise sources are assumed to be additive, such that $d(t) = n(t) + s(t)$, where n is the sum of all noise sources. The assumption that n is a zero-mean stationary Gaussian random process is often made. This implies that different Fourier components of the detector noise are uncorrelated. Denoting the Fourier transform of a time-domain quantity with a tilde, we have

$$E[\tilde{n}(f)\tilde{n}^*(f')] = \frac{1}{2}S_n(|f|)\delta(f - f')\,. \tag{2.10}$$

Here $S_n(f)$ is the noise power spectral density, and $E[\cdot]$ denotes the expected value. Note that we have assumed that the time domain data is infinitely long, going from $-\infty$ to ∞, which is why $E[|\tilde{n}(f)|^2]$ is infinite. From now on, we assume instead a finite amount of data. Taking the noise to be Gaussian leads to a very simple expression for likelihood of particular noise realizations once the noise spectral density is known. Defining an inner product

$$(x|y) = 4\,\mathrm{Re}\left[\int_0^\infty \frac{\tilde{x}(f)\tilde{y}(f)^*}{S_n(f)}df\right], \tag{2.11}$$

the probability density of obtaining a given noise configuration n is

$$\log\mathcal{P}[n] = -\frac{1}{2}(n|n) + C, \tag{2.12}$$

where C is a normalization constant. We can see this by discretizing the frequency domain into bins with uniform spacing df. We then treat $\tilde{n}(f_i)$ as an evaluation of a Gaussian random variable with mean 0 and variance $S_n(f_i)/(2df)$, so the probability density of obtaining $\tilde{n}(f_i)$ is

$$\mathcal{P}[\tilde{n}_i] = \sqrt{\frac{df}{\pi S_n(f_i)}}e^{-df\frac{\tilde{n}(f_i)\tilde{n}(f_i)^*}{S_n(f_i)}} \tag{2.13}$$

and the probability density for obtaining $\tilde{n}$ is the product of this quantity over all frequency bins i, since they are independent. Taking the log of this product yields a sum

$$\log \mathcal{P}[n] = C - df \sum_i \frac{\tilde{n}(f_i)\tilde{n}(f_i)^*}{S_n(f_i)}\,, \tag{2.14}$$

which becomes an integral, giving us the previous result in the limit $df \to 0$, since the integrand is an even function of frequency.

Suppose we want to check if the unknown signal s is the GW strain h. Then the noise configuration would be $n = d - h$, and the log-likelihood for h is then

$$\begin{aligned} \log \mathcal{L}[h] &= C - \frac{1}{2}(d - h|d - h) && (2.15)\\ &= C' + (d|h) - \frac{1}{2}(h|h)\,, && (2.16) \end{aligned}$$

where we have moved the $(d|d)$ term into the normalization constant since it is independent of h. We could determine the likelihood of many different candidate signals $h_\lambda(t)$ and determine the most likely signal(s). This should then be compared to the likelihood of no signal ($s(t) = 0$), which is simply C' above. We are projecting the data onto expected signals, which is known as matched filtering [59, 60]. For additive noise, matched filters are the optimal linear filters for maximizing the signal to noise ratio (SNR)

$$\rho = \sqrt{(h|h)}\,. \tag{2.17}$$

The SNR is an important quantity for discriminating the presence of a signal from pure noise. Using a lower threshold for the SNR to determine when a data set should be considered to contain a signal and be labeled as a “detection" will determine a false alarm rate (FAR), or how often pure noise is incorrectly identified as containing a signal. Higher SNR thresholds lead to lower FARs. Real noise in a detector is not Gaussian and contains “glitches" [142, 143], which results in pure noise achieving moderate and high SNRs more often than expected.

Related to the problem of detecting GWs is parameter estimation: once we establish that the data contain a GW signal, we wish to know the type of source of the GW and the related parameters. In other words, we wish to determine which parameters λ give signals $h_\lambda(t)$ with high likelihoods. For a full Bayesian analysis, we should take into account prior astrophysical distributions and use likelihoods to determine the posterior distribution for the parameters of the source. For a non-eccentric BBH system, the eight intrinsic parameters are the two component masses and spin vectors. There are also extrinsic parameters: the time of coalescence, the distance between the source and the Earth, two parameters describing the direction of the Earth in a binary-centered coordinate system (or the inclination angle and orbital phase at coalescence in an Earth-centered coordinate system), and three angles describing the orientation of the GW detector in a binary-centered coordinate system (or the sky-location and polarization angle in an Earth/detector-centered coordinate system). All together, this makes 15 parameters. A full analysis would evaluate the signal $h_\lambda(t)$ that would be seen in the detector for all possible values of λ in the 15 dimensional parameter space, and evaluate the likelihood as a function of this parameter space. Evaluating even a computationally inexpensive function in a reasonably dense way is not feasible, so in pracice [20], Markov chain Monte Carlo [7] methods or nested sampling [8] methods are used. If $h_\lambda(t)$ can be expressed as a linear sum of coefficients depending only on λ multiplying basis vectors depending only on t, faster methods that evaluate inner products with the basis vectors can be used [9, 144, 145].

While checking for detections or performing parameter estimation, we must have accurate templates $h(t)$. In general, our template waveforms will have some error, whether it is modeling error or due to finite resolution in a NR simulation. Suppose $h(t)$ is the waveform predicted by GR, and our waveform model predicts $H(t) = h(t) + \epsilon g(t)$ where we will consider the limit $\epsilon \to 0$. We can rewrite this as

$$H(t) = (1 + \epsilon_{||}) \left(h(t) + \frac{\epsilon}{1 + \epsilon_{||}} g_{\perp}(t) \right) , \tag{2.18}$$

where we have

$$g_{\perp}(t) = g(t) - \frac{(g|h)}{(h|h)} h(t) , \tag{2.19}$$

$$\epsilon_{||} = \epsilon \frac{(g|h)}{(h|h)} . \tag{2.20}$$

By allowing a small fractional bias of $\epsilon_{||}$ in the measured distance, we can simplify the calculations by defining

$$\tilde{\epsilon} = \frac{\epsilon}{1 + \epsilon_{||}} = \epsilon + \mathcal{O}(\epsilon^2) \tag{2.21}$$

and instead considering

$$\tilde{H}(t) = h(t) + \frac{\epsilon}{1 + \epsilon_{||}} g_{\perp}(t) = h(t) + \tilde{\epsilon} g_{\perp}(t) . \tag{2.22}$$

The difference in log-likelihoods when using $\tilde{H}(t)$ compared to using $h(t)$ will then be

$$\log\mathcal{L}[\tilde{H}] - \log\mathcal{L}[h] = (d|\tilde{H}) - (d|h) - \frac{1}{2}\left((\tilde{H}|\tilde{H}) - (h|h)\right) \tag{2.23}$$

$$= \tilde{\epsilon}(d|g_\perp) - \frac{1}{2}\tilde{\epsilon}^2(g_\perp|g_\perp) \tag{2.24}$$

$$= \tilde{\epsilon}(n|g_\perp) - \frac{1}{2}\tilde{\epsilon}^2(g_\perp|g_\perp)\,. \tag{2.25}$$

The noise term $(n|g_\perp)$ will depend on the particular noise realization, but its expected value is 0. The expected decrease in the log-likelihood is then $\frac{1}{2}\tilde{\epsilon}^2(g_\perp|g_\perp)$, which is $\frac{1}{2}\epsilon^2(g_\perp|g_\perp)$ up to order ϵ^3 corrections. Defining the match, which is also called the overlap, as

$$\mathcal{M}(x,y) = \frac{(x|y)}{\sqrt{(x|x)(y|y)}} = (\hat{x}|\hat{y})\,, \tag{2.26}$$

and the mismatch as $1 - \mathcal{M}(x,y)$, we have

$$\mathcal{M}(h,H) = \mathcal{M}(h,\tilde{H}) = 1 - \frac{1}{\sqrt{1 + \tilde{\epsilon}^2\frac{(g|g)}{(h|h)}}} \tag{2.27}$$

$$= -\frac{1}{2}\epsilon^2\frac{(g|g)}{\rho^2} + \mathcal{O}(\epsilon^3)\,. \tag{2.28}$$

We therefore have that the expected loss in log-likelihood is

$$E[\delta\log\mathcal{L}] = \rho^2(1 - \mathcal{M}(h,H)) + \mathcal{O}(\epsilon^3)\,. \tag{2.29}$$

We can tolerate waveform errors with mismatches much less than ρ^{-2}, as they will not noticeably affect the results of parameter estimation. As the loudest SNR could be ~ 50 once 40 detections are made [146], mismatches of less than 4×10^{-4} might be needed in order for waveform modeling to have a completely negligible effect on parameter estimation. Detector calibration errors place a lower bound on the mismatches we should strive to achieve in waveform models [147]. For GW150914, these calibration errors are $\sim 5\%$ in the amplitude A of the signal and $\sim 3^\circ$ in the phase φ of the signal [10]. The corresponding mismatches are roughly

$$1 - \mathcal{M} \sim \left(\frac{\delta A}{A}\right)^2 + (\delta\phi)^2 \sim 0.005\,, \tag{2.30}$$

so waveform model mismatches below ~ 0.002 should lead to a negligible contribution to parameter estimation errors. If detector calibration improves in the future, lower mismatches would be needed. Currently, typical waveform models have mismatches of up to a few percent with NR simulations, while NR resolution studies and comparisons between codes [148] estimate NR waveforms have mismatches with GR of $\mathcal{O}(10^{-4})$ for typical resolutions.

It is possible to have a waveform model where mismatches when compared to the GR prediction for the same parameters are large, but when we minimize the mismatch with the GR prediction over the model waveform's parameters we achieve small mismatches. The match maximized over all parameters is called the effectualness [149]. Using such a model would lead to a negligible loss in the detection rate, and would not erronously identify a GR waveform as being inconsistent with GR, but would yield biased parameter measurements. If mismatches can be made small by optimizing over only some parameters which we are not interested in measuring, such as the exact time of arrival or the orbital phase at some reference frequency, we will still obtain unbiased measurements for the other parameters of interest such as the black hole masses. This partially-minimized mismatch is often called the unfaithfulness [149], with the maximized match called the faithfulness. An unfaithful model could still yield reasonable results – if the modeling error somehow had the effect of reducing all likelihoods by some constant factor, the most likely parameters would still be correct. It is therefore useful to determine parameter biases explicitly – for a given waveform h_λ, find the model waveform $H_{\tilde{\lambda}}$ which maximizes the match, and report the parameter differences $\lambda - \tilde{\lambda}$ [71]. If the model is sufficiently faithful, this parameter difference will be less than the width of the likelihood peak by the triangle inequality.

2.5 Waveform Models

Several methods of generating gravitational waveforms already exist. Numerical relativity, which can be sufficiently accurate for use in advanced generation GW data analysis, is far too slow to be used directly. Other models are very fast to evaluate - some even produce waveforms in the frequency domain which eliminates the need to perform a fast Fourier transform (FFT) on time-domain data before taking frequency domain inner products with the detector signal. While some of these models are sufficiently accurate for some subset of the parameter space, they all have significant limitations. Waveform generation methods are summarized in Table 2.1.

Post-Newtonian (PN) [61, 62] waveforms are some of the most widely used. By expanding both the orbital dynamics of the BBH system and the gravitational strain emitted in powers of the orbital velocity divided by the speed of light, PN is extremely accurate during the early inspiral when velocities are slow and the gravitational interactions are weak, but becomes increasingly inaccurate as the coalescence progresses. The PN equations then fail catastrophically near merger, and PN waveforms do not contain a merger or ringdown phase. Instead, the PN GW amplitude becomes infinite at some finite time. Typically the waveforms are truncated at some earlier time, for example at the inner-most stable circular orbit. The time when PN stops being sufficiently accurate depends on the BH parameters. For an equal-mass non-spinning BBH system, PN waveforms can be accurate to within ~ 1 radian even 2 orbits before merger, while for other systems PN can become inaccurate hundreds of orbits before merger [72, 150]. Several time-domain PN approximants exist, which expand different quantities in terms of the PN expansion parameter [151]. There is also the frequency-domain approximant TaylorF2 [150] for non-precessing waveforms, which assumes that the waveform amplitudes and phase time derivatives evolve slowly to make use of a stationary phase approximation (SPA). The result is an expansion of the frequency-domain phases and amplitudes in terms of powers of the frequency.

The effective-one-body (EOB) model [11, 67–70, 152] associates the BH dynamics with the dynamics of a single particle in some external effective spacetime, expanding the spacetime as in PN. It chooses the coefficients such that the resulting inspiral waveforms agree with PN up to the chosen PN order. The dynamics terminate with a plunge of the particle into a BH, onto which a ringdown phase consisting of decaying quasi-normal modes is attached. EOB waveforms therefore have the advantage over PN of having merger and ringdown phases, although they are somewhat more computationally expensive to integrate. EOB has been shown to be more accurate than pure PN when compared with NR waveforms [16, 72]. It has also been improved by the inclusion of unknown PN coefficients, which are then calibrated using NR waveforms. The transition from a plunge into a ringdown can also be calibrated to NR waveforms. These calibrated models are called EOBNR or SEOBNR, where the 'S' indicates the model includes spin. Different particular calibrations are denoted by EOBNRv1, EOBNRv2 etc. The most current calibrated non-precessing EOB is SEOBNRv4 [11]. There is also a precessing EOB model SEOBNRv3 [70]. While EOBNR models contain some subdominant modes [68], SEOBNR models only include the dominant $\ell = 2$, $|m| = 2$ modes. They can therefore become inaccurate for certain directions of GW emission near merger where other modes are no longer negligible [153]. This causes EOB waveforms to perform poorly for higher mass systems which contain fewer cycles in the aLIGO sensitivity band, making the merger phase the dominant contribution to the SNR [269].

Waveform Method	NR	PN	EOBNR	SEOBNRv4
Domain	Time	Either	Time	Time
$q_{\max}$	20	∞	100 (6)	100 (8)
Spins	All	All	None	Aligned
$\lvert\vec{\chi})_{\max}$	~ 0.99	1	0	0.98
Modes	$\ell \leq 8$	Any	$(2,1)$, (ℓ,ℓ), $\ell \leq 5$	$(2,\pm 2)$
Merger	Yes	No	Yes	Yes
Cost	Weeks	$\sim 1\,\mathrm{s}$	$\sim 10\,\mathrm{s}$	$\sim 10\,\mathrm{s}$
Typical Mismatch	10^{-4}	0.1	0.01	< 0.01

Waveform Method	SEOBNRv3	PhenomD	PhenomP	
Domain	Time	Freq.	Freq.	
$q_{\max}$	∞ (8)	∞ (18)	∞ (18)	
Spins	All	Aligned	Some precession	
$\lvert\vec{\chi})_{\max}$	1 (0.98)	1 (0.98)	1 (0.98)	
Modes	$\ell = 2$	$(2,\pm 2)$	$\ell = 2$	
Merger	Yes	Yes	Yes	
Cost	$\sim 10\,\mathrm{s}$	$< 1\,\mathrm{s}$	$< 1\,\mathrm{s}$	
Typical Mismatch	0.01	0.01	0.01	

Table 2.1: Capabilities of different waveform generation methods. For the maximum mass ratio q_{Max} and spin $|\vec{\chi}|_{\mathrm{Max}}$, the maximum parameters used for calibration are in parenthesis where applicable. Typical mismatches are order-of-magnitude estimates for BBH waveforms from sources similar to GW150914, with detector noise similar to GW150914.

Phenomenological waveform models [63–66] are designed with GW data analysis in mind. They are evaluated in the frequency domain, and typically agree with TaylorF2 at small frequencies. After investigating the structure of NR waveforms, a model is designed and its coefficients are fit to a set of NR calibration waveforms. As they are very cheap to evaluate, they are suitable for detection searches as well as rapid parameter estimation studies where a fast result is needed. The most recent non-precessing Phenom model is PhenomD [65], which is calibrated using NR waveforms up to mass ratio 18 and aligned spins up to 0.98. Instead of being exactly TaylorF2 at small frequencies, it is designed to agree with uncalibrated SEOBNRv2 at small frequencies. Within its calibration range, it typically achieves mismatches of the $(\ell,m) = (2,2)$ mode of less than 3%. There is also the precessing waveform model PhenomP [66], which uses aligned spin PhenomC or PhenomD waveforms and “twists" them up using PN precession dynamics.

2.6 Reduced Order Modeling

The field of reduced order modeling (ROM) [73–76] treats the space of all solutions to a given difficult or computationally expensive problem as a (potentially non-linear) subspace of a high dimensional linear space, and seeks to find a linear space of much lower dimension which can represent all solutions up to some given tolerance. The space of BBH waveforms can be thought of some 7-dimensional curved submanifold of the space of all possible linear combinations of BBH waveforms. While this linear space has a very high (or infinite) dimension, it can be well approximated by a linear space of much lower dimension. We will construct a reduced basis [73, 75] to represent the lower-dimensional linear space (compression in parameter space), build the corresponding empirical interpolant [74, 76], which is an application-specific time interpolant constructed from data (compression in time), and finally construct a surrogate model, which rapidly generates solutions to the problem at arbitrary parameters with a negligible loss in accuracy. These steps are all fairly simple, but since there is a lot of jargon and notation we will step through a toy problem. A basic flow chart of the steps is shown in Fig 2.1.

Suppose we seek time-dependent solutions $f(t;\lambda)$ to a parametrized problem where $t \in T$ is time and $\lambda \in \Lambda$ is one or more parameters. Suppose we have a method of obtaining solutions for any given value of $\lambda \in \Lambda$, but that our method is computationally expensive. For example, λ could consist of two BH masses and spin vectors, and $f(t;\lambda)$ could be the $(\ell,m) = (2,2)$ mode of the GWs emitted from a BBH system with masses and spins given by λ, aligned such that its peak amplitude occurs at $t = 0$. For our toy problem, we will use leading order PN $(\ell,m) = (2,2)$ mode waveforms with a total mass of 1 parametrized by the symmetric mass ratio $\nu = m_A m_B/M^2$. Setting $t = 0$ to be the “merger" when the PN expressions become infinite, we will use $T = [-5000,-100]$ and $\Lambda = [0.1, 0.25]$ with $\lambda = \nu$. Our solutions are then described analytically by

$$f(t;\lambda) = A(t;\lambda)e^{i\varphi(t;\lambda)}\,, \tag{2.31}$$

$$A(t;\lambda) = \left(\frac{16\pi^2\lambda^3}{-5t}\right)^{1/4}\,, \tag{2.32}$$

$$\varphi(t;\lambda) = \frac{2}{5}\left(\frac{5}{\lambda}\right)^{3/8}(-t)^{5/8}\,. \tag{2.33}$$

While this toy problem is not suitable for ROM since the known solution is computationally cheap to evaluate, it is a useful example for demonstration purposes. Evaluations of f are shown in Fig. 2.2.

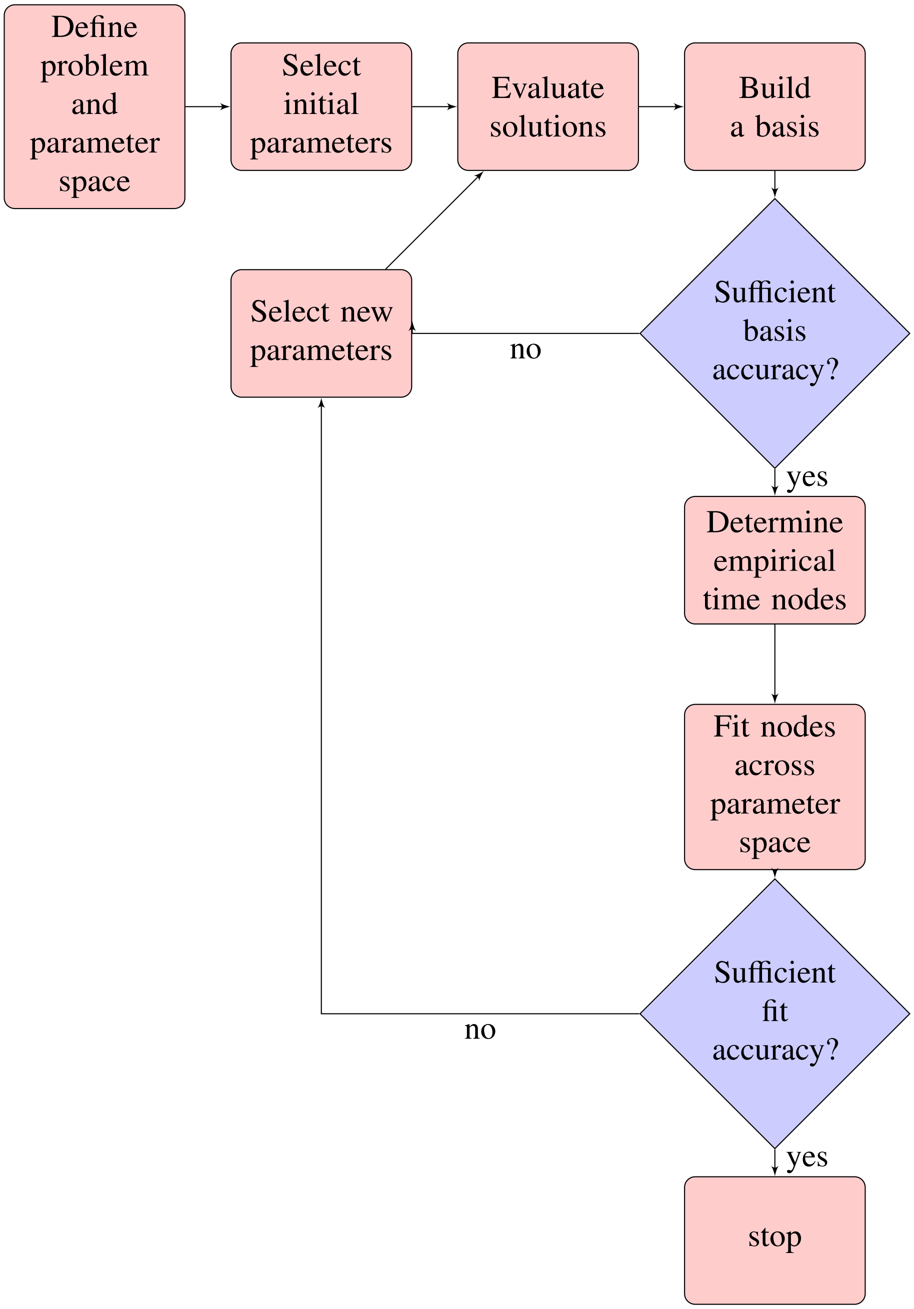

Figure 2.1: Basic flow chart for building a surrogate model.

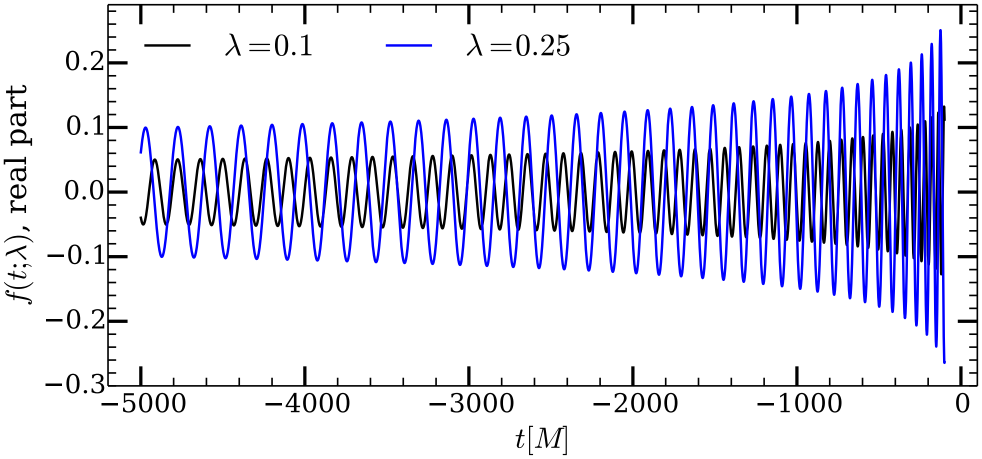

Figure 2.2: Evaluations of the toy problem at the extremal values of the parameter $\lambda = \nu$.

Our first step is to construct a reduced basis, which is an orthonormal basis for a low-dimensional linear space which can represent all solutions up to some given tolerance. The standard method of constructing a reduced basis is to first construct a training set (TS) of solutions: we will use the uniformly spaced TS parameters $\lambda_i = 0.1+0.15i/(N_{TS}-1)$, where $i = 0, 1, \ldots, N_{TS}-1$ and we will choose $N_{TS} = 500$. We must also choose an inner product and norm: we will use

$$\langle u(t), v(t)\rangle = \int_{-5000}^{-100} u(t)v(t)^* dt\,, \tag{2.34}$$

which gives us the L^2 norm $||u||^2 = \langle u, u\rangle$. We first evaluate and normalize all TS solutions:

$$\hat{f}_i = \frac{f(t;\lambda_i)}{||f(t;\lambda_i)||}\,. \tag{2.35}$$

We must then initialize our reduced basis: we will seed it with the first solution, $\hat{f}_0$, and denote the basis by $B_1 = \{\hat{b}_1\} = \{\hat{f}_0\}$ where the subscript of B indicates the dimension of the basis and the subscript of $\hat{b}$ indicates the ordering of the basis vectors. We can determine how accurately our basis represents our normalized TS solutions by computing projection errors. Denoting the projection operator for the basis $B_n = \{\hat{b}_j\}_{j=1}^n$ by P_n, the projection errors are

$$\epsilon_n^i = ||\hat{f}_i - P_n\hat{f}_i|| = ||\hat{f}_i - \sum_{j=1}^{n}\langle\hat{f}_i, \hat{b}_j\rangle\hat{b}_j||\,, \tag{2.36}$$

and the maximum projection error is $\epsilon_n = \max_i(\epsilon_n^i)$. The projection errors of our seed basis, ϵ_1^i, are shown in black in Fig. 2.3.

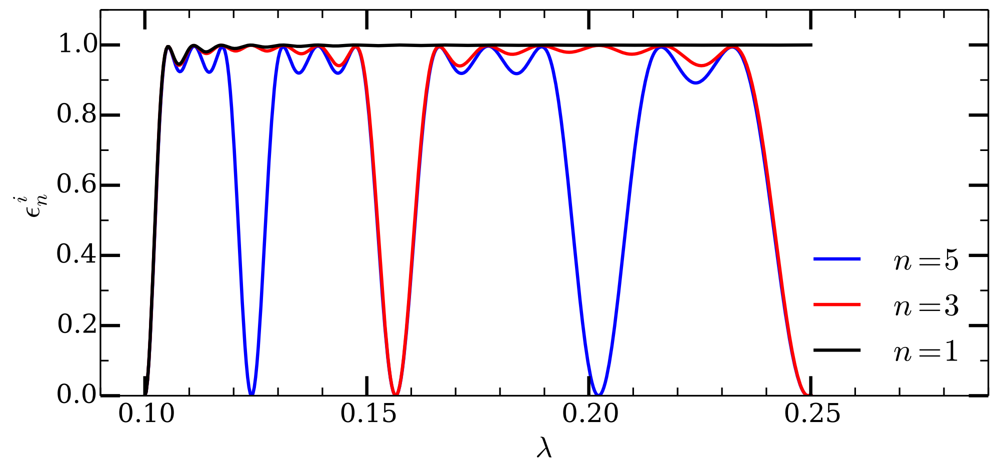

Figure 2.3: Projection errors for reduced bases of sizes 1, 3, and 5. For small basis sizes, we usually have maximum projection errors of 1 and the basis is only accurate very close to the chosen greedy parameters.

We now run a greedy algorithm which iteratively increases the dimension and accuracy of the reduced basis by adding the TS solution with the worst projection error after orthonormalizing it with the current basis. Specifically, we set

$$i_n^* = \arg\max_i \epsilon_n^i \,, \tag{2.37}$$

$$b_{n+1} = \hat{f}_{i_n^*} - P_n \hat{f}_{i_n^*} \,, \tag{2.38}$$

$$\hat{b}_{n+1} = \frac{b_{n+1}}{\|b_{n+1}\|} \,. \tag{2.39}$$

We iterate this greedy step until either ϵ_n becomes small enough, we have used up all TS solutions, or i_n^* has already been used in a previous greedy step (indicating we have reached numerical roundoff errors). Some resulting basis vectors for our toy problem are shown in Fig. 2.4, and the projection errors as a function of basis size are shown in black in Fig. 2.5. We see that once the reduced basis reaches a certain size, the projection errors fall off exponentially with basis size until numerical roundoff errors become relevant. This is true for all smooth problems. We wish to use our reduced basis to represent not only TS solutions, but solutions with any parameter $\lambda \in \Lambda$. To ensure our basis is sufficiently accurate, we perform *a posteriori* verifications by sampling the parameter space more densely than our training set and computing the maximum projection errors (shown in red and blue in Fig. 2.5 . Since the errors do not change noticeably as we increase our sampling rate, we conclude that the computed TS projection errors are a reliable estimate of the maximum error over the whole parameter space and that 40 basis elements are sufficient to represent

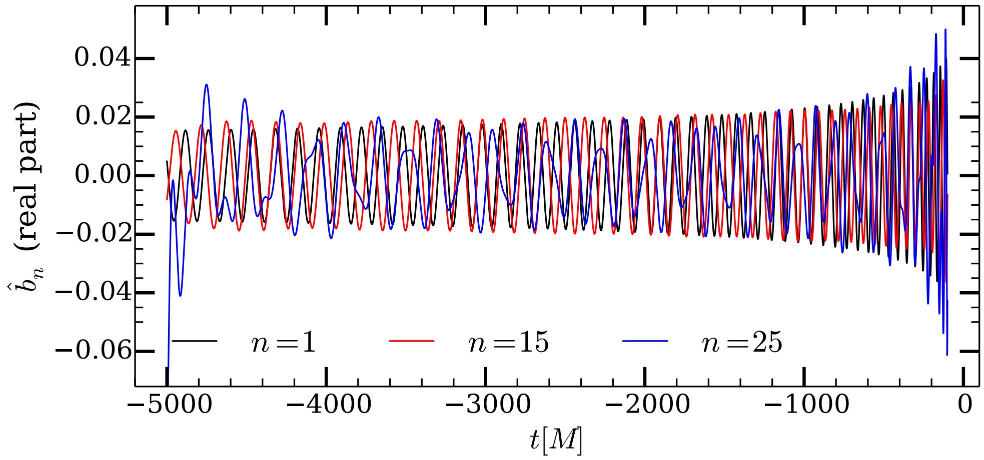

Figure 2.4: Some reduced basis vectors are shown.

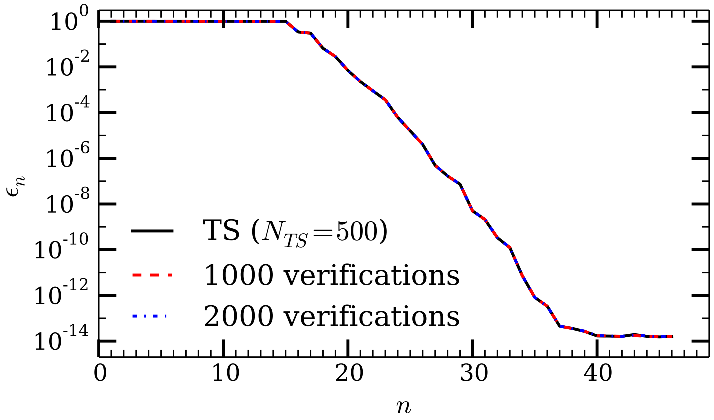

Figure 2.5: Projection errors as a function of the reduced basis dimension n. The maximum projection error over all training space parameters is shown in black. A posteriori verifications are done using denser sampling to ensure the basis is accurate for all parameters and not just the training space parameters (red and blue).

our solutions up to projection errors of less than 10^{-13}. Note that to build our training set and perform the *a posteriori* verification, we had to evaluate solutions $f(t;\lambda)$ many times. Typically this is done in an *offline* stage, and the basis is used in an *online* stage where little computational time is available. For the case of BBH GWs, computing solutions for a dense TS using NR would be prohibitively expensive, even in an offline stage. We will therefore resort to other methods of determining the greedy parameters $\{\lambda_{i_n^*}\}$ described later, and then we may evaluate the solutions for the greedy parameters using numerical relativity to construct the reduced basis.

For our toy problem, we will choose to use the reduced basis with size $N = 40$, meaning we can write

$$f(t;\lambda) = \sum_{i=1}^{N} c_i(\lambda)\hat{b}_i(t)\,, \tag{2.40}$$

where the basis coefficients are given by

$$c_i(\lambda) = \langle f(t;\lambda), \hat{b}_i(t)\rangle\,. \tag{2.41}$$

We have omitted projection errors, under the assumption that they are negligible. Once we have our reduced basis B_N, we can construct an empirical interpolant. This allows us to drastically downsample our solutions in time and then use the interpolant to upsample with little or no loss in accuracy. We first note that if we know the values of $f(t;\lambda)$ at N times $\{t_j\}_{j=1}^{N}$ (where this is the same N denoting the size of the reduced basis, which is $N = 40$ for our toy problem), then we can evaluate Eq. 2.40 at each time t_j to obtain a linear system of N equations with N unknown basis coefficients:

$$\vec{f}(\lambda) = V\vec{c}(\lambda)\,. \tag{2.42}$$

Here $\vec{f}(\lambda) = \{f(t_j;\lambda)\}$, $\vec{c}(\lambda) = \{c_i(\lambda)\}$ and $V_{ji} = \hat{b}_i(t_j)$ is an $N \times N$ matrix. Provided that V is not singular, we can invert it to determine the basis coefficients:

$$\vec{c}(\lambda) = V^{-1}\vec{f}(\lambda)\,. \tag{2.43}$$

This will be our empirical interpolant – determining $\vec{f}(\lambda)$ allows us to evaluate $f(t;\lambda)$ for any $t \in T$. Explicitly, denoting our interpolant by I_N, we have

$$I_N[\vec{f}(\lambda)](t) = \vec{b}(t)V^{-1}\vec{f}(\lambda)\,, \tag{2.44}$$

where $\vec{b}(t) = \{\hat{b}_i(t)\}$. More generally, we may attempt to interpolate any function $g(t)$ by first evaluating $\vec{g} = \{g(t_j)\}_{j=1}^{N}$ and computing

$$I_N[\vec{g}](t) = \vec{b}(t)V^{-1}\vec{g}\,. \tag{2.45}$$

We note that the interpolant may be evaluated for any time $t \in T$. If $g(t)$ lies in the span of our basis B_N, which is the case for all solutions $f(t;\lambda)$, then $I_N[\vec{g}](t) = g(t)$ for all times t. Otherwise, the interpolant is only guaranteed to be exact at the N chosen times $\{t_j\}$. Since the interpolant is linear, propagating the projection errors through the interpolant is straightforward [76]. If we fix our desired time sampling for the output of the interpolant so that $\vec{b}(t) = \{\hat{b}_i(t_k)\}_{k=1}^{L}$ is a matrix, we can multiply it with V^{-1} in advance to obtain an $L \times N$ matrix R, where L is the number of output time samples. The empirical interpolant is then just

$$I_N[\vec{g}](t) = R\vec{g}\,, \tag{2.46}$$

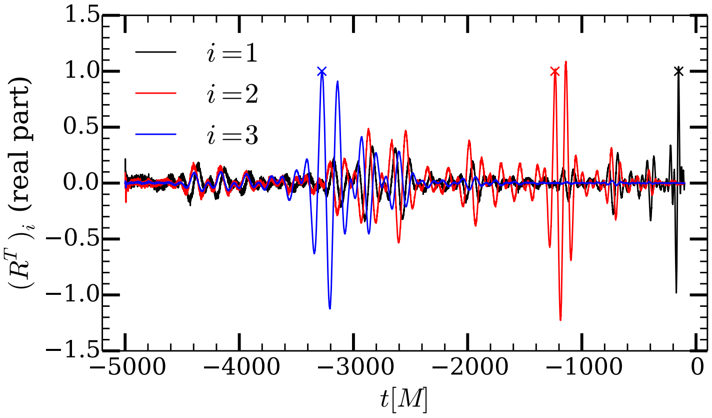

Figure 2.6: Some of the functions used in the empirical interpolant (columns of the R matrix).

and interpolates N input times onto L output times. All that is left to do is to choose the N input times $\{t_j\}$, called empirical nodes, to avoid singular systems of equations (ensure V is invertible). We will make use of another greedy algorithm to choose the empirical nodes. We seed the algorithm by choosing $t_1 = \arg\max_t |b_1(t)|$. Once we have chosen n times, we may construct an empirical interpolant I_n using B_n. Since $\hat{b}_{n+1}(t)$ does not lie in the span of B_n, I_n will not successfully interpolate $\hat{b}_{n+1}(t)$ and we may compute the interpolation error:

$$E_n(t) = \hat{b}_{n+1}(t) - I_n[\{\hat{b}_{n+1}(t_j)\}_{j=1}^{n}](t)\,. \tag{2.47}$$

We then choose $t_{n+1} = \arg\max_t |E_n(t)|$. Had we chosen t_{n+1} such that $E_n(t_{n+1}) = 0$, we would have obtained a singular system by increasing the dimensions of the V matrix by one without increasing its rank.

Some of the 40 columns of R for our toy problem, which have length equal to the number of output time samples (chosen to be 10000 in our case), are shown in Fig. 2.6. The interpolant is application-specific, and performs poorly when interpolating functions which have large projection errors with the reduced basis. In Fig. 2.7, we attempt to use our empirical interpolant to interpolate a constant function. Since a constant function does not lie in the span of the reduced basis, the interpolant is unable to accurately reproduce the original function. In Fig. 2.8 we show that the empirical interpolant can very accurately interpolate solutions $f(t;\lambda)$ for all $\lambda \in [0.1, 0.25]$. It is interesting to note that all of our solutions with $\lambda \in [0.1, 0.25]$ have between 36 and 50 GW cycles, yet we can accurately interpolate them with just 40 points – this is fewer than the Nyquist rate of 2 points per cycle!

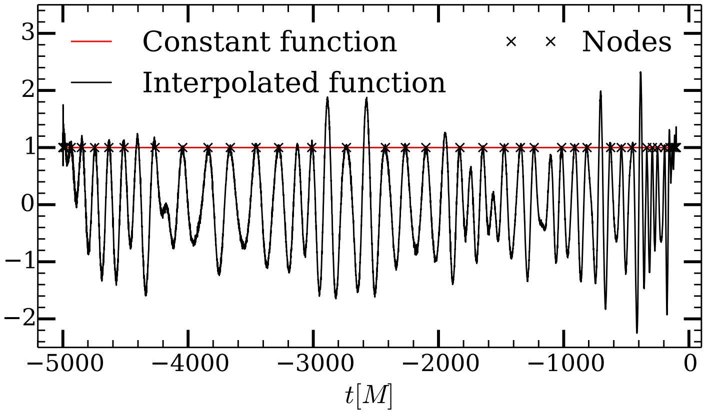

Figure 2.7: Using the empirical interpolant to interpolate a constant function leads to poor results since it does not lie in the span of the reduced basis

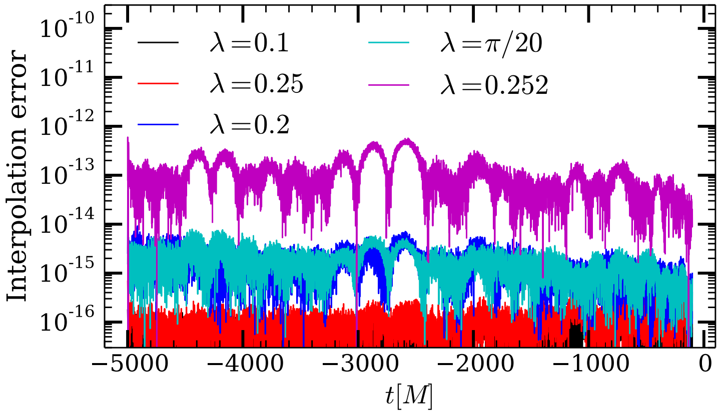

Figure 2.8: The interpolation errors $|f(t;\lambda) - I_{40}[\vec{f}(\lambda)](t)|$ are shown for several values of λ. The errors in the interpolant for parameters in the training range originate from numerical roundoff errors. The interpolant is accurate for parameters not included in the training set such as $\pi/20$, but not for those outside the training range such as 0.252. Extending further beyond the training range of $[0.1, 0.25]$ would lead to larger errors, eventually becoming order unity.

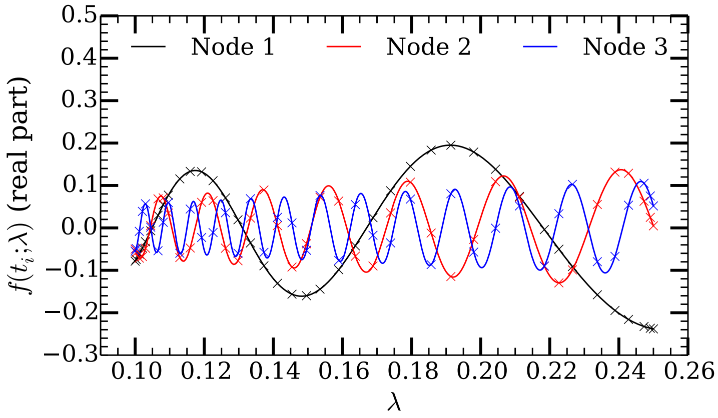

Figure 2.9: Solution data at a few empirical nodes. The 'x' marks are located at values of λ selected by the greedy algorithm for the reduced basis. In this case, the nodes oscillate rapidly and would require additional data points to obtain an accurate surrogate.

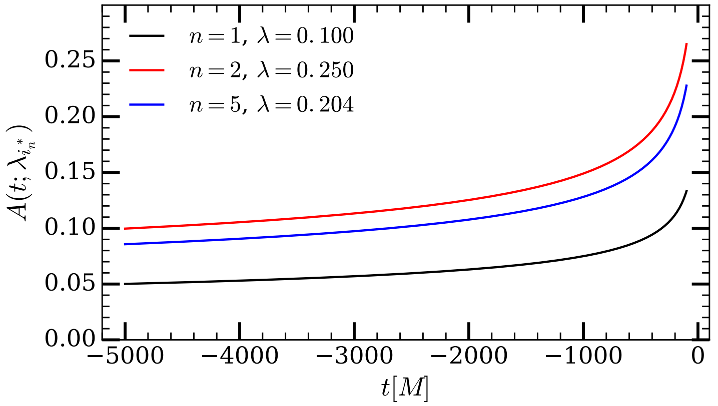

Figure 2.10: Solution amplitudes at selected greedy parameters.

Our final step will be to build a surrogate model. If we can accurately model the behavior of the solutions at the empirical nodes $f(t_i;\lambda)$, then we can predict solutions for any $\lambda \in \Lambda$ without computing it the "expensive" way (in the case of our toy model, that would be evaluating it according to Eq. 2.31). In the case of our toy model, the nodes oscillate rapidly as we vary λ, as shown in Fig. 2.9. If we try to fit or interpolate the complex node data across the parameter space Λ using only the 40 data points we have at the greedy parameters, we will obtain large errors. However, if we take our 40 known solutions and decompose them into an amplitude function and a phase function, after aligning the initial phases (corresponding to a coordinate rotation of the BBH systems about the direction of orbital angular momentum), we obtain non-oscillatory functions shown in Fig. 2.10

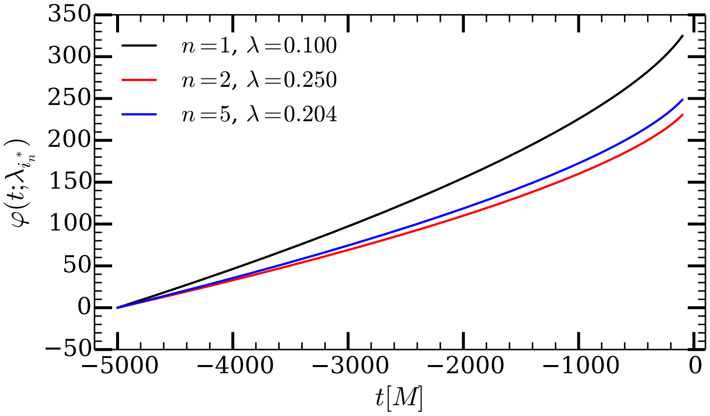

Figure 2.11: Solution phases at selected greedy parameters. Each phase function is shifted such that it begins at 0, which corresponds to rotating the coordinate system of the BBH system about the orbital angular momentum axis.

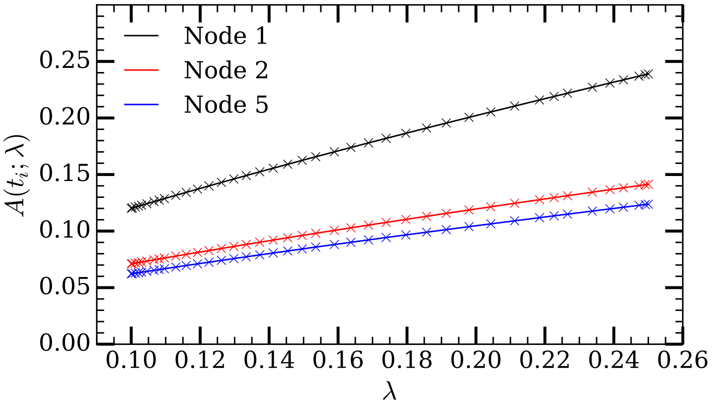

Figure 2.12: Amplitude at empirical nodes – solid curves are the analytic solutions, and ‘x’ ticks are the greedy parameters chosen for the reduced basis. The amplitude varies slowly across parameter space and a polynomial fit to the 40 known data points should perform well.

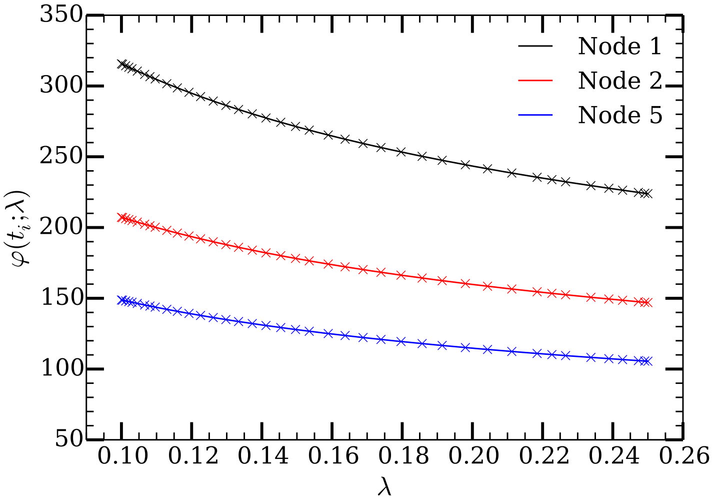

Figure 2.13: Phase at empirical nodes – solid curves are the analytic solutions, and 'x' ticks are the greedy parameters chosen for the reduced basis. The phase also varies slowly across parameter space and a polynomial fit to the 40 known data points should perform well.

and Fig. 2.11 with correspondingly simple node variation across parameter space shown in Fig. 2.12 and Fig. 2.13. We can fit these node functions to polynomials in λ using our 40 known data points at the greedy parameters $\lambda_{i_n^*}$. We then have a surrogate, with an accuracy determined by the accuracy of the fits, empirical interpolant, and reduced basis. To evaluate the surrogate at a new parameter λ, we first evaluate the 40 amplitude node fits and 40 phase node fits. We combine these to obtain 40 complex node values. Finally, we use the empirical interpolant to obtain the solution for all times $t \in T$. The surrogate method is shown schematically in Fig. 2.14. Note that we could have decomposed our data into amplitudes and phases from the beginning, building two separate reduced bases and two separate empirical interpolants. The dimensionality of the reduced bases would not need to be as high since the amplitudes and phases vary more slowly in time. The downside of this approach is that we no longer have a linear basis for the solutions, preventing us from building reduced-order quadratures [154, 155].

Using the standard method for constructing a reduced basis described above involves many solution evaluations to form a training set. If we wish to build a surrogate for SpEC waveforms, building a training set is currently computationally impossible. What we do instead is to use a different GW model such as PN or EOB to determine the greedy parameters, and then we perform SpEC simulations for only the greedy parameters. This allows us to construct a reduced basis and empirical interpolant

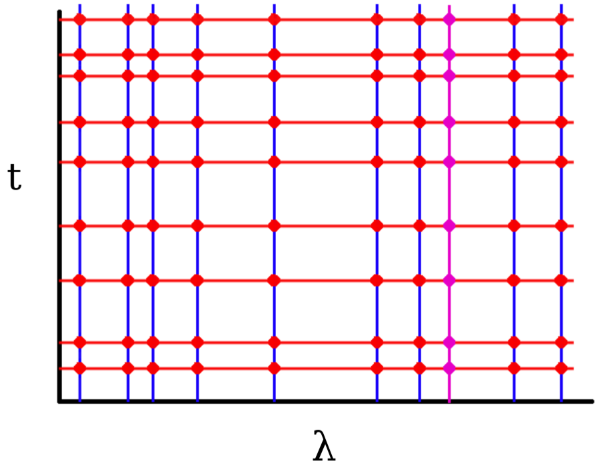

Figure 2.14: A schematic diagram of a surrogate model, with $N = 9$. In the (λ, t) plane, a solution for a particular value of λ is a vertical line. First, N greedy parameters are chosen and the solutions are obtained (vertical blue lines). Next, N empirical time nodes are chosen (horizontal red lines) and an empirical interpolant is generated. For a fixed λ, the empirical interpolant interpolates data from the time nodes onto all times. Finally, the known data at the empirical nodes (red dots) is fitted across parameter space. To evaluate the surrogate at a new value of λ, we first evaluate the fits of the empirical nodes (magenta dots). Then we use the empirical interpolant to obtain the solution for all times (magenta vertical line).

using SpEC waveforms, at which point there is no further need for input from PN or EOB – they are only used to determine the relevant points in parameter space.

3 A sparse representation of gravitational waves from precessing compact binaries

3.1 Background and Contributions

This chapter was published in Physical Review Letters on July 7 2014 [248], in issue 2 of volume 113 which can be found at `https://link.aps.org/doi/10.1103/PhysRevLett.113.021101`. A preprint can be found at `https://arxiv.org/abs/1401.7038`.

This chapter represents the first step in the project of building a surrogate model of numerical relativity waveforms in the full $7d$ non-eccentric parameter space of binary black hole mergers. Building a surrogate model [76] begins with the construction of a reduced linear basis spanning the space of all solutions, up to some error tolerance. These reduced bases had been found to require over 1000 waveforms for parameter dimensions $2 \leq d \leq 4$ [156]. While the basis size increases only mildly as the dimension increases for these non-precessing cases, it was unclear if this would remain the case when precession is present and the waveforms have additional phenomenology. Even if the increase in basis size remains mild, performing several thousand numerical relativity simulations would be quite expensive. The key insight for this project was to avoid constructing a reduced basis for the rapidly oscillating waveform modes, and to instead construct bases for more slowly-varying functions which describe the waveforms such as unit quaternions representing the precessional dynamics. This allows an accurate representation of precessing waveforms using as few as 50 judiciously selected waveforms.

Most of the insights and new methods in this paper were developed collaboratively between all authors. I contributed to the random resampling method and using the coprecessing frame. I implemented the basis construction and determined the errors, and contributed to the writing and editing of the manuscript.

A sparse representation of gravitational waves from precessing compact binaries

Jonathan Blackman,[1] Bela Szilagyi,[1] Chad R. Galley,[1] and Manuel Tiglio[2, 1]

[1] *Theoretical Astrophysics, California Institute of Technology, Pasadena, CA, 91125, USA*
[2] *Center for Scientific Conputation and Mathematical Modeling, and Joint Space Sciences Institute, Maryland Center for Fundamental Physics, Department of Physics, University of Maryland, College Park, MD 20742, USA*

Many relevant applications in gravitational wave physics share a significant common problem: the seven-dimensional parameter space of gravitational waveforms from precessing compact binary inspirals and coalescences is large enough to prohibit covering the space of waveforms with sufficient density. We find that by using the reduced basis method together with a parametrization of waveforms based on their phase and precession, we can construct ultra-compact yet high-accuracy representations of this large space. As a demonstration, we show that less than 100 judiciously chosen precessing inspiral waveforms are needed for 200 cycles, mass ratios from 1 to 10 and spin magnitudes ≤ 0.9. In fact, using only the first 10 reduced basis waveforms yields a maximum mismatch of 0.016 over the whole range of considered parameters. We test whether the parameters selected from the inspiral regime result in an accurate reduced basis when including merger and ringdown; we find that this is indeed the case in the context of a non-precessing effective-one-body model. This evidence suggests that as few as ~ 100 numerical simulations of binary black hole coalescences may accurately represent the seven-dimensional parameter space of precession waveforms for the considered ranges.

3.2 Introduction

Gravitational radiation produced by stellar-mass compact binaries of neutron stars and/or black holes are expected to be the main signals detected by the advanced generation of gravitational wave detectors [35, 36, 157, 158]. Detecting these signals and estimating the parameters of their sources require the ability to sufficiently sample the space of precessing compact binary waveforms. A compact binary intrinsically depends on its mass ratio and the spin angular momentum components of each body, which forms a 7-dimensional space for gravitational waveforms [1].

Much progress has been made in sampling the subspace of non-spinning compact binary waveforms over the last decade. However, many relevant applications, from data analysis for gravitational wave searches and parameter estimation studies to numerical relativity simulations of binary black hole coalescences, face a common challenge. In particular, going from the non-spinning subspace to the full 7d space naively requires prohibitively more samples for the same coverage simply because the volume of the space grows exponentially with dimension. As a result, the general consensus in the gravitational wave community is that the computational complexity associated with building template banks for matched-filter searches, with making parameter estimation studies, and with modeling precessing compact binaries by expensive numerical simulations is intractable (e.g., see [159]). This phenomenon entails what is called the *curse of dimensionality* [160].

In this paper, we show how to beat the curse of dimensionality for precessing compact binary inspirals. We find that only 50 judiciously chosen waveforms are needed to represent the entire $7d$ space with an accuracy better than 10^{-7} for 200 cycles, mass ratios $q \in [1, 10]$, dimensionless spin magnitudes $\|\vec{\chi}_{1,2}\| \leq 0.9$, and through $\ell = 8$ spherical harmonic modes. Using only the first 10 of these select waveforms yields a maximum representation error $\lesssim 1\%$. The results of this paper suggest that for any given parameter range a remarkably small number of numerical relativity simulations of precessing binary black holes, if judiciously chosen, are sufficient to *accurately represent any other precession waveform in that range*. We expect these results to be useful also for gravitational wave matched-filter searches and parameter estimation studies for compact binary coalescences.

---

[1] Gravitational wave detectors have a finite frequency bandwidth that introduces a total mass scale, thus adding an 8th parameter that we will ignore. We also focus on quasi-circular inspirals.

3.3 Beating the curse of dimensionality

Previous studies have shown that non-precession subspaces of the full 7d waveforms space $\mathcal{W}$ can be represented by linear spaces spanned by a relatively compact set of inspiral [75, 156], ringdown [161], and inspiral-merger-ringdown (IMR) [76] waveforms, which form a *reduced basis* (RB). The reduced basis waveforms are found by training a *greedy algorithm* [162, 163] to learn from a given discretization of $\mathcal{W}$ which are the most relevant waveforms for representing elements of $\mathcal{W}$ with regard to a given error measure. See [75] for more details. The number of RB waveforms for non-precessing inspirals hardly grows from two to four parameter dimensions thereby explicitly beating the curse of dimensionality [156]. Of further interest is that *precession* subspaces of $\mathcal{W}$ carry significant redundancy and are amenable to dimensional reduction as found in [164]. Those results strongly suggest that one may beat the curse of dimensionality in the full 7d waveform space.

In this paper, we outline how to construct a very compact but highly accurate RB of precession waveforms. We consider the following specifications on the 7d waveform space:

$$q \in [1, 10]\,, \quad \|\vec{\chi}_{1,2}\| \in [0, 0.9]\,, \quad 200 \text{ cycles}\,, \tag{3.1}$$

where $q = m_1/m_2 \geq 1$. These were chosen based partially on practical limitations of binary black hole simulations. However, the general message of this paper does not depend on our choice.

3.4 Key ingredients

Our construction of a very compact (or sparse) reduced basis representation of precession waveforms depends on several key steps [2]: (1) a randomized resampling strategy [166] for training the greedy algorithm on the 7d waveform space; (2) a frame that rotates with the binary's precession; and (3) a physically motivated parametrization of precession waveforms.

The first key ingredient is a modification of the standard greedy algorithm [75]. In its simplest inception, the greedy algorithm learns which waveforms can linearly span the space of interest in a nearly optimal way [162, 163], starting from a sufficiently dense set of waveforms called a training set or space. However, the curse

---

[2] The last two ingredients can be viewed as aspects of nonlinear dimensional reduction and manifold learning, which aim to reveal the intrinsic dimensionality of large amounts of data (e.g., see [165]).

of dimensionality prevents us from sampling the waveform space with sufficient coverage to build a reliable training set. To overcome this, we randomly resample the 7d space from a uniform distribution using a fixed number K of waveforms at each iteration of the greedy algorithm. These waveforms constitute the training set at the current iteration. Because the 7d space is resampled at each iteration by different waveforms, the maximum error from projecting waveforms onto the current basis at the $j^{\rm th}$ step is actually measuring this error over an effective training set with $j \times K$ randomly distributed waveforms. As more iterations are made, more of the 7d space is sampled and the more accurate the RB becomes. This is a simple implementation of more powerful techniques introduced in Ref. [166].

For our studies, we randomly and uniformly resampled $K \leq 36{,}000$ waveforms at each iteration of the greedy algorithm. We began our studies with small K and increased each sample size up to $K = 36{,}000$, for which the total number of RB waveforms was robust and independent of K. The largest training set used in our studies included more than 3×10^6 randomly selected waveforms.

The second key ingredient is to work in the binary's precessing frame instead of the usual inertial one. Specifically, we generate post-Newtonian (PN) precession waveforms in the time-domain using the *minimally rotating* frame of Refs. [167, 168]. In this frame, a precession waveform appears qualitatively similar to waveforms from non-spinning binaries in their inertial frame [66, 159, 168–171]. In the minimally rotating frame, waveforms have a weaker dependence on parameters than they do in the inertial frame. The rotation involved in going from the minimally rotating frame to the inertial one and vice versa can be accounted for by any convenient representation of the $SO(3)$ group.

The third key ingredient, and perhaps the most crucial, is that we choose to parametrize precession waveforms by their phase instead of by time or frequency. To motivate this choice we momentarily consider the frequency-domain waveform (in the stationary phase approximation) for a non-spinning binary inspiral at leading order ("0PN") in the PN approximation,

$$h(f;\mathcal{M}) = A\mathcal{M}^{5/6} f^{-7/6} e^{i\Phi_0(f;\mathcal{M})}, \tag{3.2}$$

where $\mathcal{M} = M\nu^{3/5}$ is the chirp mass, M is the total mass, ν is the symmetric mass ratio, A is a constant independent of the binary's intrinsic parameters, and

$$\Phi_0(f;\mathcal{M}) \equiv \frac{3}{128}(\pi\mathcal{M}f)^{-5/3}. \tag{3.3}$$

Reparametrizing (3.2) by its phase, now taken as the independent variable, gives

$$H(\varphi; \mathcal{M}) \equiv h(F(\varphi); \mathcal{M}) = A' \mathcal{M}^2 \varphi^{7/10} e^{i\varphi} \tag{3.4}$$

with $A' = A\pi^{7/6}(128/3)^{7/10}$ and $F(\varphi)$ from solving $\Phi_0(f{=}F) = \varphi$. In this phase-domain, all waveforms are proportional to each other, thus constituting a 1d space. In fact, performing the greedy algorithm analytically (versus numerically) returns a single basis waveform that *exactly* represents all such waveforms in the continuum. This is the intrinsic dimensionality of the problem as has long been known because 0PN waveforms only depend on the chirp mass. Therefore, a single reduced basis waveform spans the whole 0PN waveform space. To close the system, we also need to represent the mapping between the phase and frequency domains,

$$F(\varphi; \mathcal{M}) = \frac{1}{\pi \mathcal{M}} \left(\frac{128\varphi}{3} \right)^{-3/5}, \tag{3.5}$$

using a separate basis. As we see again, the frequencies for different chirp masses are all proportional to each other. Therefore, any 0PN waveform, as a function of frequency, is represented by *one* reduced basis waveform through the non-linear transformation in (3.3).

For the sake of comparison, we implemented a standard greedy algorithm following [75] using 0PN waveforms parametrized by frequency (not phase) for binaries with a fixed total mass and with mass ratios and number of cycles as listed in (3.1). We found that 152 RB waveforms are required to reach numerical round-off errors in representing any waveform in this 1d space. Even to reach an error of about 1% requires 138 RB waveforms. Therefore, using the phase parametrization results in a single RB waveform for *exact* representation whereas a frequency parametrization can yield a much larger RB for *approximate* representation. Part of the reason why using waveforms in the phase domain (or φ-domain) is advantageous is because the oscillations in two waveforms *always* cancel in the scalar product used to measure the projection error onto the RB in the greedy algorithm,

$$\langle H_{\mathcal{M}_1}, H_{\mathcal{M}_2} \rangle_\varphi \equiv \int_{\varphi_{\min}}^{\varphi_{\max}} d\varphi \, H(\varphi; \mathcal{M}_1) H^*(\varphi; \mathcal{M}_2). \tag{3.6}$$

For 0PN waveforms this results in a very smooth dependence on the chirp masses since (3.6) is $\propto \mathcal{M}_1^2 \mathcal{M}_2^2$. Similarly, the waveform frequency as a function of phase (3.5) has a very smooth dependence on them as well.

Higher PN orders include more physics, such as the nonlinearity of general relativity and spin-orbit, spin1-spin2, and self-spin interactions, that depend on all 7 intrinsic parameters. These contributions add more structure to the waveforms but only weakly depend on the parameters. This is especially true in the φ-domain and, as discussed below, we also find this holds through the merger and ringdown phases where the PN expansion parameter is no longer small. Consequently, the inverse function $F(\varphi)$ (or $T(\varphi)$ if in the time domain) retains the weak dependence on intrinsic parameters. As there is thus only ever a weak parameter dependence, one may expect to find only a relatively small number $\mathcal{N}$ of RB waveforms, possibly as few as $\mathcal{N} = O(d)$.

3.5 Method outline

In this paper, we use 3.5PN precessing inspiral waveforms. We solve the PN equations (see Ref. [172] and references therein) using the approach of Refs. [167, 168] where the waveforms themselves are solved in a frame that minimizes the binary's precession, along with a rotation operator represented by unit quaternions to track this frame relative to the fiducial inertial frame. All waveforms in this minimally rotating frame are normalized to unity, and the initial orbital phases are aligned. It is convenient to decompose the waveform into spin-weighted spherical harmonic modes [173] characterized by (ℓ, m). We use the phase associated with the $(\ell, m) = (2, 2)$ mode to parametrize the waveform but other choices are possible. A precession waveform $h(t)$ in the inertial frame is thus decomposed in the following way,

$$h(t) \to (\{H_{\ell m}(\varphi)\}, T(\varphi), Q(\varphi)) , \tag{3.7}$$

where $H_{\ell m}$ is a spin-weighted spherical harmonic mode in the minimally rotating frame, $T(\varphi)$ is the function relating the $(2,2)$ phase to the time coordinate, and Q is the unit quaternion describing the rotation back to the inertial frame. We take into account all modes up to $\ell = 8$ and cut all waveforms off at a dimensionless frequency of 0.2 in the $(2,2)$ modes. Finally, all waveforms contain 200 wave cycles. We build a RB for each component in the decomposition (3.7). It is natural to use the scalar product in (3.6) for the T and Q functions but to integrate the minimally rotating waveform over the 2-sphere so that, upon using the orthogonality of the spin-weighted spherical harmonics,

$$\langle H_{\lambda_1}, H_{\lambda_2} \rangle_\varphi \equiv \sum_{\ell,m} \int_{\varphi_{\min}}^{\varphi_{\max}} d\varphi \, H_{\ell m}(\varphi; \lambda_1) H^*_{\ell m}(\varphi; \lambda_2) , \tag{3.8}$$

where λ_i is a tuple of parameter values. Executing a greedy algorithm on each component in (3.7) would result in a selection of parameter values that are different for each element. In order to choose the same parameters for all three reduced bases, we define a total projection error ϵ_φ through,

$$\epsilon_\varphi(\lambda) \equiv 8 \times 10^{-6} \|\delta T_\lambda\|_\varphi^2 + 0.5\|\delta H_\lambda\|_\varphi^2 + 0.0031\|\delta Q_\lambda\|_\varphi^2 \tag{3.9}$$

so as to receive approximately equal contributions from each component. Here, $\lambda = (q, \vec{\chi}_1, \vec{\chi}_2)$ is a tuple of 7d parameter values, $\delta X_\lambda = X_\lambda - P_X[X_\lambda]$ with X one of the elements in (3.7), and P_X is the projection operator onto the basis for X. The numerical coefficients are fixed to give approximately equal contributions to the mismatch in the time domain and inertial frame in the case of small random perturbations. Binaries with periods near $200M$ lead to a small coefficient for the time function.

3.6 Results for precessing binary inspirals

We implemented a greedy algorithm using the three key ingredients discussed above to find RB representations for the space of precession waveforms for the ranges given in (3.1). The left panel of Fig. 3.1 shows the maximum of the total projection error (3.9) found at each iteration of the greedy algorithm. We observe a power-law decay with exponent ≈ -8. The total error is not monotonically decreasing because of the constant resampling at each iteration. We observe that the maximum normalized projection error over the training set is 10^{-2} using 10 basis waveforms and $\approx 4\times10^{-8}$ for 50. Also shown are the contributions to the total error from the projections onto the basis of each component in (3.7).

To measure the error in the time-domain inertial frame between a waveform h and its RB approximation h_{app} we use the standard *mismatch*

$$\text{Mismatch} = 1 - \text{Re}\,\langle h, h_{\text{app}}\rangle_t\,, \tag{3.10}$$

where for two functions A, B the time-domain complex scalar product is $\langle A, B\rangle_t \equiv \int_{t_{\min}}^{t_{\max}} dt\, A(t)B^*(t)$. In order to measure the quality of the RB approximation itself, we do not extremize the mismatch with respect to the relative phase and time shift between h and h_{app}.

It is not obvious that the basis generated using (3.9) from the minimally rotating frame and φ-domain will be accurate for inertial frame waveforms expressed in the time-domain. Nevertheless, we find that the φ-domain, precessing basis is highly accurate for representing time-domain, inertial-frame waveforms. The right panel in

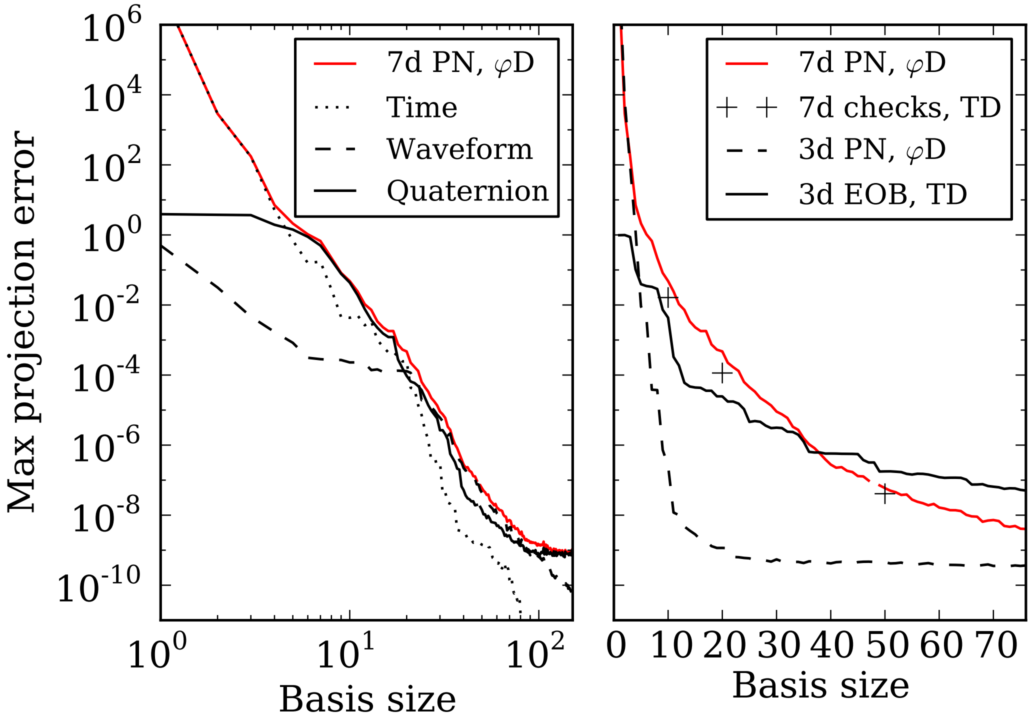

Figure 3.1: **Left**: Maximum φ-domain (φD) projection error (red) from (3.9) for 7d post-Newtonian precession waveforms versus basis size. The contributions from the time function (dotted), waveform in the minimally rotating frame (dashed), and quaternion (solid) are also shown. **Right**: Maximum time-domain, inertial frame mismatches from (3.10) for 10^7 randomly selected waveforms (+) using the first 10, 20, and 50 reduced basis waveforms. Also plotted are φ-domain projection errors for non-precessing PN waveforms (dashed) and the time-domain (TD) projection errors from using the latter parameter values selected by the greedy algorithm to represent EOB waveforms (solid black), which additionally include merger and ringdown phases.

Fig. 3.1 shows the mismatch (+) from using the first 10, 20, and 50 basis functions to represent more than 10^7 randomly chosen waveforms for the same specifications in (3.1). Figure 3.2 shows the distribution count of waveforms with a given error using the first 10, 20, and 50 RB functions. The latter distribution has median 3.5×10^{-9}, mean value 4.2×10^{-9}, and a maximum representation error of 4.1×10^{-8}. Using the first 10 RB functions, the maximum mismatch is 0.016 over more than 10^7 randomly selected waveforms.

Table 3.1 shows that we beat the curse of dimensionality since, for a given error, the number of RB waveforms needed to accurately represent the subspace of $\mathcal{W}$ with the indicated dimension d grows approximately linearly with d, not exponentially.

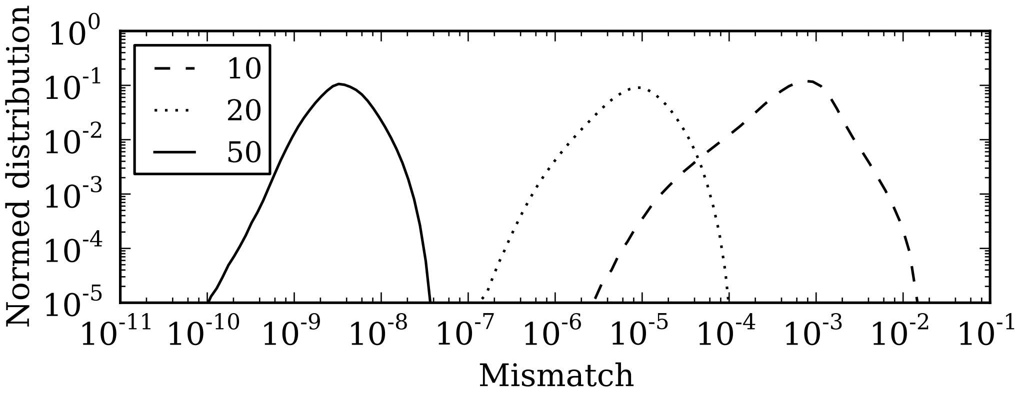

Figure 3.2: Distribution of mismatches from (3.10) for more than 10^7 randomly selected waveforms in the time-domain and inertial frame using the first 10 (dashed), 20 (dotted), and 50 (solid) reduced basis waveforms. Distributions are normalized by their total samples.

Error	Basis size			
	1d	2d	3d	7d
$\lesssim 10^{-2}$	4	6	7	13
$\lesssim 10^{-4}$	4	7	8	20
$\lesssim 3 \times 10^{-8}$	6	15	23	50

Table 3.1: The number of basis waveforms required for a desired maximum mismatch scales approximately *linearly* with the dimension thus beating the curse of dimensionality. The first three dimensions considered are from mass ratio q and z-components of the spin vectors $\vec{\chi}_{1,2}$ with 1d → (q), 2d → (q, χ_{1z}), and 3d → $(q, \chi_{1z}, \chi_{2z})$.

Figure 3.3 shows the first 90 parameters selected by our greedy algorithm and presented according to which component – time, minimally rotating waveform, quaternion – is the dominant contribution to the total representation error in the left panel of Fig. 3.1. The spins' components are taken at the initial time where the inertial and minimally rotating frames are equal. The mass ratios are heavily weighted towards the endpoints of the considered interval in (3.1). Both spins' magnitudes tend to be in $[0.8, 0.9]$. The projections of the spins onto the initial orbital angular momentum seem to be *anti*-correlated, at least when the waveform contribution to (3.9) is dominant. We also see that the x-y components of the spins tend to lie on a circle for the smaller mass m_2 while there is less clear structure for the larger mass m_1.

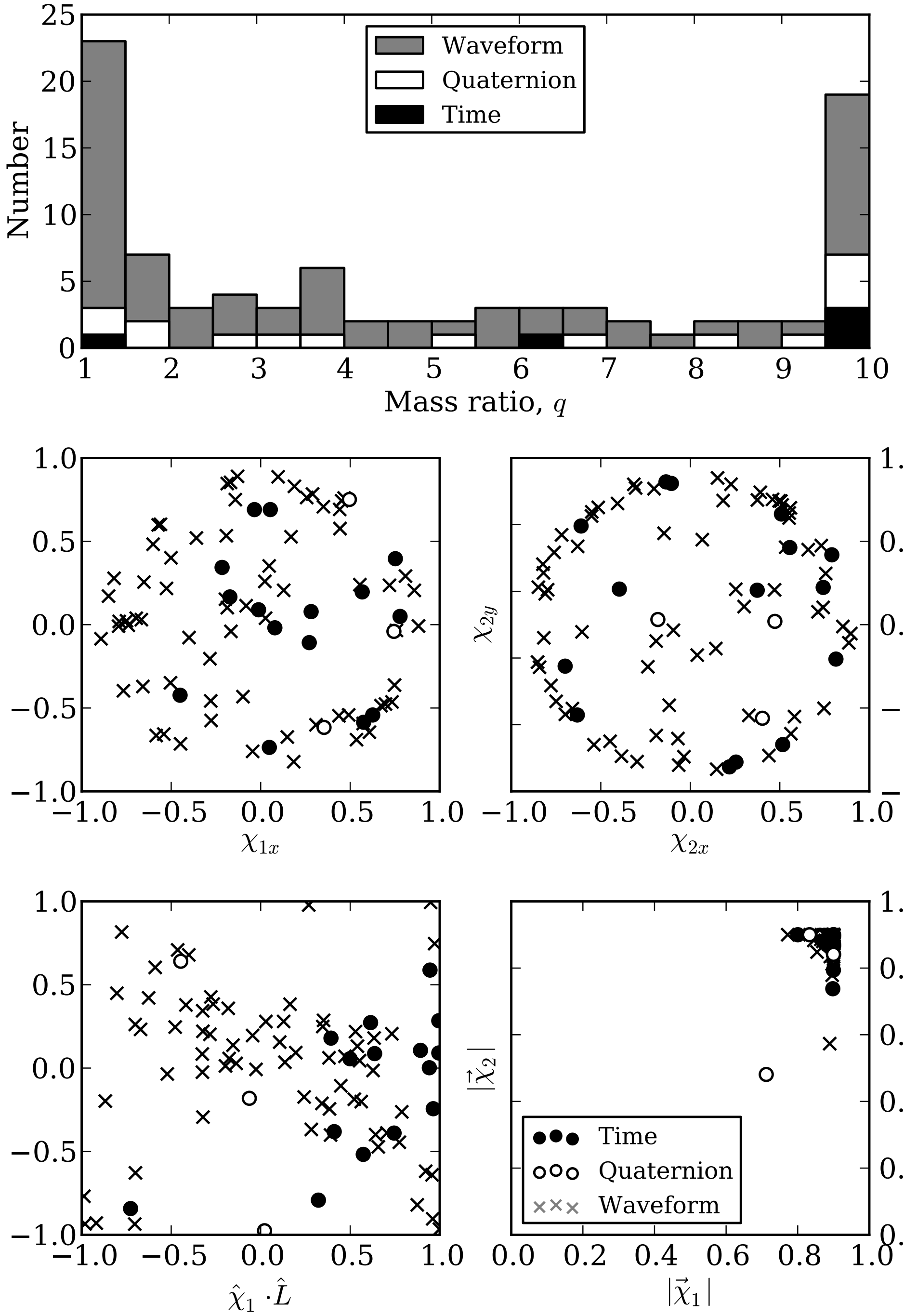

Figure 3.3: Distribution of mass ratios (top), x-y components of both spins (second row), projection of spins onto initial orbital angular momentum unit vector (bottom left), and both spin magnitudes (bottom right) as selected by our greedy algorithm. The spins' components are given in the inertial frame at the initial time.

3.7 From inspiral to coalescence

Next, we test whether the parameters selected from the inspiral regime result in an accurate reduced basis when including merger and ringdown. This issue has immediate relevance for building a RB for expensive numerical relativity simulations of precessing binary black hole mergers that, in turn, has important ramifications for data analysis applications with gravitational wave detectors and for modeling expensive merger simulations with reduced-order/surrogate models [76].

Currently, we can answer the above question for spinning but *non-precessing* binary black hole coalescences, which involves only the three parameters q, $|\vec{\chi}_1|$, and $|\vec{\chi}_2|$, for which an effective-one-body (EOB) semi-analytical model of IMR is available [67, 69]. We first used our greedy algorithm to find the parameters for building a RB for the non-precessing inspiral PN waveforms using the φ-domain error in (3.9). We then generated a basis using the EOB non-precessing coalescence waveforms evaluated at those selected parameters. Lastly, we randomly generated more than $10^{d=3}$ EOB waveforms and computed the time-domain inertial frame mismatch from (3.10). The results of this study are shown as the solid black curve in the right panel of Fig. 3.1. For the first 20 inspiral RB waveforms, the maximum mismatch of the EOB waveforms is about 3×10^{-5} while for the first 50 it is about 2×10^{-7}.

3.8 Outlook

Based on traditional methods to sample the waveform space, which scale exponentially with dimension [60, 174–176], it has been perceived that an intractable number of numerical relativity simulations would be needed to represent the space of binary black holes for any given number of orbits. However, we have found evidence that a remarkably small number of numerical relativity binary black hole simulations may actually be needed, if *judiciously* chosen, to build a high accuracy reduced basis to represent the whole space of interest.

Based on the non-precessing EOB results presented above, performing numerical simulations of binary black hole mergers for the first 50-90 parameters selected by our greedy algorithm may be sufficient to represent the precession waveforms of any other coalescences in the parameter ranges of (3.1). This constitutes less than one tenth of the number of randomly chosen simulations performed over the last few years by the numerical relativity community [171, 177–179]. In addition, this work suggests that an unexpectedly small number of low-mass inspiral waveforms may represent the frequency and parameter ranges of interest to gravitational wave detectors, which may also enable very compact reduced-order quadratures [144, 180] of overlap integrals for fast parameter estimation studies. Finally, this work opens the door for building surrogate models [76] of precessing inspiral waveforms that can be useful for multiple query applications in place of solving a large number of parametrized ordinary differential equations.

3.9 Acknowledgements

We thank Scott Field and Rory Smith for comments and discussions. We thank Michael Boyle for permission to use his Triton code for solving precessing PN waveforms, now replaced by his public open source code available from [181]. We have also used the LAL (LSC Algorithm Library) EOB code version 6.11.0.1 [182]. This project was supported in part by the Fairchild Foundation, NSF grants PHY-1068881, CAREER PHY-0956189, and PHY-1005655 to Caltech, NASA grant NNX10AC69G, and NSF grants PHY-1208861, PHY-1316424 and PHY-1005632 to the University of Maryland. Computations were performed on the Zwicky cluster at Caltech, which is supported by the Sherman Fairchild Foundation and by NSF award PHY-0960291.

4 Numerical relativity reaching into post-Newtonian territory: a compact-object binary simulation spanning 350 gravitational-wave cycles

4.1 Background and Contributions

This chapter was published in Physical Review Letters on July 16 2015 [16], in issue 3 of volume 115, which can be found at `https://link.aps.org/doi/10.1103/PhysRevLett.115.031102`. A preprint can be found at `https://arxiv.org/abs/1401.7038`.

This chapter presents a numerical relativity (NR) simulation of a binary black hole merger which begins 176 orbits before merger, performed with the Spectral Einstein Code (SpEC). The duration of the simulation is nearly an order of magnitude longer than other simulations that had been performed at the time of publication, and it begins in the early inspiral where post-Newtonian approximations are accurate. The unprecedented length of the simulation exposed an instability in the outer boundary conditions in SpEC, which leads to a slow unphysical exponential acceleration of the center of mass towards the outer boundary. The timescale of this unwanted drift was found to depend on the outer boundary radius of the simulation, such that it can be made negligibly small by using a sufficiently large computational domain. After hybridizing with a shorter NR simulation of the same physical system to fix the merger which was corrupted by this drift, various waveform models were tested across the whole frequency band of gravitational wave detectors.

I performed the long simulation and helped debug the center of mass drift. I determined the exponential behavior of the drift, and performed the additional simulations used to measure the dependence of the drift rate on the outer boundary radius of the simulation. I also reproduced the mismatches calculated by Andrea Taracchini to double check their validity, and assisted in editing the manuscript.

Numerical relativity reaching into post-Newtonian territory: a compact-object binary simulation spanning 350 gravitational-wave cycles

Béla Szilágyi,[1,2] Jonathan Blackman,[1] Alessandra Buonanno,[3,4] Andrea Taracchini,[3] Harald P. Pfeiffer,[5,6] Mark A. Scheel,[1] Tony Chu[7,5] Lawrence E. Kidder,[8] and Yi Pan[4]

[1] *TAPIR, Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, CA 91125, USA*

[2] *Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91106, USA*

[3] *Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, Potsdam-Golm, 14476, Germany*

[4] *Department of Physics, University of Maryland, College Park, MD 20742, USA*

[5] *Canadian Institute for Theoretical Astrophysics, 60 St. George Street, University of Toronto, Toronto, ON M5S 3H8, Canada*

[6] *Canadian Institute for Advanced Research, 180 Dundas St. West, Toronto, ON M5G 1Z8, Canada*

[7] *Department of Physics, Princeton University, Jadwin Hall, Princeton, NJ 08544, USA*

[8] *Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853, USA*

We present the *first* numerical-relativity simulation of a compact-object binary whose gravitational waveform is long enough to cover the *entire* frequency band of advanced gravitational-wave detectors, such as LIGO, Virgo, and KAGRA, for mass ratio 7 and total mass as low as $45.5\,M_{\odot}$. We find that effective-one-body models, either *uncalibrated* or *calibrated* against substantially shorter numerical-relativity waveforms at smaller mass ratios, reproduce our new waveform remarkably well, with a negligible loss in detection rate due to modeling error. In contrast, post-Newtonian inspiral waveforms and existing calibrated phenomenological inspiral-merger-ringdown waveforms display greater disagreement with our new simulation. The disagreement varies substantially depending on the specific post-Newtonian approximant used.

4.2 Introduction

The upgraded ground-based interferometric gravitational-wave (GW) detectors LIGO [31, 32] and Virgo [183] will begin scientific observations in mid 2015, and are expected to reach design sensitivity by 2019 [184]. Furthermore, a new Japanese detector, KAGRA [35], is under construction. Direct detection of GWs by the end of this decade is therefore very likely. The most promising GW sources are compact-object binaries, wherein each partner is either a stellar-mass black hole (BH) or a neutron star (NS) [185]. The detection of GWs from compact-object binaries, as well as the determination of source properties from detected GW signals, relies on the accurate knowledge of the expected gravitational waveforms via matched-filtering [186] and Bayesian methods [187].

The need for accurate waveforms has motivated intense research. Early waveform models, based on the post-Newtonian (PN) formalism [188], were limited to the early inspiral. The effective-one-body (EOB) formalism [189, 190] extended waveform models to the late inspiral, merger and ringdown. Since 2005 research has greatly benefited from numerical-relativity (NR) simulations [191–193][1]. Current inspiral-merger-ringdown (IMR) waveform models [70, 199–201] combine information from analytical-relativity (AR) calculations (best suited for the inspiral, when comparable-mass binaries have characteristic velocities smaller than the speed of light) and direct NR simulations (the best means to explore the late inspiral and the merger).

However, there is a gap between the portion of the binary evolution that is described by analytical methods, and the portion that is accessible by NR. For example, waveforms computed at the currently-known PN order become unreliable possibly hundreds of orbits before merger for unequal-mass binaries [202, 203], and even earlier when one of the objects is spinning [150] [2] yet NR simulations have been able to cover only tens of orbits [56, 178, 204] until now. This gap has emerged as one of the most important sources of uncertainty in present IMR waveform models. It is possible to construct IMR models by extending analytical waveforms across the gap [70, 199–201], in some cases obtaining IMR models that are faithful to longer numerical waveforms when extrapolated beyond their limited range of

---

[1] Besides its importance for GW astronomy, NR has also deepened the understanding of general relativity in topics such as binary BH recoil [194, 195], gravitational self force [196], high-energy physics, and cosmology [197, 198].

[2] Several PN waveforms (or approximants) with different PN-truncation error are available in the literature. These PN approximants can differ from each other during the last hundreds of cycles before merger.

Initial Data				
D_0/M	$10^3 M\Omega_0$	$10^6 \dot{a}_0 M$	$E_{\rm ADM}/M$	$J_{\rm ADM}/M^2$
12.2	21.1541	−47.99	0.996211	0.4510
27	6.7930	0	0.998112	0.6123

Inspiral				Remnant properties	
m_1/m_2	$10^5\epsilon$	T/M	N	M_f/M	S_f/M_f^2
6.99997(2)	< 6	4,100	20	0.98771(1)	0.32830(3)
7.00000(1)	34	106,000	176	0.98762(14)	0.32827(2)

Table 4.1: Properties of the two NR simulations: The first block lists initial separation D_0, orbital frequency Ω_0, radial velocity $\dot{a}_0$, ADM energy E_{ADM} and angular momentum J_{ADM} in units of total mass $M = m_1 + m_2$. The middle block lists mass ratio m_1/m_2, eccentricity ϵ, time duration T and number of orbits N until merger. The final block lists remnant mass M_f and spin S_f.

calibration [205]. However, so far these procedures have been tested using NR simulations with only 30 orbits, too few to close the gap.

The time duration T of an inspiral waveform starting at initial GW frequency f_{ini} scales as $T \propto \nu^{-1} f_{\rm ini}^{-8/3}$, where $\nu = m_1 m_2/M^2$ is the symmetric mass ratio of the binary with component masses $m_{1,2}$ and total mass $M = m_1 + m_2$. Therefore, reducing f_{ini} by a factor of 2 increases T sevenfold. Halving the symmetric mass ratio ν (e.g., from $m_1/m_2=2$ to $m_1/m_2=7$) doubles T. Increasing the simulation length T is difficult: it becomes harder to preserve phase coherency, the outer boundary of a simulation is in causal contact for a larger fraction of the simulation, and existing codes would require many months or even years of wall-clock time. Therefore progress toward longer simulations has been sluggish, with T increasing by only about a factor of 2 to 3 during the last five years [177, 178, 204, 206, 207]. The duration T needed to close the gap depends on the binary parameters and the detector bandwidth. Here we start addressing the issue of the gap by focusing on the nonspinning case and high mass ratio, $q = m_1/m_2 = 7$, for which the PN approximants can greatly differ [203, 208]. We present a new NR simulation that extends T by a factor of 20 and reduces the initial frequency f_{ini} by a factor of 3. With its comparatively high mass ratio, the new simulation probes an astrophysically relevant parameter regime for BH-BH and NS-BH binaries and for certain total masses covers the entire frequency band of advanced LIGO (aLIGO) and Virgo. We describe challenges involved in carrying out this new simulation, most notably an unexpected motion of the binary center of mass (CoM), and we suggest improvements for future long simulations. We then compare the new simulation with existing analytical waveform models to assess the

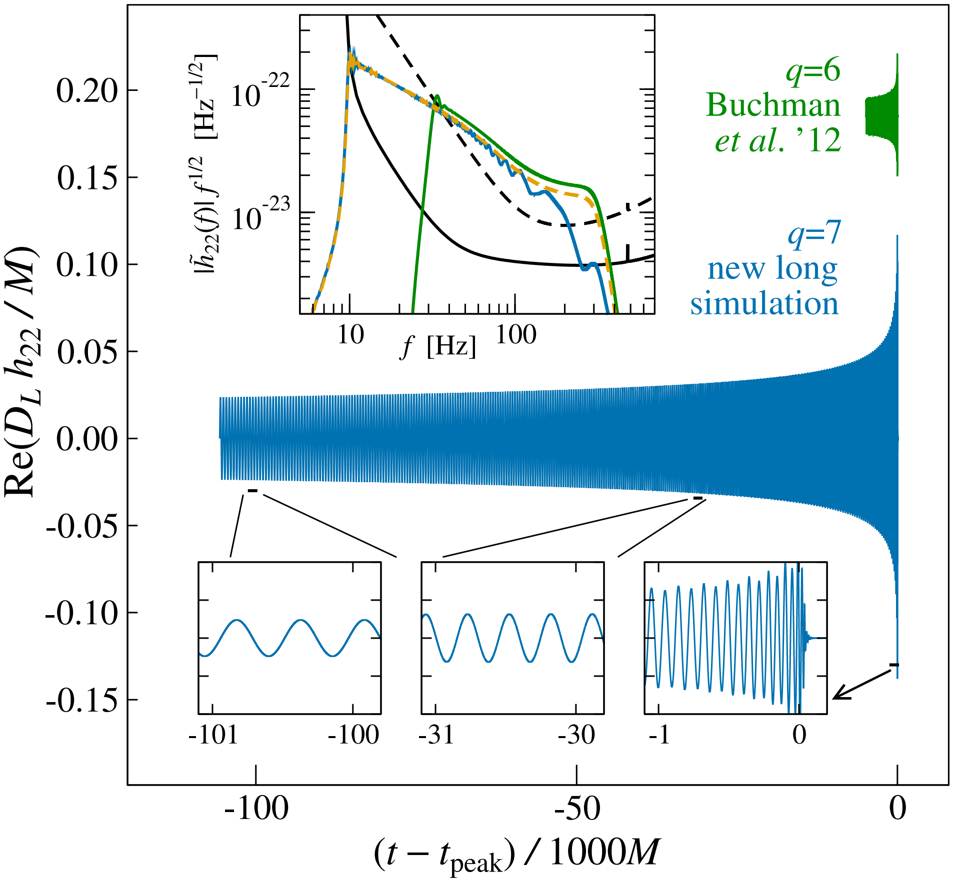

Figure 4.1: Overview of the new very long simulation. The main panel shows the $(2,2)$ spherical-harmonic mode of the GW strain, with enlargements in the lower insets. The top inset shows the Fourier spectra of the new waveform in blue and the NR-NR hybrid waveform (used for comparisons with analytical models) in yellow, overlaid with noise power spectral densities of aLIGO at the early (dashed black) and design (solid black) sensitivity [184]. The waveforms in the inset are scaled to total mass $M = 45.5\,M_\odot$ and luminosity distance $D_L \approx 1.06\,$Gpc. For comparison, an older $q = 6$ waveform [56] of representative length is shown in the main panel (offset vertically for clarity) and in the power-spectrum inset.

impact of waveform model errors on the detection rate of advanced detectors.

4.3 Numerical-relativity waveforms

We report on two new simulations of a nonspinning BH binary with mass ratio $q = m_1/m_2 = 7$. The *short* simulation is of typical length: 20 orbits, $T = 4{,}100\,M$. The *long* simulation, the main focus of this paper, is about 25 times longer. Both simulations are computed using the Spectral Einstein Code (SpEC) [209]. The short simulation uses established computational techniques [204]. The speed-up needed for the long simulation is the result of a series of code changes including task-based adaptive parallel load-balancing, live timing-based selection of the most efficient algorithm (when multiple implementations of the same function are available), a modified memory layout to allow more efficient calls to low-level numerical packages and a more efficient implementation of the Generalized Harmonic evolution equations. The highest resolution of the long simulation required 265 days on 48 cores. Figure 4.1 shows the new long waveform and Table 4.1 presents additional details about both simulations. Geometrized units $G = c = 1$ are used in Table 4.1 and throughout this paper. The top inset of Fig. 4.1 shows the spectra of the $(2,2)$ spherical harmonic waveform modes. The long simulation covers the entire design-aLIGO frequency range for nonspinning BH-BH binaries with $M \gtrsim 45\,M_\odot$, and covers the early-aLIGO frequency range for $M \gtrsim 11\,M_\odot$, including nonspinning NS-BH binaries[3]. In contrast, the $q = 6$ simulation plotted in green, which is representative of past simulations, starts at 3 times higher frequency, and covers a much smaller portion of the aLIGO bandwidth for a given M. Thus, we present here the first gravitational waveform covering the entire design-aLIGO frequency band for a nonspinning, compact-object binary at mass ratio $q = 7$ with a total mass as low as $M = 45.5\,M_\odot$.

The short simulation is run at three different numerical resolutions, and the long one at four resolutions. The long simulation employs dynamical spectral adaptive mesh refinement [211], so measured quantities (like BH masses or waveforms) do not always converge in a regular, predictable manner with increasing resolution, as is the case when each resolution is defined by a fixed number of grid points. Furthermore, failure to resolve initial transients caused by imperfect initial data also complicates convergence (see discussion in Sec. IIIB of Ref. [212]). Nevertheless, we find that differences in measured quantities like waveforms, masses, and spins decrease rapidly with resolution, and Table 4.1 displays a conservative error estimate obtained by taking the difference between the two highest resolutions. After the initial transients have decayed, we measure the mass ratio to be equal to 7 to within five significant digits, and the dimensionless spins to be $\lesssim 10^{-6}$. The remnant mass and spin agree to within four significant digits between the short and long simulations.

However, the long simulation encounters an unexpected problem. After a few

---

[3] For mass ratio 7, in absence of spin, we expect no observable differences in the merger signal between a BH-BH and a NS-BH binary [210].

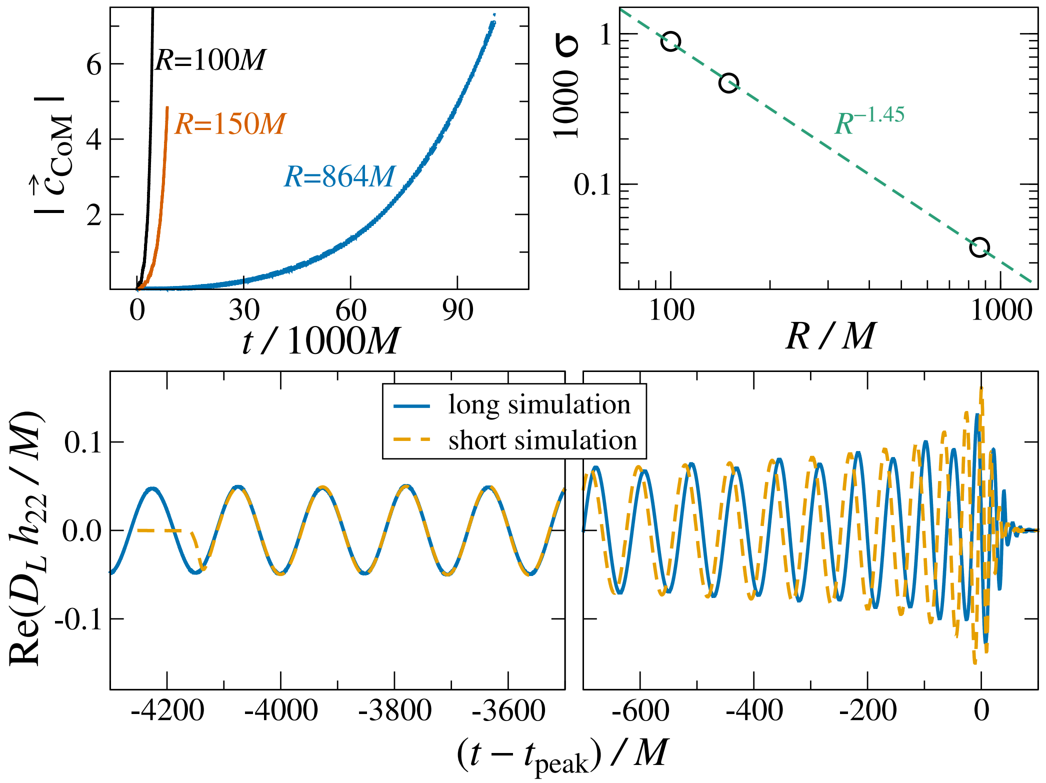

Figure 4.2: **T**op left: displacement of the CoM from the origin, for three long simulations with different outer boundary radii R. In each case, $|\vec{c}_{CoM}|$ increases exponentially. **T**op right: growth rate σ of $|\vec{c}_{CoM}|$ as a function of R with a power-law fit. **L**ower panels: GW (2,2) mode of the short and long simulations. The long simulation still agrees very well with the short one at early times, but fails to produce an accurate merger waveform.

$10{,}000\,M$ of evolution, the coordinate CoM begins to drift away from the origin. Define the CoM as $\vec{c}_{CoM} = \vec{c}_1 m_1/M + \vec{c}_2 m_2/M$, where $\vec{c}_{1,2}$ are the coordinate centers of the apparent horizons of the two BHs. Fig. 4.2 shows that $|\vec{c}_{CoM}|$ increases *exponentially* with time. This drift is primarily a coordinate effect that only marginally affects most measurable quantities. For example, the linear momentum radiated to infinity, as computed from the waveform obtained by Cauchy-Characteristic extraction [213], is consistent with PN theory, and is too small to account for the motion of the CoM. To explore the drift in more detail, we repeat the long simulation with different values of R, the coordinate radius of the artificial outer boundary where we impose an outgoing-wave boundary condition. The top left panel of Fig. 4.2 shows the CoM drift for several values of R. Empirically, we find that the exponential growth rate σ behaves like $\sigma \propto R^{-1.45}$, c.f. Fig. 4.2. For our standard choice of $R = 864\,M$, $1/\sigma = 26{,}000\,M$; this large timescale explains why the drift was not noticed in earlier, shorter simulations.

We conjecture that the drift is caused by a coupling with the outer boundary. Such a coupling might arise through enhanced reflections of the outgoing GW at the outer boundary. Our outgoing-wave boundary conditions [214] have the smallest reflection coefficient for spherical harmonic modes with small ℓ, and a reflection coefficient of order unity when $kR/\ell \gtrsim 1$ [214], where k is the radial wavevector. With increasing $|\vec{c}_{\mathrm{Co}M}|$ the emitted GW will have increasing high-ℓ content when decomposed on the outer boundary.

From the bottom panel in Fig. 4.2, we see that the effects of the drift on the long waveform is confined to the last ~ 10 orbits before merger. In the Fourier domain (see Fig. 4.1) one can clearly see the unusual behavior of the long waveform at high frequencies. Since these orbits are covered by the short waveform, we hybridize the short waveform with the long one, thus replacing the final portion of the former. We adopt the hybridization method of Ref. [215]. We construct 9 NR-NR hybrids by combining three versions of the long simulation and three versions of the short simulation. Each version may differ by the numerical resolution (which we label by "Lev$\mathcal{N}$", where the adaptive mesh refinement aims for truncation errors $\propto e^{-\mathcal{N}}$), or by the degree of the polynomial used to extrapolate the waveform to infinity [213, 216] (which we label by "N$\mathcal{M}$", where $\mathcal{M}$ is the polynomial degree.) For example, the Lev3 simulation uses $\sim 90^3$ collocation points near merger. In particular, we use (Lev3, N3), (Lev3, N2), and (Lev2, N3) for the long simulation, while we use (Lev5, N3), (Lev5, N2), and (Lev4, N3) for the short simulation. The fiducial NR-NR hybrid is built from the long (Lev3, N3) and the short (Lev5, N3) simulations; this pair of waveforms is blended over the interval $t - t_{peak} \in [-3252, -2252]M$. In the top panel of Fig. 4.1, we show in yellow the spectrum of the fiducial NR-NR hybrid. This spectrum behaves as expected close to merger, and is devoid of oscillations, just like the spectrum of the $q = 6$ simulation. Since we cannot estimate the impact of the coordinate drift on the phase error of the long waveform, we cannot make statements about the phase disagreement between the long waveform and analytical waveform models. Nevertheless, we can compare the analytical models to the hybrid NR-NR waveform, and investigate how the results change when we vary the blending window where the hybridization is done.

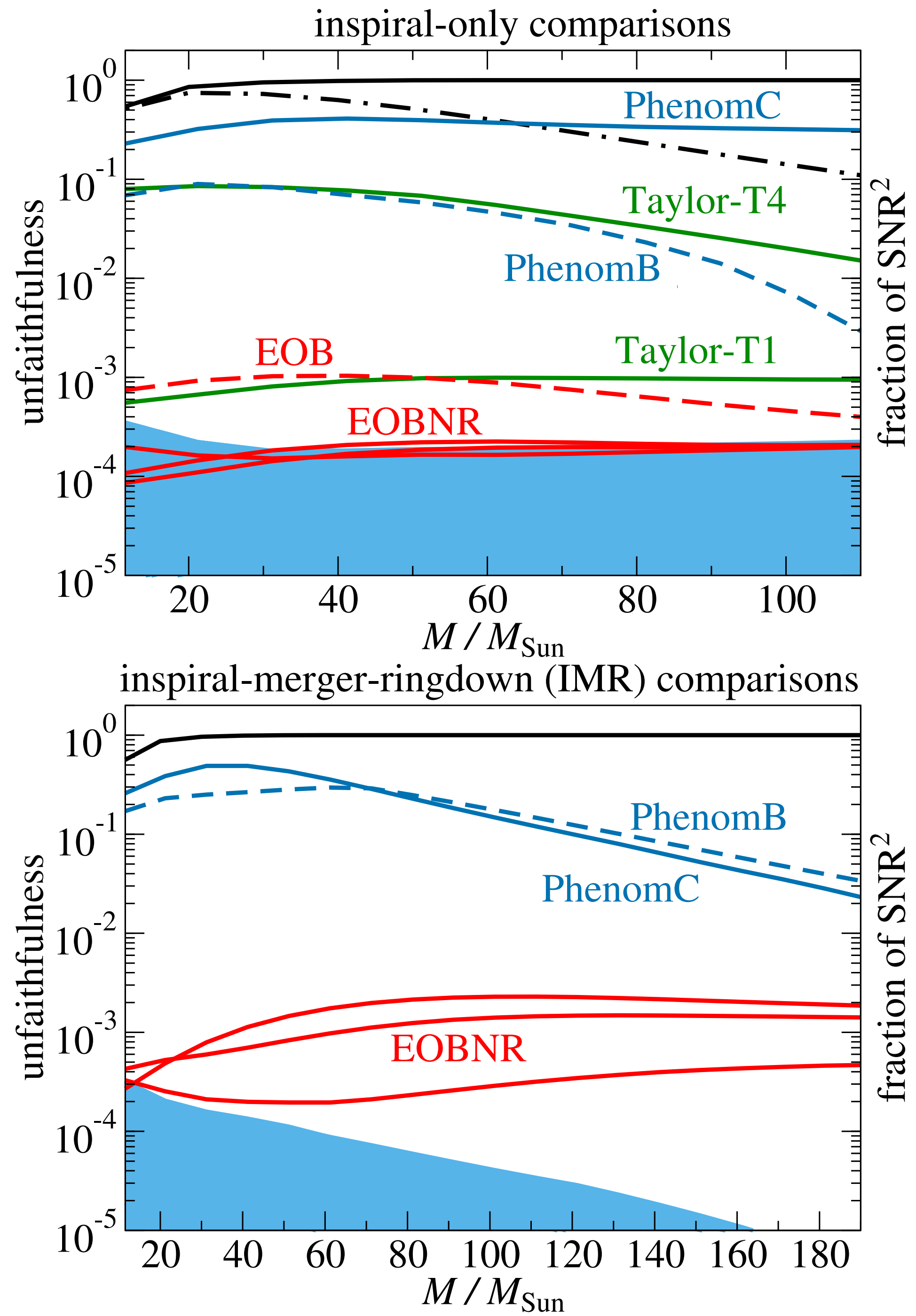

Figure 4.3: Unfaithfulness of the hybrid NR-NR waveform against analytical models. **Left:** inspiral-only comparisons. **Right:** IMR comparisons. Also shown in the left panel are $\mathrm{SNR}^2_{insp}/\mathrm{SNR}^2_{full-insp}$ (solid black line) and $\mathrm{SNR}^2_{insp}/\mathrm{SNR}^2_{full-IMR}$ (dot-dashed black line), and in the right panel $\mathrm{SNR}^2_{IMR}/\mathrm{SNR}^2_{full-IMR}$ (solid black line). The blue area indicates the NR error.

4.4 Comparison to analytical-relativity waveforms

Figure 4.3 summarizes our comparisons between various analytical waveforms h_{22}^{AR} with the hybrid NR-NR waveform h_{22}^{NR}. Shown is the unfaithfulness $\bar{\mathcal{F}}$, defined as $\bar{\mathcal{F}} \equiv 1 - \max_{t_0,\phi_0} \langle h_{22}^{\mathrm{AR}}, h_{22}^{\mathrm{NR}} \rangle / ||h_{22}^{\mathrm{AR}}|| / ||h_{22}^{\mathrm{NR}}||$. Here t_0 and ϕ_0 are the initial time and phase, $||h|| \equiv \sqrt{\langle h, h \rangle}$, and $\langle h_1, h_2 \rangle \equiv 4\,\mathrm{R}e \int_{f_{\mathrm{m}in}}^{f_{\mathrm{m}ax}} \tilde{h}_1(f)\tilde{h}_2^*(f)/S_n(f)\,\mathrm{d}f$, where $S_n(f)$ is the zero-detuned, high-power noise power spectral density of aLIGO [217]. We consider the following analytical waveform models from the LIGO Algorithm Library (LAL): the inspiral-only PN Taylor approximants [218] in the time domain (Taylor-T1, T2, T4) and in the frequency domain (Taylor-F2), an inspiral EOB model (obtained from Ref. [70] by dropping any NR information, thus uncalibrated), the IMR EOBNR models that were obtained by calibrating the EOB model to NR simulations [68, 70, 219] (denoted in LAL as EOBNRv2, SEOBNRv1, and SEOBNRv2), and the IMR phenomenological models that were built combining PN and NR results [199, 200] (denoted in LAL as PhenomB and PhenomC). All the time-domain IMR waveforms are tapered using a Planck windowing function [220], both at the beginning and at the end. We generate the model waveforms from an initial GW frequency of $M\omega_{22} = 0.01317$, and use $f_{\mathrm{m}in} = 12\,\mathrm{Hz}\,(45.5\,M_\odot/M)$. For inspiral-only models, we set $f_{\mathrm{m}ax} = 52.5\,\mathrm{Hz}\,(45.5\,M_\odot/M)$, the minimum available final GW frequency among the time-domain Taylor models, a value close to the innermost-circular-orbit value in Schwarzschild spacetime ($\approx 48.3\,\mathrm{Hz}\,(45.5\,M_\odot/M)$), whereas for the IMR comparisons $f_{\mathrm{m}ax} = \infty$. Quite interestingly, the inspiral-only comparisons give similar results when employing directly the long simulation instead of the NR-NR hybrid.

The blue shaded area in Fig. 4.3 represents the uncertainty in the NR waveforms, estimated by computing $\bar{\mathcal{F}}$ between the fiducial hybrid NR-NR waveform and the other 8 NR-NR hybrids. Because the inspiral-only and IMR curves are calculated using different portions of the hybrid NR-NR waveform, the same model may have different values in the two panels for the same total mass. We vary the prescriptions used for the hybridization (namely, position and width of the blending window), and we find changes $O(10^{-4})$ in the unfaithfulness curves for low total masses. Thus, we consider our results robust. If general relativity correctly describes the GW signals found in nature, then the unfaithfulness $\bar{\mathcal{F}}$ plotted in Fig. 4.3 yields a bound on the loss in detection rate due to modeling error. For sources uniformly distributed in space, the relative loss in detection rate is $\sim 3(d_{MM} + d_{\mathrm{E}})$ (see Sec. VB in Ref. [218]), where d_{MM} is the minimal match of the template bank and d_{E}=1 −

$\max_{\vec{\lambda}} \langle h_{22}^{AR}, h_{22}^{NR} \rangle / ||h_{22}^{AR}|| / ||h_{22}^{NR}||$ is the ineffectualness. Here $\vec{\lambda}$ describes *all* the binary parameters, not just ϕ_0 and t_0, and therefore $d_{\rm E} \leq \bar{\mathcal{F}}$. Typically, $d_{MM} = 3\%$ in LIGO searches. Thus, to achieve $\lesssim 10\%$ loss in detection rate, it suffices that $\bar{\mathcal{F}} \lesssim 1\%$ [218].

Quite remarkably, we find that the unfaithfulness of the uncalibrated *inspiral* EOB waveform is $< 0.1\%$, with a negligible loss in detection rate due to modeling error. The agreement is of course better for the inspiral EOBNR waveforms (i.e., EOBNRv2, SEOBNRv1, SEOBNRv2) ($\bar{\mathcal{F}} < 0.02\%$, left panel), and $\bar{\mathcal{F}} < 0.2\%$ for the IMR EOBNR waveforms (right panel). The closeness of all inspiral EOBNR waveforms strongly suggests that the different calibrations and variations in the dynamics and energy fluxes of those EOBNR models [68, 70, 219] do not impact the low-frequency part of the waveforms, but affect (in a minor way) only the last stages of the inspiral and the merger. The unfaithfulness of the time- and frequency-domain inspiral-only PN Taylor approximants varies between 0.1% and 10% depending on the binary's total mass and the PN approximant used [4]. In particular, Taylor-T4, which has the best agreement with NR in the equal-mass case [207], has the largest disagreement with the new long $q = 7$ NR waveform. (The approximants Taylor-T2 and Taylor-F2 are not displayed, but lie between Taylor-T1 and Taylor-T4). The PhenomB and C models were fitted to hybrids built with Taylor PN approximants and less accurate, short NR waveforms, which may in part explain the large disagreement we find.

The new long NR waveform covers the entire design-aLIGO frequency band only for total mass $M \geq 45.5 M_{\odot}$; for smaller M, the unfaithfulness calculations in Fig. 4.3 neglect the lowest frequency portion of the waveform visible to aLIGO, down to $\sim$ 10 Hz. To understand the significance of the missing GW cycles in the low-frequency portion of the bandwidth, we compute the signal-to-noise ratio (SNR) accumulated within the frequency range $f_{\rm min} \leq f \leq f_{\rm max}$ (SNR_{insp} and SNR_{IMR} for the left and right panel of Fig. 4.3, respectively), and compare it with the SNR accumulated over the *entire* inspiral ($\mathrm{SNR}_{full-insp}$) and the *entire* IMR ($\mathrm{SNR}_{full-IMR}$), i.e., down to 10 Hz. To cover the entire design-aLIGO bandwidth we use the calibrated EOB model of Ref. [70]. Suitable (squared) ratios of these quantities, which represent the fraction of total SNR that is accessible to our comparisons, are plotted in Fig. 4.3. These ratios are < 1 whenever GW cycles are missing in the range $10\,\mathrm{Hz} \leq f \leq f_{\rm min}$ or, in the case of $\mathrm{SNR}^2_{insp}/\mathrm{SNR}^2_{full-IMR}$, also when the merger-ringdown signal is in

---

[4] The large unfaithfulness of some of the PN Taylor approximants is due to differences in the evolution of the frequency and its first time derivative during the late inspiral phase.

band. We find that, even for total masses $< 45.5\, M_\odot$, the unfaithfulness can still be a meaningful assessment of the quality of the analytical models, since a large fraction of SNR is accumulated. Because the merger-ringdown portion of the waveforms becomes increasingly important at higher masses, the inspiral-only comparisons at high mass cover only a small fraction of the entire SNR, as illustrated by the steep decline of $\mathrm{SNR}^2_{insp}/\mathrm{SNR}^2_{\mathrm{full}-IMR}$ in the left panel of Fig. 4.3.

4.5 Conclusions.

To detect and extract unique, astrophysical information from coalescing compact-object binaries, GW instruments employ model waveforms built by combining analytical and numerical-relativity predictions [70, 199–201]. Currently, the main uncertainty of those waveform models is caused by the gap between the regimes of applicability of those methods. This uncertainty can be addressed and, eventually, solved by running much longer NR simulations. In this work we start to tackle this issue by producing a BH-BH simulation 20 times longer than previous simulations. Because of an unexpected drift of the CoM during the last 40 GW cycles, we construct the full NR inspiral, merger and ringdown waveform by hybridizing the long NR waveform with a new, short NR simulation. The hybrid NR-NR waveform covers the entire band of advanced GW detectors for total mass $\geq 45.5\, M_\odot$. Comparing to analytical waveform models, we find strong evidence that — at least for non-spinning binaries — the EOB formalism accurately describes the inspiral dynamics in the so-far unexplored regime of 20 to 176 orbits before merger, and combined with previous work [205] provides accurate waveforms beyond the limited range of calibration. Quite remarkably, the excellent agreement of EOBNR waveforms holds also for uncalibrated inspiral EOB waveforms. PN-approximants have larger errors and more importantly the errors vary substantially depending on the specific PN approximant used.

4.6 Acknowledgments.

We thank Alejandro Bohé for useful discussions. A.B. acknowledges partial support from NSF Grant No. PHY-1208881 and NASA Grant NNX12AN10G. T.C. and H.P. gratefully acknowledge support from NSERC of Canada, the Canada Chairs Program, and the Canadian Institute for Advanced Research. L.K. gratefully acknowledges support from the Sherman Fairchild Foundation, and from NSF grants

PHY-1306125 and AST-1333129 at Cornell. J.B. gratefully acknowledges support from NSERC of Canada. M.S., B.Sz., and J.B. acknowledge support from the Sherman Fairchild Foundation and from NSF grants PHY-1440083 and AST-1333520 at Caltech. Simulations used in this work were computed with the `SpEC` code [209]. Computations were performed on the Zwicky cluster at Caltech, which is supported by the Sherman Fairchild Foundation and by NSF award PHY-0960291; on the NSF XSEDE network under grant TG-PHY990007N; on the Orca cluster supported by Cal State Fullerton; and on the GPC supercomputer at the SciNet HPC Consortium [221]. SciNet is funded by the Canada Foundation for Innovation under the auspices of Compute Canada, the Government of Ontario, Ontario Research Fund–Research Excellence, and the University of Toronto.

5 Fast and accurate prediction of numerical relativity waveforms from binary black hole mergers using surrogate models

5.1 Background and Contributions

This chapter was published in Physical Review Letters on September 18 2014 [269], in issue 12 of volume 115 which can be found at `https://link.aps.org/doi/10.1103/PhysRevLett.115.121102`. A preprint can be found at `https://arxiv.org/abs/1502.07758`.

This chapter presents the first surrogate model of numerical relativity (NR) waveforms. The surrogate model covers the $1d$ parameter subspace of non-spinning black holes, with mass ratios up to 10. While black hole spins have a significant impact on the gravitational waveforms, making this model unsuitable for a full parameter estimation of gravitational wave detections, it addresses some challenges of applying surrogate modeling to NR waveforms. The surrogate model presented is also as accurate as the underlying NR waveforms, demonstrating the ability of NR surrogate models to be both very fast and very accurate.

I performed most of the 22 NR simulations used to build the model using the Spectral Einstein Code (SpEC). I implemented and built the model that ended up being used in the paper, although Scott Field and Chad Galley also built models with similar accuracy and we collaborated to determine the best methods. I performed the error calculations and generated the corresponding figures, and assisted in writing and editing the manuscript.

Fast and accurate prediction of numerical relativity waveforms from binary black hole mergers using surrogate models

Jonathan Blackman,[1] Scott E. Field,[2] Chad R. Galley,[1] Béla Szilágyi,[1]
Mark A. Scheel,[1] Manuel Tiglio,[3] and Daniel A. Hemberger[1]

[1] *TAPIR, Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, CA 91125, USA*
[2] *Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853, USA*
[3] *Center for Astrophysics and Space Sciences, Center for Computational Mathematics, San Diego Supercomputer Center, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0424, USA*

Simulating a binary black hole coalescence by solving Einstein's equations is computationally expensive, requiring days to months of supercomputing time. In this paper, we construct an accurate and fast-to-evaluate surrogate model for numerical relativity (NR) waveforms from non-spinning binary black hole coalescences with mass ratios from 1 to 10 and durations corresponding to about 15 orbits before merger. Our surrogate, which is built using reduced order modeling techniques, is distinct from traditional modeling efforts. We find that the full multi-mode surrogate model agrees with waveforms generated by NR to within the numerical error of the NR code. In particular, we show that our modeling strategy produces surrogates which can correctly predict NR waveforms that were *not* used for the surrogate's training. For all practical purposes, then, the surrogate waveform model is equivalent to the high-accuracy, large-scale simulation waveform but can be evaluated in a millisecond to a second depending on the number of output modes and the sampling rate. Our model includes all spherical-harmonic ${}_{-2}Y_{\ell m}$ waveform modes that can be resolved by the NR code up to $\ell = 8$, including modes that are typically difficult to model with other approaches. We assess the model's uncertainty, which could be useful in parameter estimation studies seeking to incorporate model error. We anticipate NR surrogate models to be useful for rapid NR waveform generation in multiple-query applications like parameter estimation, template bank construction, and testing the fidelity of other waveform models.

Since the breakthroughs of 2005 [191–193], tremendous progress in numerical relativity (NR) has led to hundreds of simulations of binary black hole (BBH) coalescences [58, 171, 177, 178, 204, 206, 222]. This progress has been driven partly by data analysis needs of advanced ground-based gravitational wave detectors like LIGO [223] and Virgo [224]. Recent upgrades to these detectors are expected to yield the first direct detections of gravitational waves (GWs) from compact binary coalescences [225].

Despite the remarkable progress of the NR community, a single high-quality simulation typically requires days to months of supercomputing time. This high computational cost makes it difficult to directly use NR waveforms for data analysis, except for injection studies [206, 222], since detecting GWs and inferring their source parameters may require thousands to millions of accurate gravitational waveforms. Nevertheless, a first template bank for nonspinning binaries in Advanced LIGO has been recently constructed from NR waveforms [226]. Furthermore, NR waveforms have been used successfully in calibrating inspiral-merger-ringdown (IMR) effective-one-body (EOB) [68, 189, 201, 219, 227–229] and phenomenological [66, 200, 230, 231] models. These models have free parameters that can be set by matching to NR waveforms and are suitable for certain GW data analysis studies [232]. However, these models can have systematic errors since they assume *a priori* physical waveform structure and are calibrated and tested against a small set of NR simulations.

In this Letter, we present an *ab initio* methodology based on surrogate [76, 233] and reduced order modeling techniques [234–238] that is capable of accurately predicting the gravitational waveform outputs from NR *without* any phenomenological assumptions or approximations to general relativity. From a small set of specially selected non-spinning BBH simulations performed with the Spectral Einstein Code (SpEC) [209, 211, 212], we build a surrogate model for SpEC waveforms that can be used in place of performing SpEC simulations. The techniques are general, however, and directly apply to other NR codes or even analytical waveform models. The surrogate model constructed here generates non-spinning BBH waveforms with mass ratios q in $[1, 10]$, contains 25–31 gravitational wave cycles before peak amplitude, and includes many spherical-harmonic modes (see Table 5.2 and its caption). These choices are made based on available NR waveforms and are not limitations of the method. Our surrogate model has errors close to the estimated numerical error of the input waveforms. An example comparison of the surrogate output to an NR waveform can be seen in Fig. 5.1. This simulation took 9.3 days using 48 cores but

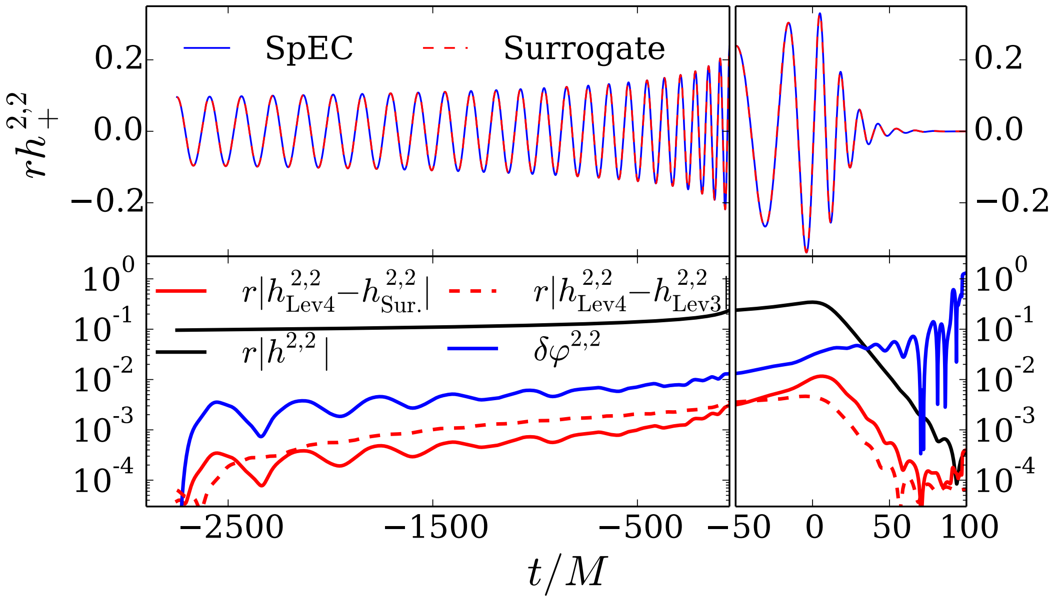

Figure 5.1: **Top**: The + polarization (2,2) mode prediction for $q = 2$, the surrogate model's worst prediction over q from a "leave-one-out" surrogate that was *not* trained with this waveform (see below). Our full surrogate, trained on the entire data set, is more accurate. **Bottom**: Phase $\delta\varphi^{2,2}$ and waveform differences between the surrogate and highest resolution (Lev4) SpEC waveforms. Also shown is the SpEC numerical truncation error found by comparing the two highest resolution (Lev4 and Lev3) waveforms.

only ~ 0.01 sec for the surrogate evaluation of the (2,2) mode.

Previous work [76, 239] built surrogates for EOB waveforms; building and assessing surrogate models of NR waveforms have unique challenges associated with input waveforms that are expensive to compute. We summarize next the construction of our model, focusing on steps that were not addressed in [76] but are required for NR surrogates.

5.2 Parametric sampling

Typically, a surrogate model is trained on a dense set of waveforms known as the *training set*. In the case of NR, however, we cannot afford to generate a large number of waveforms. Instead, we generate a dense set of non-spinning waveforms using an EOB model [229], as implemented in [182], which contains the $(\ell, m) = \{(2,2),(2,1),(3,3),(4,4),(5,5)\}$ spin-weight -2 spherical-harmonic modes and captures the robust features of NR waveforms. The EOB training set waveforms are computed for times in $[-2750, 100]M$, which is the interval over which we build our surrogates and where M is the total mass.

Next, on this training set we apply a greedy algorithm to expose the most relevant mass ratio values [75, 162]. SpEC simulations of non-spinning BBH mergers are then performed for these mass ratios. The resulting NR waveforms are used to build our surrogates without any further input from the EOB model.

We seeded the greedy algorithm with 5 publicly available SpEC simulations of non-spinning BBH mergers [58, 68] (see Table 5.1), and the next 17 (ordered) mass ratio values are the algorithm's output based on the EOB model. The final ≈ 10 mass ratios are included to improve the surrogate if necessary, since we can assess the surrogate model's accuracy only after it is built (cf. Fig. 5.6). Our method for building surrogates is hierarchical [75, 76]; additional NR waveforms can be included to improve the model's accuracy.

5.3 Generating the NR waveforms

Table 5.1 summarizes the 22 SpEC simulations used in this paper. See, e.g., Ref. [212] for the numerical techniques used in SpEC. The numerical resolution is denoted by "Lev*i*", where i is an integer that controls the truncation error allowed by adaptive mesh refinement in SpEC; larger numbers correspond to smaller errors and more computationally-expensive simulations. Two to five levels of resolution are simulated for each mass ratio. To achieve quasi-circular orbits, initial data is subject to an iterative eccentricity reduction procedure resulting in eccentricities $\lesssim 7 \times 10^{-4}$ [132, 133, 240].

SpEC numerically solves an initial boundary value problem defined on a finite computational domain. An observation in the radiation zone, however, is well approximated as taking place at future null infinity $\mathscr{I}^+$. To obtain waveforms at $\mathscr{I}^+$, we use the Cauchy characteristic extraction (CCE) method [213, 241–244]. Using the PittNull code [241–243], we compute the Newman-Penrose scalar Ψ_4 at $\mathscr{I}^+$ and finally obtain the gravitational wave strain h through two temporal integrations. We minimize the low-frequency, noise-induced "drifts" [244] by using frequency cut-offs[1]

---

[1] We integrate Ψ_4 twice in the (dimensionless) frequency domain by dividing $-\Psi_4^{\ell,m}(f)$ by $[2\pi \max(f, 2f_0/3)]^2$, where f_0 is the initial GW mode frequency. For the (2,0) mode, which has high power from low frequencies, we hybridize $\Psi_4^{2,0}$ with a corresponding Taylor T4 post-Newtonian inspiral waveform mode from $600M$ to $100M$ before merger where $\Psi_4^{2,0}$ begins to rise above the SpEC truncation error. Our surrogate is built from and compared to $h^{2,0}$ stemming from hybridized

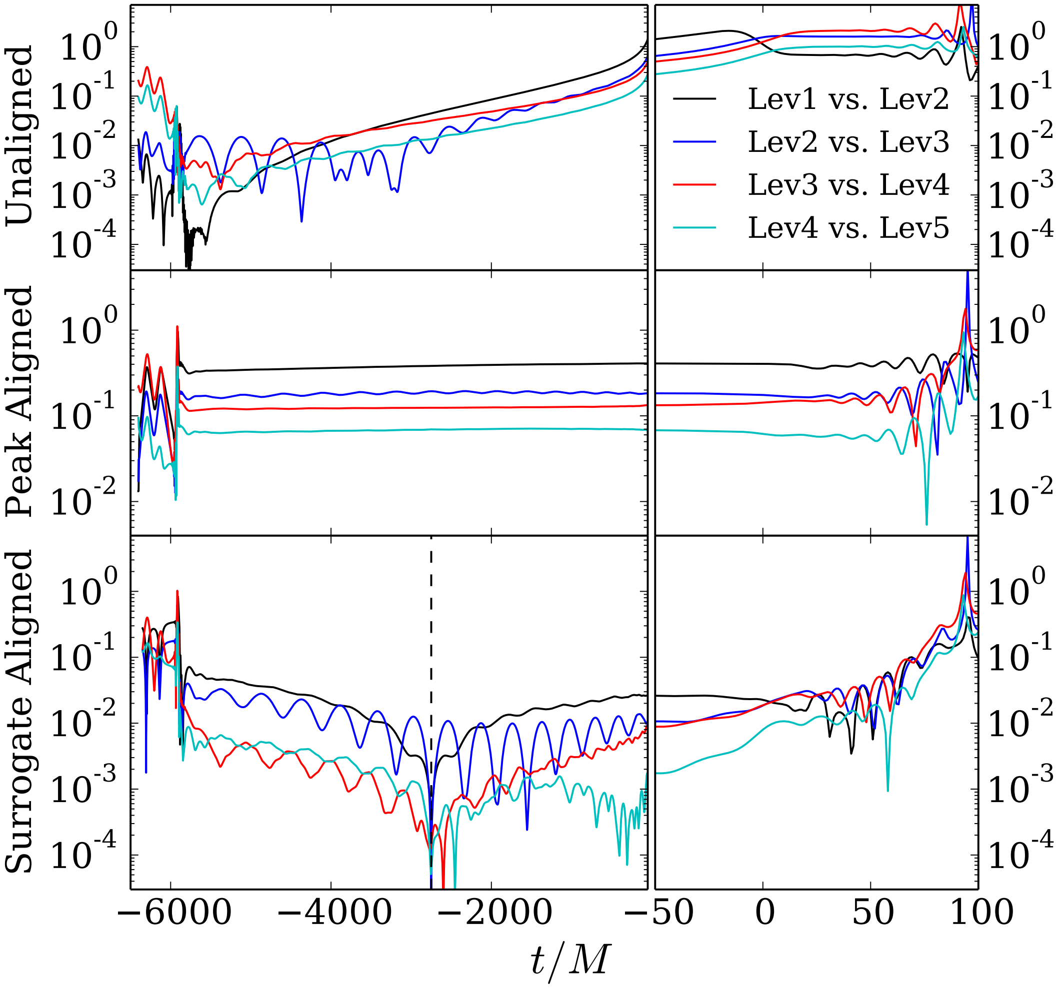

Figure 5.2: The relative error, $|h_i^{22} - h_{i+1}^{22}|/|h_{i+1}^{22}|$, of successive resolutions SpEC Lev$i$ for the (2,2) mode of simulation 19 in Table 5.1. **Top**:Waveform output as directly given by SpEC (“Unaligned"). **Middle**: Peak-aligned waveforms where a Lev-dependent time shift is applied to set all peaks to $t = 0$. **Bottom**: “Surrogate aligned,” which involves a rotation of the binary around the z-axis to align the waveform phases at the surrogate’s initial time $t_i = -2750M$ together with the peak alignment scheme.

#	ID	q	e_{-5}	T/M	Orbs	#	ID	q	e_{-5}	T/M	Orbs
1	180	1.00	5.1	9867	28.2	12	191	2.51	65	6645	22.5
2	181	6.00	5.8	7056	26.5	13	192	6.58	4.0	5149	21.1
3	182	4.00	12	3840	15.6	14	193	3.50	3.0	5242	19.6
4	183	3.00	4.8	4008	15.6	15	194	1.52	74	5774	19.6
5	184	2.00	15	4201	15.6	16	195	7.76	22	5226	21.9
6	185	9.99	31	5817	24.9	17	196	9.66	23	5330	23.1
7	186	8.27	16	5687	23.7	18	197	5.52	25	5061	20.3
8	187	5.04	3.0	4807	19.2	19	198	1.20	17	6315	20.7
9	188	7.19	15	5439	22.3	20	199	8.73	8.5	5302	22.6
10	189	9.17	13	6019	25.2	21	200	3.27	36	5507	20.2
11	190	4.50	2.5	5199	20.1	22	201	2.32	15	5719	20.0

Table 5.1: Properties of the highest resolution SpEC simulations used for building binary black hole waveform surrogates. The orbital eccentricity $e_{-5}/10^5$ is measured after initial transients have has left the computational grid. The duration T/M and number of orbits (Orbs) are also given. The SpEC simulations are available in the public waveform catalog [58] under the name "SXS:BBH:*ID*."

Figure 5.2 shows the convergence typically observed in our simulations. Although textbook convergence is difficult to achieve in production SpEC simulations (see the discussion in Sec. IIIB of [212]), waveform differences typically decrease quickly with increasing resolution. Let

$$\delta h^{\ell,m}(q) \equiv \frac{\|h_1^{\ell,m}(\cdot;q) - h_2^{\ell,m}(\cdot;q)\|^2}{\sum_{\ell,m} \|h_2^{\ell,m}(\cdot;q)\|^2} \tag{5.1}$$

be the disagreement between two waveform modes $h_1^{\ell,m}$ and $h_2^{\ell,m}$ where $\|h^{\ell,m}(\cdot;q)\|^2 = \int dt\,|h^{\ell,m}(t;q)|^2$. We estimate the numerical truncation error of each mode when h_1 and h_2 are waveforms computed at the two highest resolutions. The full waveform error for a given mass ratio is $\delta h(q) = \sum_{\ell,m} \delta h^{\ell,m}(q)$. We also report errors after an overall simulation-dependent time shift and rotation (called "surrogate aligned", described in the next section), which are physically unimportant coordinate changes. The resulting estimated numerical truncation errors of the dominant $(2,2)$ modes, using our surrogate alignment scheme, are shown in Fig. 5.3 (black circles).

Additional error sources are non-zero eccentricity in the (intended to be circular) NR initial data, and an imperfect procedure for integrating $\Psi_4^{\ell,m}$ to obtain $h^{\ell,m} \equiv A^{\ell,m}\exp(-i\varphi^{\ell,m})$. These both cause small oscillations in the waveform amplitudes $A^{\ell,m}(t)$ and phases $\varphi^{\ell,m}(t)$ [244, 245] that we model following [245]. We also

---

data $\Psi_4^{2,0}$.

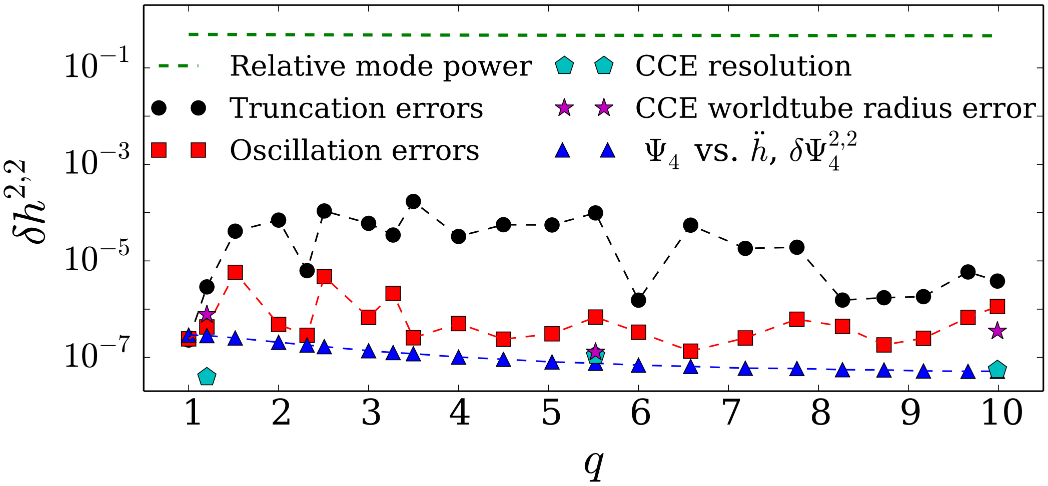

Figure 5.3: Numerical truncation errors (black) dominate all other sources of error for the (2,2) mode, except for simulation 1 ($q \approx 1$), where the truncation errors are already very small. For some weaker modes, systematic amplitude oscillations primarily due to eccentricity may become more relevant.

compute the error in the strain integration scheme by comparing $\Psi_4^{\ell,m}$ to two time derivatives of $h^{\ell,m}$, as well as estimates for numerical errors in the CCE method [213]. For the $(2,2)$ mode, these additional errors are negligibly small compared to SpEC truncation errors (cf. Fig. 5.3).

5.4 Preparing NR waveforms for surrogate modeling

We apply a simulation-dependent time shift and physical rotation about the z-axis so that all the modes' phases are aligned. This reveals the underlying parametric smoothness in q that will be useful for building a surrogate. Our time shifts set each waveform's *total* amplitude

$$A(t;q)^2 \equiv \int_{S^2} d\Omega \, |h(t,\theta,\phi;q)|^2 = \sum_{\ell,m} |h^{\ell,m}(t;q)|^2 \,, \tag{5.2}$$

to be maximum at $t = 0$. After enforcing this alignment scheme we interpolate the waveform mode amplitudes and phases onto an array of uniformly spaced times in $[-2750, 100]M$, with $\Delta t = 0.1M$. Finally, we align the initial gravitational wave mode phases by performing a simulation-dependent, constant (in time) physical rotation about the z-axis so that $\varphi^{2,2}(t_i) = \varphi^{2,-2}(t_i)$, which fixes a physical rotation up to multiples of π. We resolve the ambiguity by requiring $\varphi^{2,1}(t_i) \in (-\pi, 0]$. Waveform truncation errors, after performing this *surrogate alignment* scheme, are shown in Fig. 5.2. In what follows, we call “truncation error after surrogate alignment" simply “truncation error."

5.5 Building the surrogate

Each $m \geq 0$ mode, $h^{\ell,m}(t;q)$, is modeled separately while (due to reflection symmetry about the orbital plane) $m < 0$ modes are evaluated using $h^{\ell,-m}(t;q) = (-1)^{\ell} h^{\ell,m}(t;q)^*$. We do not model modes which do not improve the predictive power of the surrogate. Table 5.2 lists our modeled modes and their errors.

Our complete surrogate waveform model is defined by

$$h_{\mathrm{S}}(t,\theta,\phi;q) = \sum_{\ell,m} h_{\mathrm{S}}^{\ell,m}(t;q)_{-2}Y_{\ell m}(\theta,\phi) \ , \tag{5.3}$$

where

$$\begin{aligned} h_{\mathrm{S}}^{\ell,m}(t;q) &= A_{\mathrm{S}}^{\ell,m}(t;q) e^{-i\varphi_{\mathrm{S}}^{\ell,m}(t;q)} \ , \\ X_{\mathrm{S}}^{\ell,m}(t;q) &= \sum_{i=1}^{N_X} B_{X,i}^{\ell,m}(t) X_i^{\ell,\mathrm{m}}(q) \ , \quad X = \{A, \varphi\}. \end{aligned} \tag{5.4}$$

Unlike Ref. [76], we construct a reduced basis representation for the waveform amplitudes and phases separately, instead of the waveforms themselves [239]. Here, the $\{B_{X,i}^{\ell m}\}_{i=1}^{N_X}$ are computed off-line from the SpEC waveforms [76]. At a set of N_X specially selected times $\{T_{X,i}^{\ell m}\}_{i=1}^{N_X}$, which are the empirical interpolant nodes [76, 246], the functions $X_i^{\ell m}(q) \approx X^{\ell m}(T_{X,i}^{\ell m};q)$ approximate the parametric variation of the amplitudes and phases (via fitting, see below). A thorough discussion of surrogate model building steps is presented in [76]. When evaluating the surrogate at a particular mass ratio, the fits are evaluated first to determine the amplitudes and phases at their respective interpolating times $\{T_{X,i}^{\ell m}\}_{i=1}^{N_X}$. The remaining operations yield the surrogate model prediction, $h_{\mathrm{S}}(t,\theta,\phi;q)$.

To find each $X_i^{\ell m}(q)$ we perform least-squares fits to the 22 data points

$$\{X^{\ell m}(T_{X,i}^{\ell m};q_j)\}_{j=1}^{22}. \tag{5.5}$$

All fits except odd m mode amplitudes use 5th degree polynomials in the symmetric mass ratio, $\nu = q/(1+q)^2$. For odd m modes, the amplitude approaches 0 and its derivative with respect to ν diverges as $q \to 1$ (or $\nu \to 1/4$). Consequently, we use $A_i^{\ell m}(\nu) = \sum_{n=1/2,1}^{5} a_n^{\ell m}(1-4\nu)^n$ to account for this behavior. The waveform phases of odd m modes at $q = 1$, which are undefined, are excluded when fitting for each $\varphi_i^{\ell m}(q)$.

(ℓ,m)	Surrogate		NR		(ℓ,m)	Surrogate		NR	
	Max	Mean	Max	Mean		Max	Mean	Max	Mean
$(2,2)$	0.36	0.07	0.36	0.08	$(3,2)$	100	17	1.7	0.43
$(2,1)$	29	3.4	4.1	0.54	$(4,4)$	7.4	2.2	20	2.1
$(2,0)$	2.1	0.27	0.04	0.01	$(5,5)$	5.4	3.1	1.0	0.33
$(3,3)$	53	4.1	11	0.94	All	0.41	0.12	0.40	0.10

Table 5.2: Relative mode errors, reported as $10^3 \times \|h_{\mathrm{S}}^{\ell,m}(q) - h^{\ell,m}(q)\|^2/\|h^{\ell,m}(q)\|^2$, from the leave-one-out surrogates. Only those modes which contribute greater than 0.02% to the full waveform's time-domain power are used in the computation of the max and mean, except for 'All' which is just δh. Our surrogate also includes the $(3,1)$, $(4,[2,3])$, $(5,[3,4])$, $(6,[4,5,6])$, $(7,[5,6,7])$, and $(8,[7,8])$ modes. Weaker modes typically have relative errors between 1% and 35%.

5.6 Assessing surrogate errors

We next assess the surrogate's predictive quality. To quantify the error in the surrogate model itself, as opposed to its usage in a data analysis study, we do *not* minimize the errors over relative time and phase shifts for any of these studies.

A first test is a consistency check to reproduce the 22 input SpEC waveforms used to build the surrogate. These errors are shown in Fig. 5.4 (red squares) and are comparable to SpEC truncation errors (black circles).

A more stringent test is the leave-one-out cross-validation (LOOCV) study [247]. For each simulated mass ratio q_i, we build a temporary *trial surrogate* using the other 21 waveforms, evaluate the trial surrogate at q_i, and compare the prediction with the SpEC waveform for q_i. Therefore, the trial surrogate's error at q_i should serve as an upper bound for the full surrogate trained on all 22 waveforms. Repeating this process for all possible 20 LOOCV tests[2] results in Fig. 5.4 (blue triangles). Despite the ith trial surrogate having no information about the waveform at q_i, the errors remain comparable to the SpEC truncation errors. The LOOCV errors are typically twice as large as the full surrogate ones confirming the former as bounds for the latter. Relative errors for selected modes are shown in Table 5.2. While weaker modes have larger relative errors, their power contribution is small enough that the error in the full surrogate waveform, δh, is nearly identical to the SpEC resolution error.

A third test is to compare the surrogate waveforms to those of a second surrogate,

---

[2] We omit the smallest and largest mass ratios here as the corresponding trial surrogates would extrapolate to their values.

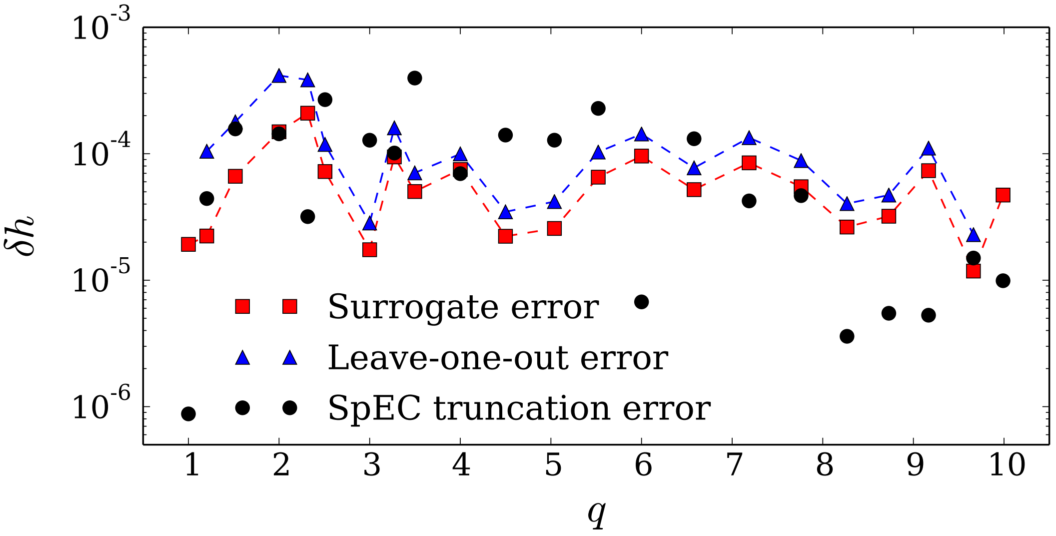

Figure 5.4: Waveform differences between the two highest SpEC resolutions (black circles), the full surrogate and SpEC (red squares), and leave-one-out trial surrogates and SpEC (blue triangles). The largest surrogate error is for $q = 2$, for which the $(2, 2)$ mode is shown in Fig. 5.1.

built from the second highest resolution SpEC waveforms. The resulting comparison (solid lines) is shown for the full waveform and several of its modes in Fig. 5.5. These errors are comparable to SpEC waveform truncation errors (circles). We find that the surrogate building process is robust to resolution differences. Furthermore, the surrogate can be improved using NR waveforms of higher accuracy.

We perform a final test by constructing surrogates using the first N selected mass ratios (from Table 5.1) as input waveforms. This leaves $22 - N$ mass ratios with which to test. For these smaller surrogates, we use fits with $\min(N - 1, 6)$ coefficients where N is the number of input waveforms. The resulting errors are shown in Fig. 5.6. We see that the total waveform error becomes comparable to the SpEC truncation error for surrogates using at least 7 waveforms. However, some modes (e.g., $(3, 3)$) require more waveforms to be fully resolved.

5.7 Discussion

We have built a surrogate model for NR non-spinning BBH merger waveforms generated by SpEC. On a standard 2015 single core computer, all 77 modes with $2 \leq \ell \leq 8$ are evaluated in ≈ 0.5 sec (≈ 0.01 sec for a single mode) providing a factor of $\sim 10^{6-8}$ speedup compared to SpEC. Importantly, this is achieved *without*

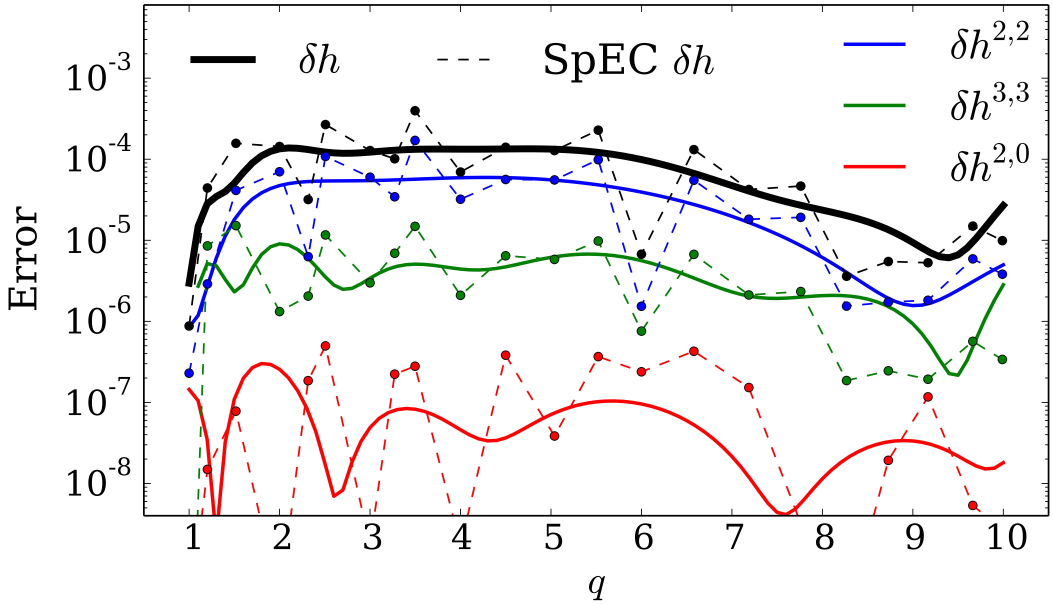

Figure 5.5: Difference between the full, high-resolution surrogate and a surrogate built using the *second* highest resolution SpEC waveforms (solid curves), each evaluated at 91 mass ratios. Differences between the two highest-resolution SpEC waveforms are shown as circles and connected by a dashed line for visual assistance.

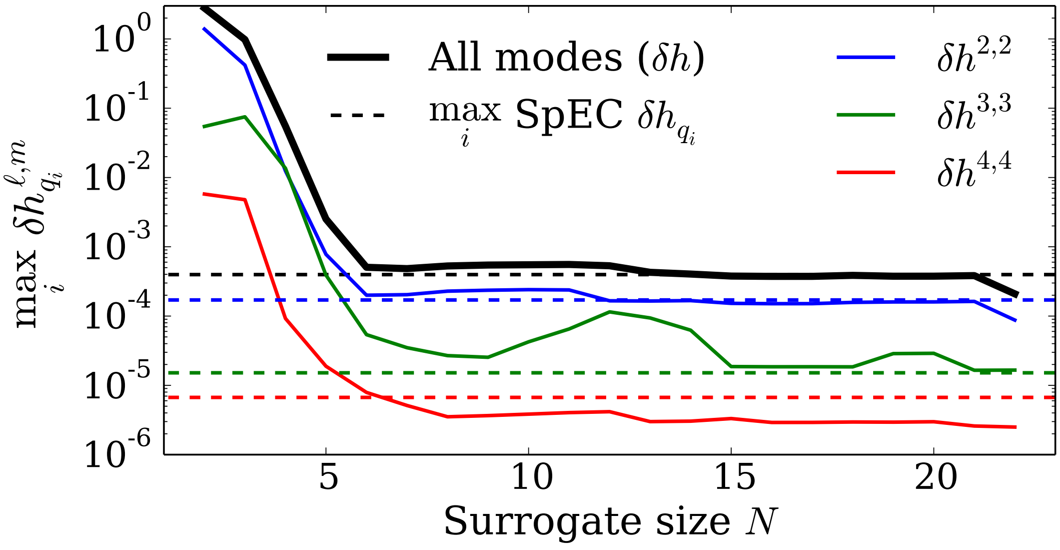

Figure 5.6: Maximum errors over all 22 mass ratios for surrogates built from N SpEC waveforms. Convergence is achieved after $\approx$ 7 waveforms. The observed exponential error decay is expected for any parametrically smooth waveform family [162].

sacrificing accuracy. Like other data-driven modeling strategies, our surrogate is valid only within the training intervals, namely, $q \in [1,10]$ and $t/M \in [-2570,100]$. Therefore, within the training intervals, our surrogate model generates BBH merger waveforms that are equivalent to SpEC outputs *within numerical error*.

NR surrogates can be used for multiple-query applications in gravitational wave data analysis such as detector-specific template-bank (re-)generation and parameter estimation. Our surrogate, and more generally the results of this paper, open up the exciting possibility of performing, for example, parameter estimation with multi-modal NR waveforms (with hybridization, if needed). Parameter estimation studies seeking to incorporate model error may benefit from the surrogate's relatively straightforward characterization and assessment of uncertainty from a combination of the surrogate's and SpEC's systematic and numerical errors. We anticipate NR surrogate modeling to complement traditional strategies [68, 189, 200, 201, 219, 227–232] by providing unlimited high-fidelity approximations of NR waveforms with which to calibrate, refine and make comparisons. Building NR surrogates of precessing BBH merger waveforms, which may be modeled from the parameters specially selected in [248], offer a promising avenue for modeling the full 7 dimensional BBH parameter space. The surrogate model described in this paper will be available for download after peer review at [249]. We thank Mike Boyle, Alessandra Buonanno, Jan Hesthaven, Jason Kaye, Geoffrey Lovelace, Lee Lindblom, Tom Loredo, Christian Ott, Yi Pan, Harald Pfeiffer, Rory Smith, and Nicholas Taylor for many useful discussions throughout this project. This work was supported in part by NSF grants CAREER PHY-0956189, PHY-1068881, PHY-1005655, PHY-1440083, PHY-1404569, and AST-1333520 to Caltech, NSF grants PHY-1306125 and AST-1333129 to Cornell University, NSF grant PHY-1500818 to the University of California at San Diego, NSF grants PHY-1208861 and PHY-1316424 to the University of Maryland (UMD), NSERC of Canada, and the Sherman Fairchild Foundation. Computations were performed on the Zwicky cluster at Caltech, which is supported by the Sherman Fairchild Foundation and by NSF award PHY-0960291. Portions of this research were carried out at the Center for Scientific Computation and Mathematical Modeling cluster at UMD.

6 A Surrogate Model of Gravitational Waveforms from Numerical Relativity Simulations of Precessing Binary Black Hole Mergers

6.1 Background and Contributions

This chapter was published in Physical Review D on May 17 2017 [14], in issue 10 of volume 95 which can be found at `https://link.aps.org/doi/10.1103/PhysRevD.95.104023`. A preprint can be found at `https://arxiv.org/abs/1701.00550`.

This chapter presents a numerical relativity (NR) surrogate model for a $4d$ parameter subspace including precession. The methods of chapter 5 are significantly expanded to accomodate the additional waveform phenomenology due to precession. Difficulties in keeping the NR spin directions within this $4d$ parameter space lead to the need for an analytic approximation and surrogate errors that are much larger than the NR waveform errors. Nevertheless, our model is signifiantly better within its range of validity than other waveform models.

I built the post-Newtonian surrogate models used to determine the parameters for the NR simulations, and helped perform some of the simulations. I determined the waveform decomposition to be used, implemented and built the surrogate models, performed the error studies, and generated most of the figures. I wrote the intial draft of the manuscript, and contributed to editing the manuscript.

A Surrogate Model of Gravitational Waveforms from Numerical Relativity Simulations of Precessing Binary Black Hole Mergers

Jonathan Blackman,[1] Scott E. Field,[2,3] Mark A. Scheel,[1] Chad R. Galley,[1]
Daniel A. Hemberger,[1] Patricia Schmidt,[1,4] and Rory Smith[1,4]

[1] *TAPIR, Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, CA 91125, USA*

[2] *Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, NY 14853, USA*

[3] *Mathematics Department, University of Massachusetts Dartmouth, Dartmouth, MA 02747, USA*

[4] *LIGO Laboratory, California Institute of Technology, MS 100-36, Pasadena, California 91125, USA*

We present the first surrogate model for gravitational waveforms from the coalescence of precessing binary black holes. We call this surrogate model NRSur4d2s. Our methodology significantly extends recently introduced reduced-order and surrogate modeling techniques, and is capable of directly modeling numerical relativity waveforms without introducing phenomenological assumptions or approximations to general relativity. Motivated by GW150914, LIGO's first detection of gravitational waves from merging black holes, the model is built from a set of 276 numerical relativity (NR) simulations with mass ratios $q \leq 2$, dimensionless spin magnitudes up to 0.8, and the restriction that the initial spin of the smaller black hole lies along the axis of orbital angular momentum. It produces waveforms which begin ~ 30 gravitational wave cycles before merger and continue through ringdown, and which contain the effects of precession as well as all $\ell \in \{2,3\}$ spin-weighted spherical-harmonic modes. We perform cross-validation studies to compare the model to NR waveforms *not* used to build the model, and find a better agreement within the parameter range of the model than other, state-of-the-art precessing waveform models, with typical mismatches of 10^{-3}. We also construct a frequency domain surrogate model (called NRSur4d2s_FDROM) which can be evaluated in 50 ms and is suitable for performing parameter estimation studies on gravitational wave detections similar to GW150914.

6.2 Introduction

With two confident detections of gravitational waves (GWs) from binary black hole (BBH) systems [250, 251], an exciting era of gravitational wave astronomy has begun. Once a signal has been detected, the masses and spins of the black holes (BHs), and their uncertainties, can be determined by comparing the signal to waveforms predicted by general relativity (GR) [20]. Similarly, by comparing the signal to predictions, tests of GR can now be performed in the regime of strong-field dynamics with relativistic velocities [252].

Parameter estimation and tests of GR typically require the computation of predicted gravitational waveforms for a large set of different source parameters (e.g. black hole masses and spins). A typical Bayesian parameter estimation analysis, for example, evaluates millions of waveforms [187]. Therefore, in order to obtain reliable results on realistic timescales, the GW model must be fast to evaluate. Additionally, the waveform model must be accurate not only during the weak-field perturbative binary inspiral, but also in the strong-field, large-velocity regime. Otherwise the model may introduce biases in parameter estimation and inaccuracies in tests of GR. Waveform accuracy will become increasingly important in future GW measurements, because higher signal-to-noise-ratio detections are anticipated as detector technology improves.

Numerical relativity (NR) is now in a sufficiently mature state that there are a number of codes [191, 209, 253–257] capable of accurately simulating the late inspiral, merger and ringdown of a BBH system, and the resulting GWs, even for somewhat extreme spins [212, 258] and high mass ratios [259, 260]. While the resulting waveforms are quite accurate, the simulations can take weeks or months, thereby precluding them from being directly used in most data analysis studies. Therefore, data analysis studies currently use approximate NR-tuned waveform models that are fast to evaluate [65, 66, 70, 152, 199, 200, 261, 262].

For the analysis of GW150914 [20, 263], the first GW detection by Advanced LIGO [32], waveform models built within the effective-one-body (EOB) [70, 152, 227, 262, 264, 265] and the phenomenological (Phenom) [65, 66, 261] frameworks were used [20, 263]. All models necessarily introduce some systematic error, however small, which are often quantified either by comparing to NR simulations directly [65, 72, 266–268] or by performing parameter estimation with many different waveform models and monitoring the discrepancies. In the case of GW150914, the systematic error for the black hole masses was estimated to be smaller than the statistical uncertainty. However, estimating a model's systematic error in this way is complicated by the fact that the waveform models make similar simplifications. For example, the models ignore spin-weighted spherical-harmonic (SWSH) modes with $\ell > 2$, which may be significant since the signal's power is dominated by the late inspiral and merger. Recent studies continue to investigate this systematic parameter estimation bias through the use of newer waveform models including additional physics [263] and by comparing to NR waveforms [268].

In this paper, we use a surrogate model, which we call NRSur4d2s, to compute waveforms approaching the accuracy of NR simulations. A surrogate model [76, 239, 265, 269] is a way to substantially accelerate the evaluation of a slower but accurate waveform model (in our case, NR), while largely retaining the accuracy of the original model. This is done by through an expensive *offline* stage, where we perform many accurate NR simulations for different input parameter values and subsequently build and validate the surrogate model on this set of simulations. The waveforms from these simulations are then “interpolated" in parameter space in an inexpensive *online* stage. The resulting model can be used in place of performing additional NR simulations. Surrogates can be used to accelerate other analytical models, and have been used to successfully speed up non-spinning EOB models with multiple SWSH modes [76], and spin-aligned EOB models that include only the $\ell = 2$ modes [239, 265]. Most recently, surrogates have been used to speed up non-spinning BBH waveforms from NR simulations including 40 SWSH modes [269].

The surrogate model we develop here is based on NR simulations using the Spectral Einstein Code (SpEC) [209, 211, 270–274]. It extends previous NR surrogate models [269] to include precessing binaries. The number of NR simulations required to build a surrogate model increases with parameter space size, and NR simulations become more expensive as the mass ratio and spin magnitudes grow. To reduce the computational cost, we restrict to a subspace of the full precessing parameter space. The initial spin direction of the smaller black hole is restricted to be parallel to the orbital angular momentum. We also restrict the mass ratio of the black holes to $1 \leq q \leq 2$ and the dimensionless spin magnitudes to be at most 0.8. The duration of each NRSur4d2s waveform is equal to that of the NR simulations, which begin $4500M$ before merger, corresponding to ~ 30 gravitational wave cycles.

It has been shown that waveforms from precessing systems closely resemble waveforms from non-precessing systems when viewed in a suitable non-inertial, coprecessing frame [169, 171]. We use this relationship to simplify the construction of the surrogate model by decomposing each precessing waveform into a simpler waveform measured in a coprecessing frame [167, 275, 276] plus a time-dependent rotation that characterizes the precession. Additional simplification is achieved by further decomposing each waveform into a set of functions that are slowly-varying in parameter space and thus easier to model (cf. Fig. 6.6). The model is evaluated by "interpolating" these slowly-varying functions to a desired point in parameter space, and then using the interpolated functions to reconstruct the waveform in the inertial source frame of the binary.

The NRSur4d2s surrogate model just described produces a waveform in the time domain, and takes approximately one second to evaluate. While this is much faster than computing a waveform using NR, it is still too slow for many applications; furthermore many LIGO analyses are more easily performed in the frequency domain rather than the time domain. Therefore, we build a second surrogate model in the frequency domain, called NRSur4d2s_FDROM, using NRSur4d2s as input. NRSur4d2s_FDROM does not employ complicated decompositions of its input waveforms, so it requires significantly more waveforms to build (an offline cost), but because of its simplicity it is significantly faster, and can be evaluated in about 50 ms.

We compute errors in both our time-domain and frequency-domain surrogate models by comparing the resulting waveforms with selected NR waveforms that were *not* used to build the models; see Section 6.7 for details. While these errors are larger than the numerical truncation error of the underlying NR simulations, we find that the agreement between NR and our surrogate models is better than that between NR and other precessing waveform models. The accuracy of the surrogate models could be further improved by incorporating additional NR waveforms.

Section 6.3 describes the surrogate modeling methods that have been used previously, and our modifications to them for this work. The NR simulations, as well as their parameters and waveforms, are described in Section 6.4. Section 6.5 describes how the NR waveforms are decomposed into simple pieces, and surrogate models for each piece are built in Section 6.6. The errors of NRSur4d2s are analyzed and compared to other waveform models in Section 6.7. Section 6.8 describes the construction of NRSur4d2s_FDROM from NRSur4d2s, which reduces the computational cost by over an order of magnitude without sacrificing accuracy. Finally, Section 6.9 summarizes this work and discusses potential modifications and improvements.

6.3 Surrogate modeling methods

Compared to previous work [76, 233, 239, 265, 269, 277, 278], which focused on surrogates of analytical waveform models or on surrogates of simpler NR waveforms, surrogate models of precessing numerical relativity (NR) waveforms pose a number of new, unique challenges. First, the complicated waveform morphologies characteristic of precessing systems [279, 280] suggest that a substantially larger training set may be necessary for these systems than for simpler cases considered previously. On the other hand, NR waveforms require the solution of computationally intensive time-dependent partial differential equations; current hardware and binary black hole evolution codes are capable of producing only roughly $\mathcal{O}(1{,}000)$ simulations in about a year.

In this section we outline our method for the construction of precessing NR waveform surrogates, briefly summarizing existing techniques while focusing on solutions to the new challenges. A dimensionless, complex gravitational-wave strain[1]

$$h(t,\theta,\phi;\boldsymbol{\lambda}) = h_{+}(t,\theta,\phi;\boldsymbol{\lambda}) - i h_{\times}(t,\theta,\phi;\boldsymbol{\lambda})\,, \tag{6.1}$$

can be expressed in terms of its two fundamental polarizations h_+ and $h_\times$. Here, t denotes time, θ and ϕ are the polar and azimuthal angles for the direction of gravitational wave propagation away from the source, and $\boldsymbol{\lambda}$ is a set of parameters that characterize the waveform. For concreteness, the parameters $\boldsymbol{\lambda}$ we will use in Sec 6.6 will be the initial mass ratio and spin vectors of the black holes, but the discussion in this section applies to a general set of parameters. Gravitational waveforms considered in this paper are parameterized through their dependence on the initial data, and we shall focus on the the five-dimensional subspace described in Sec. 6.4.

When numerically generating a waveform by solving partial differential equations, one solves an initial-boundary value problem for a fixed $\boldsymbol{\lambda}$, thereby generating a waveform on a dense temporal grid. In this paper we seek to build an accurate and fast-to-evaluate surrogate gravitational-wave strain model $h_{\mathrm{S}}(t,\theta,\phi;\boldsymbol{\lambda})$ by numerically solving the Einstein equations for judicious choices of $\boldsymbol{\lambda}$. Surrogate evaluations require only simple function evaluations, matrix-vector products, and coordinate

---

[1] More precisely, we work with the distance-independent dimensionless strain Rh/M, where R is the distance from the binary's center-of-mass and M is the total Christodoulou mass [281] measured after the initial burst of junk radiation [206] has passed. In this paper we choose units so that $c = G = 1$.

transformations. In Sec. 6.8 we also build a frequency-domain surrogate model, using our time-domain surrogate model as input data, with the purpose of accelerating the evaluation of model waveforms. Evaluation of the frequency-domain model is about 20 times faster than the corresponding time-domain surrogate. Except for Sec. 6.8 our discussion will focus exclusively on time-domain surrogates.

The complex gravitational-wave strain can be written in terms of SWSHs ${}_{-2}Y_{\ell m}(\theta,\phi)$ via

$$h(t,\theta,\phi;\boldsymbol{\lambda}) = \sum_{\ell=2}^{\infty}\sum_{m=-\ell}^{\ell} h^{\ell m}(t;\boldsymbol{\lambda})_{-2}Y_{\ell m}(\theta,\phi)\ , \tag{6.2}$$

where the sum includes all SWSH modes $h^{\ell m}(t;\boldsymbol{\lambda})$. In many data analysis applications, however, one often requires only the most dominant SWSH modes. The NRSur4d2s surrogate model will include all $\ell \leq 3$ modes, while our assessment of the model's error will compare to NR waveforms with all $\ell \leq 5$ modes. Including modes in the NR waveforms which are not included in our model ensures our error studies are sensitive to the effect of neglecting higher order modes. We find that including $\ell = 4$ and $\ell = 5$ modes in our model does not significantly reduce the surrogate errors, but it increases the evaluation cost of the model. As seen in Table 6.3, however, neglecting all $\ell = 3$ modes would significantly increase the surrogate errors, which is why we include $\ell \leq 3$ modes. Other models with which we compare have $\ell = 2$ modes only. When comparing two waveforms with different available modes, missing modes are simply treated as being zero.

The basic surrogate modeling approach

Problem statement

Our surrogate modeling methods build on those outlined in [76], which we briefly describe here. Consider a physical system parameterized by $\boldsymbol{\lambda} \in \mathcal{T}$, where $\mathcal{T}$ is a compact region in the space of possible parameters. We seek quick-to-evaluate time-dependent functions $X(t;\boldsymbol{\lambda})$ that describe this system for times $t \in [t_{\min}, t_{\max}]$. In our case, $\boldsymbol{\lambda}$ will be the black hole masses and spins for a single BBH system, and $\mathcal{T}$ will extend to some maximum spin magnitude and maximum mass ratio for which we choose to compute NR waveforms. The functions $X(t;\boldsymbol{\lambda})$ will be obtained from decomposing $h^{\ell m}(t;\boldsymbol{\lambda})$ as described in Sec. 6.5, but here we discuss building a surrogate model for a single such function.

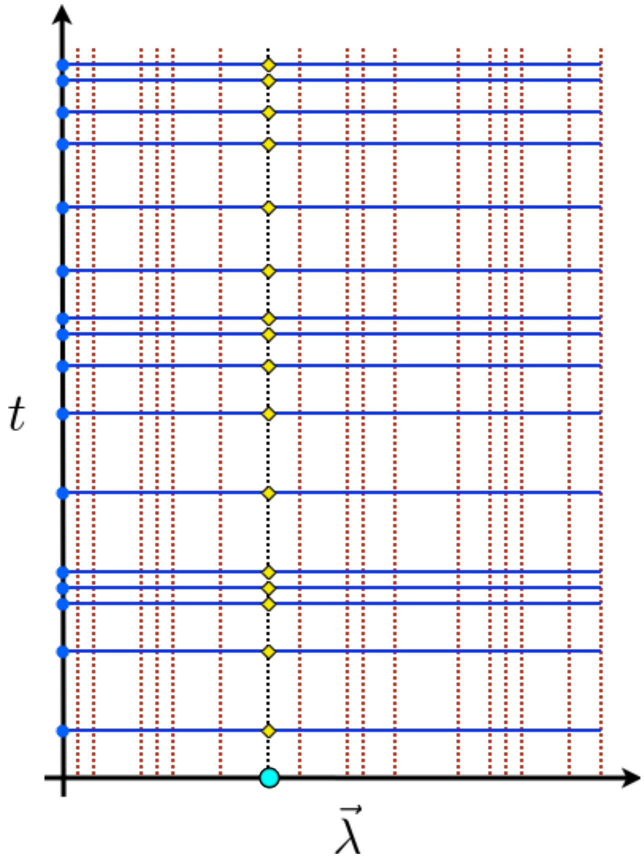

Figure 6.1: A schematic of the method for building a surrogate model for a function $X(t;\boldsymbol{\lambda})$. The red dotted lines show $X(t)$ evaluated at a selected set of greedy parameters $\vec{\Lambda}_i$ used to build a linear basis, and the blue dots show the associated empirical nodes in time from which $X_S(t;\boldsymbol{\lambda})$ can be reconstructed by interpolation with high accuracy. The blue lines indicate fits for $X(t;\boldsymbol{\lambda})$ as a function of $\boldsymbol{\lambda}$ at each of the empirical time nodes. The cyan dot shows a generic parameter $\boldsymbol{\lambda}_0$ that is not in the set of greedy parameters. To compute $X_S(t;\boldsymbol{\lambda}_0)$, each fit is evaluated at $\boldsymbol{\lambda}_0$ (the yellow diamonds), and then the empirical interpolant is used to construct $X_S(t;\boldsymbol{\lambda}_0)$ at arbitrary times (the dotted black line).

We already have a slow method of generating $X(t;\boldsymbol{\lambda})$, so we seek a faster *surrogate model*, denoted as $X_S(t;\boldsymbol{\lambda})$, which approximates $X(t;\boldsymbol{\lambda})$. The surrogate model $X_S(t;\boldsymbol{\lambda})$, whose construction is summarized in this section culminating in Eq. (6.13), is built to achieve small approximation errors $\|X(\cdot;\boldsymbol{\lambda}) - X_S(\cdot;\boldsymbol{\lambda})\|$. In our case, the slow method is performing a NR simulation, extracting $h^{\ell,m}(t;\boldsymbol{\lambda})$, and decomposing it to obtain $X(t;\boldsymbol{\lambda})$. A solution $X(t;\boldsymbol{\lambda})$ for a fixed $\boldsymbol{\lambda}$ is represented as a single (dotted red) vertical line in Fig. 6.1, which diagramatically represents the surrogate model.

Discovering representative binary configurations

The first steps in building a surrogate model are to determine a finite set of *greedy parameters*

$$G \equiv \{\mathbf{\Lambda}_i \in \mathcal{T}\}_{i=1}^{N} .$$

An NR simulation is then performed at each greedy parameter, yielding the greedy solutions $\{X(t;\mathbf{\Lambda}_i)\}_{i=1}^{N}$, shown as vertical dotted red lines in Fig. 6.1.

One strategy (described in more detail in [76]) to find the greedy parameters begins by evaluating the slow method on a densely sampled *training set*, $\mathcal{T}_{\rm TS} \subset \mathcal{T}$. This training set is input to a *greedy algorithm* (hence the name greedy parameters) that works as follows. First, the greedy algorithm is initialized by arbitrarily selecting the first few greedy parameters which are sometimes called the algorithm's *seed* [2]. The set of greedy parameters is then extended iteratively by first building an orthonormal linear basis $B_n = \{e^i(t)\}_{i=1}^{n}$ spanning the n current greedy solutions, such that

$$X(t;\mathbf{\Lambda}_j) = \sum_{i=1}^{n} c_i(\mathbf{\Lambda}_j) e^i(t) . \tag{6.3}$$

The aim of the greedy algorithm is to extend this basis such that the approximation

$$X(t;\boldsymbol{\lambda}) \approx \sum_{i=1}^{n} c_i(\boldsymbol{\lambda}) e^i(t) , \quad \boldsymbol{\lambda} \in \mathcal{T}_{\rm TS} \tag{6.4}$$

is as accurate as possible and where the coefficient $c_i(\boldsymbol{\lambda})$ is the inner product of $X(t;\boldsymbol{\lambda})$ with $e^i(t)$. Coefficients found in this way define an *orthogonal projection* of the function $X(t;\boldsymbol{\lambda})$ onto the span of the basis. We compute the projection errors

$$E_n(\boldsymbol{\lambda}) = \|X(\cdot;\boldsymbol{\lambda}) - \sum_{i=1}^{n} c_i(\boldsymbol{\lambda}) e^i(\cdot)\| \tag{6.5}$$

for each $\boldsymbol{\lambda} \in \mathcal{T}_{\rm TS}$, and the next greedy parameter $\mathbf{\Lambda}_{n+1}$ is chosen to be the one yielding the largest projection error. The next basis vector $e^{n+1}(t)$ is then obtained by orthonormalizing $X(t;\mathbf{\Lambda}_{n+1})$ against B_n, and the basis set is extended as $B_{n+1} = B_n \cup \{e^{n+1}(t)\}$. The algorithm terminates once the basis achieves an accuracy requirement $E_N(\boldsymbol{\lambda}) \leq \epsilon$, for some predetermined error tolerance ϵ, over the whole training set. With a dense enough training set and assuming X varies smoothly over $\mathcal{T}$, the projection errors outside of the training set will be only mildly larger than ϵ.

---

[2] The final set of greedy parameters selected by the greedy algorithm will depend on that choice of seed. However, the number and distribution of greedy parameters is expected to be robust to the choice of seed [75, 282].

This method unfortunately requires evaluating the slow method on each point in the (large) training set, so we make modifications as described in sections 6.3 and 6.4.

Temporal compression

We have built a linear basis B_N which can represent $X(t;\boldsymbol{\lambda})$ for any $\boldsymbol{\lambda} \in \mathcal{T}$ using Eq. 6.4, up to some small projection error. This reduces the problem of determining $X(t;\boldsymbol{\lambda})$ to determining the basis coefficients $\{c_i(\boldsymbol{\lambda})\}_{i=1}^N$. The most straightforward method of doing so would be to fit or interpolate the basis coefficients c_i over the parameter space $\mathcal{T}$ as is done in [239, 265]. We have more intuition for the behavior over $\mathcal{T}$ of the solutions $X(T;\cdot)$ evaluated at a fixed time T than we do for the basis coefficients. We will therefore pursue an *empirical interpolation* approach, described in detail in [76], which enables us to avoid fitting the basis coefficients.

An empirical interpolant makes use of the orthogonal linear basis $B_N = \{e^i(t)\}_{i=1}^N$ such that the errors given by Eq. (6.5) are small, so Eq. (6.4) continues to provide a good approximation despite using a different method to compute the coefficients. During the construction of the empirical interpolant, N empirical time nodes $\{T_j\}_{j=1}^N$ will be used. An algorithm to find these special time nodes will be described later on.

We denote an N-node empirical interpolant of a function $f(t)$ by $I_N[f](t)$. A conceptually helpful way to think of the empirical interpolant is that $I_N[f](t)$ lies in the span of B_N, passes through $f(T_j)$ at time T_j, and is nearly as accurate as the orthogonal projection. To construct the interpolant, we expand it in terms of unknown coefficients c_i,

$$I_N[f](t) = \sum_{i=1}^{N} c_i e^i(t)\,. \tag{6.6}$$

We then write a linear system of equations

$$\sum_{i=1}^{N} c_i e^i(T_j) = f(T_j)\,, \quad j = 1,\ldots,N \tag{6.7}$$

and we solve this system for all the coefficients c_i. A good choice of empirical time nodes will ensure that the matrix $V_{ij} = e^i(T_j)$ is well-conditioned, thereby allowing an accurate solution

$$c_i = (V^{-1})_{ij} f(T_j)\,. \tag{6.8}$$

We can then substitute the coefficients back into Eq. (6.6) to obtain

$$I_N[f](t) = \sum_{i=1}^{N} (V^{-1})_{ij} f(T_j) e^i(t)\,. \tag{6.9}$$

If we then define

$$b^j(t) = \sum_{i=1}^{N} (V^{-1})_{ij} e^i(t)\,, \tag{6.10}$$

we obtain

$$I_N[f](t) = \sum_{j=1}^{N} f(T_j) b^j(t)\,. \tag{6.11}$$

Here $b^j_N(t)$ is computed before evaluating the surrogate, so evaluating the empirical interpolant amounts to a matrix multiplication.

If $f(t)$ lies in the span of B_N, then $I_N[f](t) = f(t)$ for all times t. Otherwise, there will be some interpolation error. In practice, the empirical time nodes are constructed iteratively using bases B_n for $n = 1, \ldots, N$. If I_n is the nth iteration of the interpolant, then the nth empirical time node T_n is chosen to be the time t yielding the largest interpolation error when interpolating $e^n(t)$ using the previous interpolant I_{n-1}. The iteration begins with the initial interpolant chosen to be $I_0[f](t) = 0$ for all f.

Note that since the empirical interpolant is linear and V is well-conditioned, if $f(t)$ has a deviation from the span of B_N of order ϵ, then the empirical interpolation error will also be of order ϵ. Since our basis B_N is constructed such that the projection errors of $X(t; \boldsymbol{\lambda})$ onto B_N are small for all $\lambda \in \mathcal{T}$, we can use the empirical interpolant $I_N[X](t)$ to obtain $X(t; \boldsymbol{\lambda})$ for all times t given the empirical node values $\{X(T_j; \boldsymbol{\lambda})\}_{j=1}^{N}$. The remaining step is then to approximate the N functions

$$X_j(\boldsymbol{\lambda}) = X(T_j; \boldsymbol{\lambda}) \tag{6.12}$$

by fitting the available data $\{X(T_j; \boldsymbol{\Lambda}) : \boldsymbol{\Lambda} \in G\}$ over the parameter space $\mathcal{T}$. We call these *parametric fits*, and denote the fitted approximation for $X_j(\boldsymbol{\lambda})$ by $X_{jS}(\boldsymbol{\lambda})$. The parametric fits are represented by the blue horizontal lines in Fig. 6.1. The explicit form of our *surrogate model for* X is then given by

$$X_S(t; \boldsymbol{\lambda}) = \sum_{j=1}^{N} X_{jS}(\boldsymbol{\lambda}) b^j(t)\,. \tag{6.13}$$

Modifications to the basic surrogate modeling approach

A drawback of the algorithm presented in § 6.3, and of many previous surrogate modeling efforts, is the assumption that the original slow model can be evaluated an arbitrary number of times to build a dense training set. Because of the significant computational expense, this is not feasible for waveforms found by numerically

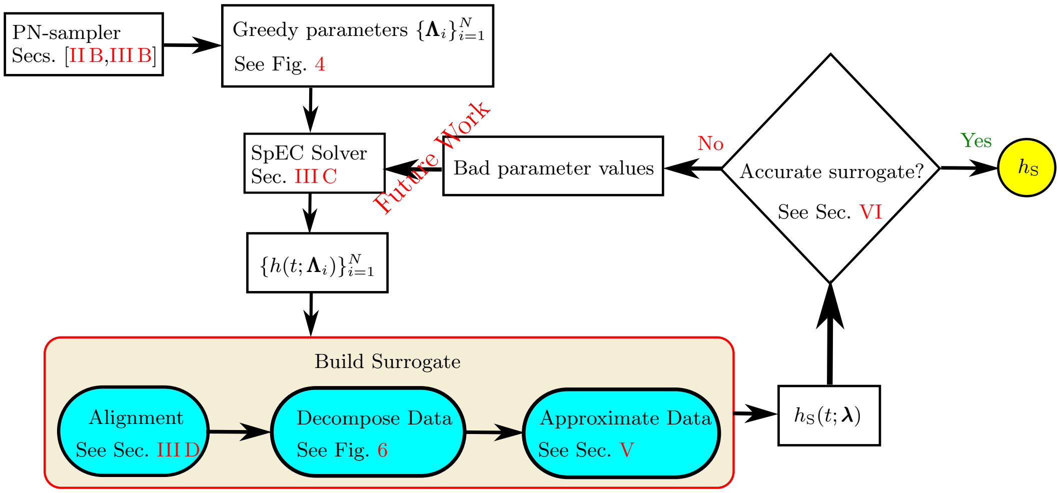

Figure 6.2: Surrogate workflow. A greedy "PN-sampler" selects the most informative parameter values $\{\mathbf{\Lambda}_i\}_{i=1}^N$ for a fixed parametric and temporal range. For each selected value $\mathbf{\Lambda}_i$, SpEC generates a gravitational waveform. A surrogate model building algorithm (cf. Fig. 6.1) is applied to a set of suitably aligned and decomposed (cf. Fig. 6.6) numerical relativity waveforms, thereby producing a trial surrogate. A handful of validation tests are performed to assess the surrogate's quality. If the surrogate performs poorly for some parameter values, one could produce additional numerical relativity waveforms near those values, and rebuild a more accurate surrogate.

solving the Einstein equations. We can neither build NR surrogate models from dense training data nor can we assess the surrogate's quality at arbitrarily many randomly chosen validation points. In previous work that used computationally inexpensive waveform models [76], thousands of nonspinning waveforms comprised the training set, yet the final surrogate required only a very small subset of greedy parameters G. If we could have predicted G in advance then dense training sets would not be required.

Since we cannot evaluate an arbitrarily large number of NR waveforms, we instead first construct a temporary *mock surrogate* using a simpler waveform model that is both fast to evaluate and is defined in the training region of interest. In this paper, for the purpose of discovering the most relevant parameter values, we build a mock surrogate using the precessing TaylorT4 Post-Newtonian (PN) waveform model as implemented in GWFrames [168, 283]. We determine the PN greedy parameters G^{PN} using a training set containing many thousands of these PN waveforms, as described in Sec. 6.4. If we then assume that the distribution of parameters selected using PN waveforms roughly mimics the distribution we would have obtained had NR waveforms been available, then G^{PN} should be a suitable set of greedy parameters for building a NR surrogate. This was found to work well for the non-spinning surrogates of Ref. [248] and, as judged by our validation studies, continues to remain applicable to the precessing waveforms considered here. Instead of PN, we could have used a different analytical waveform model [65, 66, 70, 152, 227, 261, 262, 264, 265] that contains merger and ringdown. However, these other models either omit $\ell = 3$ modes, omit precession, or yield waveforms that do not vary smoothly as a function of $\boldsymbol{\lambda}$. We find that these other considerations outweigh the inclusion of merger and ringdown.

This entire process just described is shown in the first stage of the surrogate workflow diagram (Figure 6.2) as the “PN-sampler". Once the points G^{PN} have been selected, the corresponding NR waveforms are generated, and the surrogate building proceeds as in Fig. (6.2). We emphasize that *no* PN waveforms are used to build the resulting NRSur4d2s surrogate; the PN model is used only to find the greedy parameters $G = G^{\mathrm{PN}}$.

While the PN waveforms are much cheaper to evaluate than NR waveforms, building a dense training set remains costly for high dimensional parameter spaces. In Ref. [248], it was found that an accurate basis can be achieved using small, sparse training sets if each iteration i of the greedy algorithm uses an independent randomly-sampled training set $\mathcal{T}_{\mathrm{R}}^{i}$. We extend this methodology by also including in our training sets a fixed set of parameters $\mathcal{T}_{\mathrm{B}}$ on the boundary of $\mathcal{T}$ (for example, the maximum mass ratio allowed in $\mathcal{T}$). This is motivated by the fact that the boundary of $\mathcal{T}$ carries significant weight both when building a linear basis and when performing parametric fits. At the ith greedy iteration, we then have

$$\mathcal{T}_{\mathrm{TS}}^{i} = \mathcal{T}_{\mathrm{B}} \cup \mathcal{T}_{\mathrm{R}}^{i} \tag{6.14}$$

as our training set of parameters at which we evaluate PN waveforms.

Another issue with the standard greedy algorithm is that it considers only a single function X. For modeling waveforms, we will decompose each waveform $h^{\ell m}(t; \boldsymbol{\lambda})$ into many such functions, which we call *waveform data pieces* (cf. Sec. 6.5). Rather than generate a separate set of greedy parameters G_X^{PN} for each X, we construct a single set of greedy parameters G^{PN} that can be used for all waveform data pieces X. We do so by replacing the projection errors for a single waveform data piece given in Eq. 6.5 with a single error including contributions from all waveform data pieces. This will be described explicitly in Sec. 6.4 after the waveform decomposition and error measures have been introduced.

The standard greedy algorithm guarantees that the basis yields small projection errors given by Eq. 6.5. Therefore, if we have perfect parametric fits (so that $X_{jS}(\boldsymbol{\lambda}) = X_j(\boldsymbol{\lambda})$ for all $\boldsymbol{\lambda} \in \mathcal{T}$) then the surrogate model X_S given by Eq. 6.13 will agree with X in the sense that the L^2 norm of $X_S(t;\boldsymbol{\lambda}) - X(t;\boldsymbol{\lambda})$ will be small for all $\boldsymbol{\lambda} \in \mathcal{T}$. There is, however, no corresponding guarantee that the greedy points are sufficient for producing accurate parametric fits X_{jS}. In the one-dimensional models built in Refs. [76, 269], the parametric fits performed well using the samples produced from the standard greedy algorithm. As the dimensionality of the parameter space increases, the number of greedy parameters required for an accurate basis grows slowly [284], but the number of samples required for accurate parametric fits can grow rapidly. We therefore anticipate that the standard greedy algorithm alone may lead to underresolved parametric fits in problems with high dimensionality.

We overcome this problem by first performing a greedy algorithm to obtain greedy parameters G_0^{PN} that ensure small basis projection errors, and then performing a second greedy algorithm, seeded with G_0^{PN}, that produces the final set of PN greedy parameters G^{PN}. In each iteration of the second greedy algorithm, a mock PN surrogate is constructed from PN waveforms evaluated at the current set of greedy parameters, including the parametric fits at each empirical node. To select the next greedy parameter in this second greedy algorithm, for each $\boldsymbol{\lambda} \in \mathcal{T}_{\text{TS}}^i$ we compute an error between a PN waveform evaluated at $\boldsymbol{\lambda}$ and the mock-PN surrogate evaluation at $\boldsymbol{\lambda}$. Since the basis is already accurate and in general $\boldsymbol{\lambda}$ will not have already been selected as a greedy parameter, this procedure selects points for which the parametric fits are underresolved.

Handling noise in the NR waveforms

The presence of numerical noise in the input NR waveforms complicates the construction of surrogates. The situation is simpler when building a surrogate of a waveform model that is mostly noise-free, such as post-Newtonian or EOB models that require the solution of ODEs (which can be evaluated to almost arbitrary accuracy) but not PDEs. For example, Ref. [76] demonstrates in their Fig. 15 that EOB surrogates can be made to have arbitrarily small errors, and Refs. [239, 265] use interpolation across the parameter space without needing to avoid potential pitfalls such as overfitting the noise. We do not expect this to be the case for numerical relativity waveforms which are beset by numerous error sources, some of which cannot be made arbitrarily small with current computing technology.

Systematic as well as numerical errors can influence the quality of the NR waveform. For example, when attempting to model non-eccentric binaries, the NR simulations will always have some small but non-zero orbital eccentricity. In this paper we will mostly focus on numerical truncation error. This is typically the dominant source of error in SpEC waveforms [285], and the other sources of error are expected to be significantly smaller than truncation error, and smaller than the surrogate error (see Fig. 3 of Ref. [269]). The numerical error can be quantified through standard convergence tests [285]. Following Ref. [269], we will (i) characterize SpEC waveform error across the parameter space and, if necessary, remove poorly-resolved waveforms (Sec. 6.5) (ii) avoid overfitting the noise sources (App. 6.11), and (iii) set surrogate accuracy goals based on our answer to the first question. In future work it would be interesting to study the impact of other noise sources.

Decomposing NR waveforms into simpler components

The detailed time dependence of an NR waveform is generally too complicated to model directly with an acceptable degree of accuracy. Instead, each NR waveform is decomposed into *waveform data pieces* (cf. Sec. 6.5), which are simpler, more slowly-varying functions that can be modeled more easily. A surrogate model is then built for each waveform data piece (cf Sec. 6.6), and then these models are recombined to produce a full surrogate waveform. This process is shown in the “Surrogate Build" step shown in Fig. 6.2.

Selecting the waveform data pieces is a critically important step. For example, in nonspinning [76, 269] and spin-aligned [239, 265] surrogate models, the waveform data pieces are either the real and imaginary parts of the SWSH modes, $h^{\ell m}$, or the amplitude and phase decompositions of these modes $A_{\ell m}$ and $\phi_{\ell m}$, where $h^{\ell m} = A_{\ell m} \exp\left(-\mathrm{i}\phi_{\ell m}\right)$. The idea is that it is easier to model every $A_{\ell m}$ and $\phi_{\ell m}$, which are smooth and slowly-varying functions of time, than it is to directly model the complicated waveform $h(t, \theta, \phi; \boldsymbol{\lambda})$, Eq. (6.2).

Because of the complexity of precessing waveforms, we have needed to pursue a somewhat more complicated decomposition scheme than in the nonprecessing case. Fig. 6.6 summarizes the decomposition scheme used here. Briefly, each waveform is transformed into a coordinate frame in which the binary is not precessing [167, 168, 275, 276, 286]; specifically, we use the minimal-rotation coprecessing frame of Boyle [167]. The waveform modes in this frame have a simpler structure than their inertial frame counterparts. Additional simplifications occur by applying further transformations (described in detail in Sec. 6.5) to the coprecessing-frame waveform modes. The result of these steps is a set of waveform data pieces. If $X(t, \boldsymbol{\lambda})$ is a single waveform data piece, then for that piece we build a surrogate $X_{\mathrm{S}}(t, \boldsymbol{\lambda}) \approx X(t, \boldsymbol{\lambda})$. Here X can stand for any of the decomposed waveform data pieces depicted as cyan ellipses in Fig. 6.6, for example A_{+}^{22}, ϕ_{-}^{32}, φ_p, etc. The full NRSur4d2s surrogate waveform model is defined by the individual data piece surrogates, $X_{\mathrm{S}}(t, \boldsymbol{\lambda})$, and the inverse transformations required to move back up the data decomposition diagram (Fig. 6.6) and reconstruct the waveform from all of the $X_{\mathrm{S}}(t, \boldsymbol{\lambda})$.

Tools for surrogate model validation

Here we describe a useful framework for assessing the surrogate's predictive quality when only a limited number of waveforms are available. This is a different setting from the EOB surrogates of [239, 265] where *out-of-sample* validation studies could be performed at arbitrarily many parameter values. The primary tool we shall use is *cross-validation* [247], which was also used in [269]. Cross-validation happens after the surrogate is built and determines whether or not more SpEC waveforms are needed to improve the accuracy of the model (see Fig. 6.2).

We consider the case where our full dataset is composed of N SpEC waveforms. From the full dataset, we select non-intersecting sets of trial and verification waveforms with sizes N_{t} and N_{v}, such that $N_{\mathrm{t}} + N_{\mathrm{v}} \leq N$. In the cross-validation step, a new *trial surrogate* is built solely from N_{t} trial waveforms. The remaining N_{v} verification waveforms serve as an *exact and independent* error measure of the *trial* surrogate's prediction. The key assumption, which we believe to be true in practice, is that the surrogate built from all N waveforms will have an accuracy similar to the trial surrogates, if not better. Indeed, each step of the surrogate building algorithm will be more accurate so long as parametric overfitting is kept under control. Therefore, the trial surrogate's error should serve as a useful estimate of the error associated with the full surrogate built from all N waveforms. We note, however, that when N_{v} is small or the surrogate error is dominated by some systematic source of error, the improved accuracy when including all N waveforms may not be enough to overcome the variance in the accuracy of the parametric fits seen in Fig. 6.5. In that case the full surrogate error may in fact be slightly larger than a trial surrogate error.

Two variants of cross-validation are considered. *Random* cross-validation proceeds by selecting the verification waveform set randomly. When $N_{\mathrm{v}} = 1$ this is known as the *leave-one-out* strategy. In Ref. [269], all possible leave-one-out studies were performed. In our case, N is sufficiently large and surrogate-building is sufficiently expensive that we opt to choose $N_\mathrm{v} = 10$. We can perform many resamplings of the validation subset to infer an error profile across the parameter space.

Deterministic cross-validation proceeds by selecting the verification waveforms according to a rule. For example, the greedy bases are already ranked according to a “most important" criterion. We select the first N_t greedy waveforms for our validation set. These should contribute most heavily to the surrogate’s overall predictive ability, while the last N_v verification ones are quite dissimilar from the trial waveforms due to the greedy selection process. We fix N_v to have a consistent test of our trial surrogates, and vary $N_\mathrm{t} \leq N - N_\mathrm{v}$ to estimate how the surrogate errors depend on N.

Waveform error measurements

This subsection summarizes the most commonly used tools to compare waveforms. A typical scenario is to quantify differences between waveforms, for example to compare a waveform model to NR waveforms or to estimate the numerical truncation error associated with an NR waveform.

Let $h_1(t,\theta,\phi;\boldsymbol{\lambda}_1)$ and $h_2(t,\theta,\phi;\boldsymbol{\lambda}_2)$ denote waveforms from two different models (or two NR simulations with different numerical resolution) potentially evaluated at different parameter values $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$. We assume the waveforms are already aligned according to the procedure of Sec. 6.4. Decomposing these waveforms into SWSHs we compute a time-dependent error

$$\delta h(t) = \sqrt{\sum_{\ell,m} |\delta h^{\ell m}(t)|^2}\,, \tag{6.15}$$

from the individual mode differences

$$\delta h^{\ell m}(t) = h_1^{\ell m}(t;\boldsymbol{\lambda}_1) - h_2^{\ell m}(t;\boldsymbol{\lambda}_2)\,. \tag{6.16}$$

We use the time-domain inner product

$$\langle a,b\rangle_t = \frac{1}{T}\int_{t_\mathrm{min}}^{t_\mathrm{max}} a(t)b^*(t)dt\,, \tag{6.17}$$

between any complex functions of time a and b, where $T = t_\mathrm{max} - t_\mathrm{min}$ and $*$ denotes complex conjugation. The associated norm $\|a\|^2 = \langle a,a\rangle_t$ can be used to compute mean-squared errors, and we compute the full time-domain waveform error

$$(\delta h)^2 = \int_{S^2} ||h_1(t,\theta,\phi;\boldsymbol{\lambda}_1) - h_2(t,\theta,\phi;\boldsymbol{\lambda}_2)||^2 d\Omega \tag{6.18}$$

$$= \sum_{\ell m} ||\delta h^{\ell,m}||^2 \tag{6.19}$$

$$= \frac{1}{T}\int_{t_{\min}}^{t_{\max}} \delta h(t)^2 dt \tag{6.20}$$

as a sum over individual mode errors $||\delta h^{\ell,m}||$. We note that we do not perform any time or phase shifts to minimize this error. Since waveforms with different mass ratios and spins will have different norms, the error we will use most often is defined as

$$\mathcal{E}[h_1,h_2] = \frac{1}{2}\frac{\delta h^2}{||h_1||^2}\,, \tag{6.21}$$

where h_1 is taken to be the more trusted waveform (usually the highest resolution NR waveform). The factor of $1/2$ is motivated in Appendix 6.13 and makes $\mathcal{E}$ similar to a weighted average over the sphere of overlap errors between $h_1(t,\theta,\phi;\boldsymbol{\lambda}_1)$ and $h_2(t,\theta,\phi;\boldsymbol{\lambda}_2)$, where the overlap error is $1-\mathcal{O}$ with

$$\mathcal{O} = \frac{\langle h_1, h_2\rangle}{\sqrt{\langle h_1, h_1\rangle\langle h_2, h_2\rangle}}\,. \tag{6.22}$$

We note, however, that while the overlap error vanishes if h_1 and h_2 are identical up to a constant factor, $\mathcal{E}$ does not and vanishes only when h_1 and h_2 are identical. This is important as a different normalization will lead to a bias when measuring the distance to the source of a gravitational wave.

Overlap errors are often computed in the frequency domain with a noise-weighted inner product [287]

$$\langle a,b\rangle_f = 4\mathrm{Re}\int_{f_{\min}}^{f_{\max}} \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)}df, \tag{6.23}$$

where $S_n(f)$ is the noise power spectral density of a gravitational wave detector and tildes are used to represent a Fourier transform. We define the *mismatch* as the overlap error, $1-\mathcal{O}$, minimized over one or more extrinsic parameters such as an overall time shift.

6.4 Populating the Set of NR Waveforms

Parameter space

Non-eccentric BBH systems are parametrized by the mass ratio $q = m_1/m_2 \geq 1$ as well as the two dimensionless BH spin vectors $\vec{\chi}_1$, $\vec{\chi}_2$. The total mass $M = m_1 + m_2$ scales out of the problem, and can be used to restore appropriate dimensions to times and distances. Because the spin vectors precess and are therefore time-dependent, to use them as parameters one must specify them at a particular time or frequency. We choose to specify parameters at a reference time of $t_0 = t_{\mathrm{peak}} - 4500M$, where t_{peak} is the time at which the quadrature sum of the waveform modes,

$$A_{\mathrm{tot}}(t) = \sqrt{\sum_{\ell,m} |h^{\ell,m}(t)|^2}, \tag{6.24}$$

reaches its maximum value.

We restrict to a $5d$ subspace of the parameter space where $\vec{\chi}_2$ is aligned with the Newtonian orbital angular momentum $\hat{L}_N$ at the reference time (in practice the NR simulations give us small misalignments but ignore them; see Sec. 6.4). Let θ_χ and ϕ_χ be the polar and azimuthal angles of $\vec{\chi}_1$ at the reference time. Then our five parameters are q, $|\vec{\chi}_1|$, χ_2^z, θ_χ, and ϕ_χ (see Fig. 6.3). While NR simulations can be done for nearly extremal spins [212] and large mass ratios [259], they are computationally expensive and so we restrict to $|\vec{\chi}_1| \leq 0.8$, $|\chi_2^z| \leq 0.8$ and $q \leq 2$. These bounds were also motivated by the parameters of GW150914, which was close to equal mass and did not show strong evidence of large spin magnitudes [20].

To further simplify the surrogate, we attempted to reduce the parameter subspace from 5d to 4d by restricting $\phi_\chi = 0$. While this can be done for analytic waveforms (PN, EOB, etc), it is problematic for NR waveforms. This is because it is not possible to accurately predict the amount of time between the start of an NR simulation and the peak of $A_{\mathrm{tot}}(t)$, without having carried out the simulation. Therefore, it is not possible to precisely set initial conditions of the simulation so that $\phi_\chi = 0$ at the reference time. Therefore, our NR simulations actually cover a $5d$ and not a $4d$ subspace of the parameter space, and we must include ϕ_χ as one parameter. Since we nevertheless attempt to obtain $\phi_\chi = 0$ when choosing the NR initial data parameters, the actual distribution of ϕ_χ is highly correlated with other parameters. Since we do not have full coverage of this $5d$ parameter space, we avoid including the extra dimension ϕ_χ in the NRSur4d2s surrogate model by using an analytic approximation

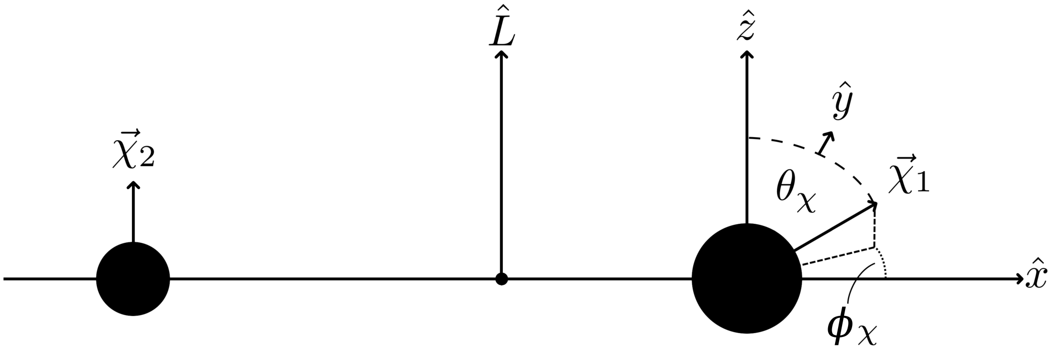

Figure 6.3: Diagram of the four spin components in the $5d$ parameter subspace. We attempt to obtain $\phi_\chi = 0$ at $t_0 = 4500M$ before peak amplitude, but in practice the NR simulations have arbitrary values of ϕ_χ.

for the ϕ_χ dependence of the model, as described in Sec. 6.5. The surrogate model can then predict waveforms for parameters in the $5d$ subspace, but the ϕ_χ dimension is entirely described by the analytic approximation.

Selection of greedy parameters

We use $\phi_\chi = 0$ while determining the greedy parameters $G = \{\mathbf{\Lambda}_i\}$, and we use PN waveforms to identify the most relevant and distinct points in parameter space as outlined in sections 6.3 and 6.3. We first seed G with the parameter space corner cases: $q \in \{1,2\}$, $|\vec{\chi}_A| \in \{0, 0.8\}$, $\theta_\chi \in \{0, \pi\}$ and $\chi_2^z \in \{-0.8, 0.8\}$. As described in Eq. (6.14), we compute training sets $\mathcal{T}_{\mathrm{TS}}^i = \mathcal{T}_{\mathrm{B}} \cup \mathcal{T}_{\mathrm{R}}^i$ consisting of a set of boundary parameters $\mathcal{T}_{\mathrm{B}}$ as well as a set of randomly sampled parameters $\mathcal{T}_{\mathrm{R}}^i$ that is resampled at each greedy iteration i. For $\mathcal{T}_{\mathrm{B}}$, we use a set of 216 points where two components of $\boldsymbol{\lambda}$ take on one of their extremal values and the other two are one of three intermediate (non-boundary) values. This results in features that can be seen in Fig. 6.4, where the two boundary values and three intermediate values occur frequently. For example, because some $2d$ projections of these special points are selected multiple times, they appear as darker points around the boundary of some of the subplots in Fig. 6.4. In addition, subplots involving ϕ_χ show an uneven distribution of stripes that occur at these special points. For $\mathcal{T}_{\mathrm{R}}^i$, we randomly sample each parameter component uniformly in its range.

Next, we add parameters to G using an initial greedy algorithm that uses basis projection errors to select greedy parameters. Given a point $\boldsymbol{\lambda}$ as a candidate that might be added to G, we compute a PN waveform h corresponding to $\boldsymbol{\lambda}$, we decompose h into waveform data pieces (see Sec. 6.5), and we project each waveform data piece onto their respective bases. Then we recombine the projected waveform data pieces to produce a waveform h_{proj}. We then compute an error $\mathcal{E}[h, h_{proj}]$ using Eq. (6.21). The point in $\mathcal{T}^i_{\mathrm{TS}}$ with the largest such error is the next point added to G. This method is different than that of [248], in which projection errors of each waveform data piece were computed separately, and then these errors were combined in a weighted sum with coefficients determined by hand. Our new method avoids the need to determine these coefficients, and automatically ensures that the most significant waveform data pieces are resolved accurately. We use this initial greedy algorithm until the error is $\mathcal{E} \leq 10^{-5}$. At this point, the number of greedy points is approximately $|G| = 30$. Thus we have built a linear basis for each waveform data piece. For each iteration of this initial greedy algorithm, we choose the number of randomly-sampled parameters to be $|\mathcal{T}^i_{\mathrm{R}}| = 10 + 2|G^i|$, where $|G^i|$ is the number of greedy parameters at the start of the ith iteration.

Finally, we add parameters to G using a second greedy algorithm that uses surrogate errors to select greedy parameters. At each iteration i, we construct a new trial PN waveform surrogate (as described in Appendix 6.15), using the greedy parameters G^i, and then for each point $\boldsymbol{\lambda} \in \mathcal{T}^i_{\mathrm{TS}}$, we evalute this surrogate and compare it to the corresponding PN waveform by computing $\mathcal{E}$. The parameter $\boldsymbol{\lambda}$ that maximizes this error is used as the next greedy parameter and is added to G. This error includes the errors in the parametric fits for each empirical node of all waveform data pieces; the parametric fits are shown as blue lines in Fig. 6.1 and are described in detail in Sec. 6.6 and appendices 6.11 and 6.15. For this step, we use $|\mathcal{T}^i_{\mathrm{R}}| = 6|G|$. The maximum errors found in each iteration of this second greedy algorithm are shown in Fig. 6.5 as a function of $|G|$. The noise is due to the random resampling of the training set, as well as the possibility of the parametric fits becoming worse by adding a data point. Because the parametric fits are restricted to a particular order, the surrogate error in Fig. 6.5 does not go below 10^{-3}. In principle one can reduce this error floor by increasing the order of the fits, but here we simply keep only the first 300 greedy parameters. We perform NR simulations for these 300 parameters, except for those parameters that can be obtained from other parameters by symmetry, for example by exchanging the black hole labels. These symmetry considerations reduce the number of simulations to 276.

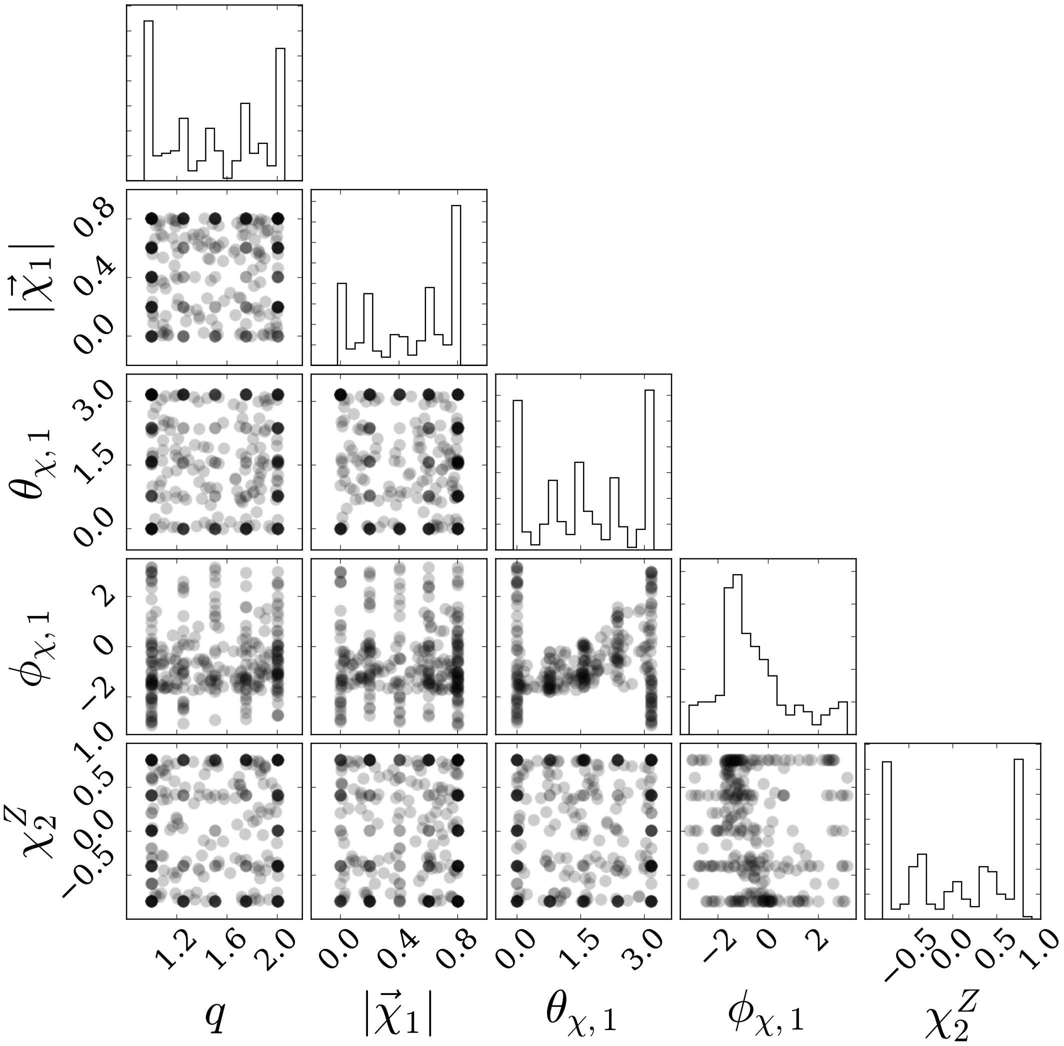

Figure 6.4: A “triangle plot" showing all possible two-dimensional projections and one-dimensional histograms of the greedy parameters G selected by the procedure of Sec. 6.4. These are the parameters used for the numerical relativity simulations. Made using the Python package corner.py [288].

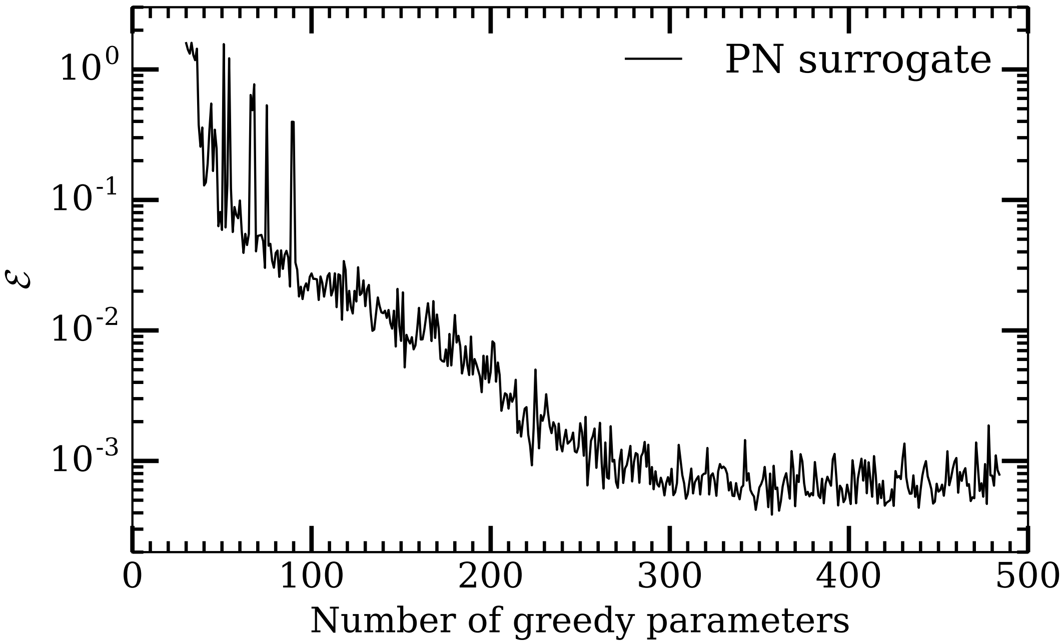

Figure 6.5: Maximum surrogate errors found during the second greedy algorithm (see Sec. 6.4) for determining $\mathbf{\Lambda}_i$ using trial PN surrogates. The noise is due to the random resampling, as well as the possibility of the parameter space fits becoming worse by adding a data point. The finite order of the fits leads to an error floor of 10^{-3}, so we keep and perform NR simulations for only the first 300 greedy parameters.

Numerical Relativity Simulations

To build our time-domain model, we use the 276 NR waveforms computed by the SXS collaboration with the Spectral Einstein Code (SpEC) described in Ref. [285]. Each NR simulation is performed at three different numerical resolutions, labeled 'Lev1', 'Lev2', and 'Lev3', in order of increasing resolution; Lev$i$ has an adaptive-mesh-refinement (AMR) error tolerance that is a factor of 4 smaller than Lev$i-1$. For each resolution, the waveform is extracted at multiple finite radii from the source, and then the waveform is extrapolated to future null infinity [216]. The extrapolation is done using an N-th order polynomial in $1/r$, where r is a radial coordinate. To estimate errors in extrapolation, we perform extrapolation with several values of N [216]. Similarly, to estimate numerical truncation error, we compare simulations that are identical except for resolution [285]. However, for building surrogates, we always use the highest available resolution (Lev3) simulations, and use the $N = 2$ extrapolated waveforms. The simulations begin at a time of $\sim 5000M$ before merger where $M = m_1 + m_2$ and m_i are the Christodoulou masses of each black hole. We ignore the small eccentricities present in the simulations, which have a median of

0.00029 and a maximum of 0.00085 for the highest resolution simulations. The masses we use are those measured after the initial burst of junk radiation [206] leaves the computational domain.

The BH spin vectors are measured on the apparent horizons of the BHs during the evolution of the NR simulation. The spin directions are therefore gauge-dependent The potential concern is that when the surrogate model is evaluated, the spin directions must be provided with the gauge used to build the model, so that the spin directions obtained in gravitational wave parameter estimation can be interpreted correctly. However, it has been found that the time-dependent spin and orbital angular momentum vectors in the damped harmonic gauge used by SpEC agree very well with the corresponding vectors in PN theory [57]. Therefore, this is of no more concern than the interpretation of spin directions with PN-based gravitational wave models.

For the purposes of surrogate modeling, we need to associate each gravitational waveform with a *single* value of the parameter vector $\boldsymbol{\lambda}$, even though some of the parameters (in particular the spin directions) are time-dependent. To do this, we measure the parameters at some fiducial time. To define this time, we (arbitrarily) equate the time coordinate of the simulation with the time coordinate of the waveform at future null infinity, offset so that the beginning of the simulation and the beginning of the NR waveform correspond to the same coordinate t. We then set $t = 0$ at the peak amplitude of the waveform, and we measure $\boldsymbol{\lambda}$ at a fiducial value of $t_0 = -4500M$. We emphasize that there is no unique way to map coordinates in the near zone to coordinates at infinity. However, choosing a different map changes nothing in the surrogate model other than the time at which $\boldsymbol{\lambda}$ is measured. Because the spin directions change only on the precession timescale and not the orbital timescale, any other choice that measures $\boldsymbol{\lambda}$ at a time near the beginning of the simulation should yield similar results.

As described above, we selected the first 300 points in parameter space chosen by the PN greedy algorithms, and we reduced this number to 276 points after removing configurations that were equivalent because of symmetries. We therefore performed 276 NR simulations. However, the total number of NR waveforms represented by these 276 simulations is greater than 276 if we use symmetry to restore additional configurations. For example, for equal mass cases with $\theta_\chi \in \{0, \pi\}$, exchanging the two black holes yields another configuration in the parameter subspace. For each of these cases, we produce the additional configuration by relabeling the black holes and rotating the coordinates by 180 degrees in the orbital plane; this results in a total of 288 NR waveforms. In addition, configurations with $|\vec{\chi}_1| = 0$ are invariant under changes in θ_χ, so we might add additional such configurations that differ only in θ_χ. In principle, we could add an arbitrary number of such configurations, but it is unclear how many to add. Also, $|\vec{\chi}_1|$ is never exactly zero for NR simulations, so we have an unambiguous choice of θ_χ for each simulation. We therefore choose not to restore these additional configurations, so we are left with a total of 288 NR waveforms.

Waveform alignment

Our surrogate model is built assuming that the waveform has peak amplitude at $t = 0$, and that the parameters $\boldsymbol{\lambda}$ (mass ratio and spin vectors) are measured at some fixed time $t = t_0$, which we choose to be $t_0 = -4500M$. Furthermore, our surrogate model assumes a coordinate system in the source frame such that at $t = t_0$, black hole 1 lies along the positive $\hat{x}$ axis, black hole 2 lies along the negative $\hat{x}$ axis, and the instantaneous Newtonian orbital angular momentum lies along the positive $\hat{z}$ axis.

Ideally, all of the input NR waveforms used in the surrogate should also have peak amplitude at $t = 0$, and each simulation's black holes should have the same orientation vector $\hat{n}$ at $t = t_0$, where $\hat{n}$ is a unit vector pointing from the large black hole to the small black hole. However, when setting up an NR simulation, the time between the beginning of the simulation until merger is *a priori* unknown, and depends on the mass ratio and the black hole spins. Furthermore, the orientation $\hat{n}$ of the black holes, and the mass and spin parameters, are chosen at the beginning of the simulation, which (because the merger time is *a priori* unknown) is not at a fixed time before merger. Therefore, for each of our 276 NR waveforms the peak amplitude occurs at a different time, and the orientation of the black holes with respect to the coordinates does not agree at any given time relative to the time of peak amplitude. Therefore, it is necessary to *align* all the NR waveforms by time-shifting them so that the maximum amplitude occurs at $t = 0$, rotating the coordinates so that the black holes are oriented in the same way at $t = t_0$, and then re-measuring the mass and spin parameters at $t = t_0$.

To align the waveforms, we shift them in time such that the peak of the total waveform amplitude as given in Eq. 6.24 occurs at $t = 0$. We then use a cubic spline to interpolate the real and imaginary parts of the waveform onto a uniformly-spaced time series with $dt = 0.1M$. Next, we rotate the waveforms to align the orientation of the binary at $t_0 = -4500M$ in two steps: first we perform an approximate rotation using the black hole trajectories, and then we perform a small correction using only the waveform. For the initial approximate rotation, we use the horizon trajectory to align the Newtonian orbital angular momentum with $\hat{z}$ and rotate about $\hat{z}$ such that black hole 1 lies along the positive $\hat{x}$ axis. We then use the waveform modes to perform an additional rotation, aligning the principal eigenvector of the angular momentum operator [167] with $\hat{z}$ and equating the phases of $h^{2,2}$ and $h^{2,-2}$ at $t = t_0$. The first coarse alignment was used since the second alignment is ambiguous; we can change the sign of the principal eigenvector and/or rotate by an additional π about $\hat{z}$, which we resolve by choosing the smallest of the rotations, since the waveform is already nearly aligned. We perform identical rotations on the spin directions and then measure them at t_0.

Post-alignment parameterization

While the initial orbital parameters were chosen using Post-Newtonian (PN) approximations such that $\vec{\chi}_2(t_0) \propto \hat{z}$ after this alignment, in practice we obtain small misalignments leading to orthogonal components of $\vec{\chi}_2$ less than 0.016 in magnitude. We ignore these spin components, leading to a $5d$ parameter space:

- $q = \frac{m_1}{m_2} \in [0.9999, 2.0005]$
- $|\vec{\chi}_1| \in [0, 0.801]$
- $\theta_\chi \equiv \cos^{-1}\left(\frac{\chi_1^z(t_0)}{|\vec{\chi}_1|}\right) \in [0, \pi]$
- $\phi_\chi \equiv \text{arctan2}\left(\chi_1^y(t_0), \chi_1^x(t_0)\right) \in (-\pi, \pi]$
- $\chi_2^z(t_0) \in [-0.8, 0.800006]$,

as shown in Fig. 6.3. We will often omit the time dependence of the last parameter and simply write χ_2^z.

6.5 Waveform Decomposition

This section describes how each input NR waveform is decomposed into a set of "waveform data pieces", which are simple functions that can be modeled easily and can be recombined to produce the original waveform. This decomposition was outlined briefly in § 6.3, and a flowchart of this process is shown in Fig. 6.6.

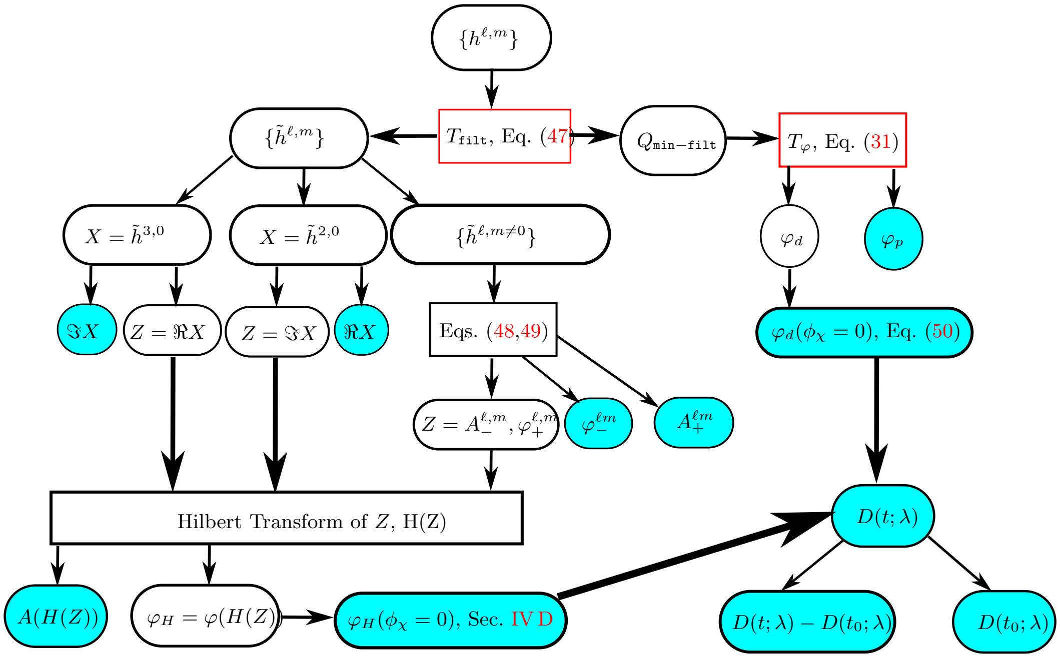

Figure 6.6: Waveform decomposition schematic. A series of decompositions are applied to a set of NR waveform modes $\{h^{\ell m}\}$ yielding easier-to-approximate waveform data pieces (shown as cyan ellipses) for which we ultimately fit. Two types of objects are shown: timeseries data as an ellipse and operators/maps as rectangles. A red outlining border identifies an object which uses a modeling approximation which will *not* go away with additional NR waveforms. These *decomposition errors* are quantified and shown to be smaller than other sources of error in Sec. 6.7. An additional source of error that will not converge away with more NR waveforms results from the assumption that each data piece transform in a simplistic way with changes of ϕ_χ.

We write each input waveform as a set of modes $H = \{h^{\ell,m}(t)\}$, with $t \in [t_{\min}, t_{\max}]$. Here $t_{\min}$ and $t_{\max}$ are chosen to be the same for all waveforms, and are selected in the following way: Recall that each waveform is time-shifted so that the maximum amplitude occurs at $t = 0$; this means that each time-shifted finite-length NR waveform H_i has a different beginning time t_i^{begin} and a different ending time t_i^{end}. We choose $t_{\min} = \max_i(t_i^{begin}) + 150M$ and $t_{\max} = \min_i(t_i^{end})$. The value $150M$ is chosen to remove the worst of the "junk radiation" [206] that results from the failure of NR initial data to precisely describe a quasiequilibrium inspiral. Although the surrogate output will cover only the smaller time interval $[t_0, t_f = 70M]$, we use waveforms over the larger time interval $[t_{\min}, t_{\max}]$ in order to mitigate edge effects that can occur in later steps in the decomposition process (filtering and Hilbert transforms, described below). Selected modes of H are shown in Fig. 6.7.

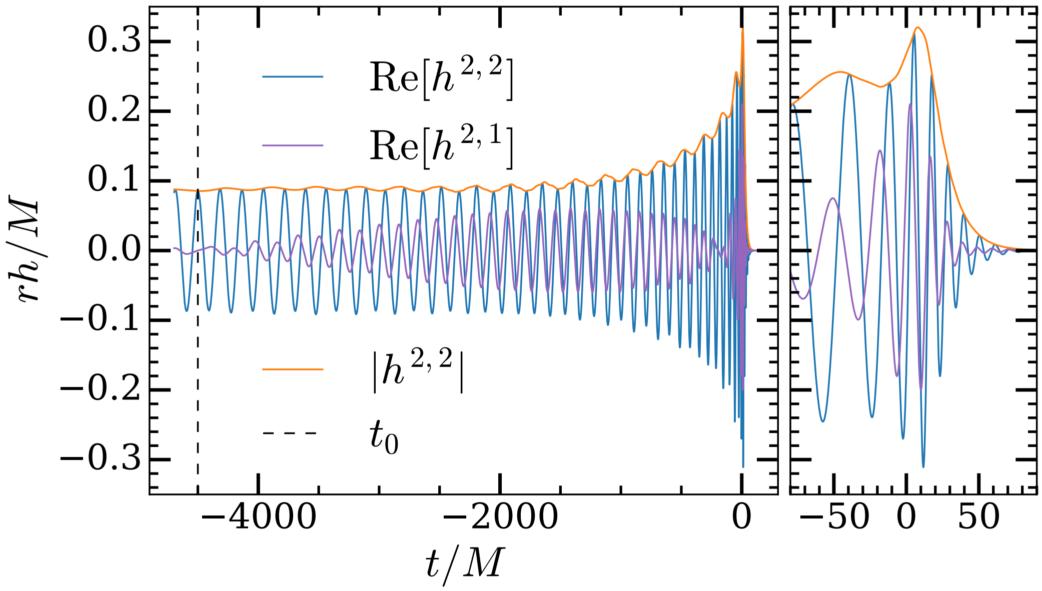

Figure 6.7: Waveform modes in the inertial frame for SXS:BBH:0338 with $q = 2$, $|\vec{\chi}_1| = 0.8$, $\theta_\chi = 1.505$, $\phi_\chi = -1.041$ and $\chi_2^z = 0.8$. For precessing systems, all $\ell = 2$ modes contain significant power in the inertial frame. The NR waveform is aligned to have the canonical orientation at $t = t_0$.

Transforming to a coprecessing frame

The first step in the waveform decomposition is transforming to a rotating coordinate frame in which the binary is not precessing. Thus the original waveform is described by a (much simpler) waveform in this coprecessing frame, plus functions that describe the time-dependent rotation. We transform [3] H to the minimally rotating coprecessing frame of Ref. [167], and thereby obtain the waveform modes $\tilde{H} = \{\tilde{h}^{\ell,m}(t)\}$ in this frame, as well as a time-dependent unit quaternion $q(t)$ that describes the rotation of the frame. Throughout this section we will use a tilde, i.e., $\tilde{h}^{\ell,m}(t)$, to denote a time-domain waveform mode in the coprecessing frame, as opposed to the Fourier transform of a waveform mode. Selected modes of $\tilde{H}$ are shown in Fig. 6.8. We denote this transformation by

$$T_C : H \to (\tilde{H}, q), \tag{6.25}$$

where the ’C’ stands for the coprecessing frame. If we also define a different transformation

$$T_Q : (H', q) \to H \tag{6.26}$$

---

[3] Throughout this work we use GWFrames [168, 283] to enact our transformations.

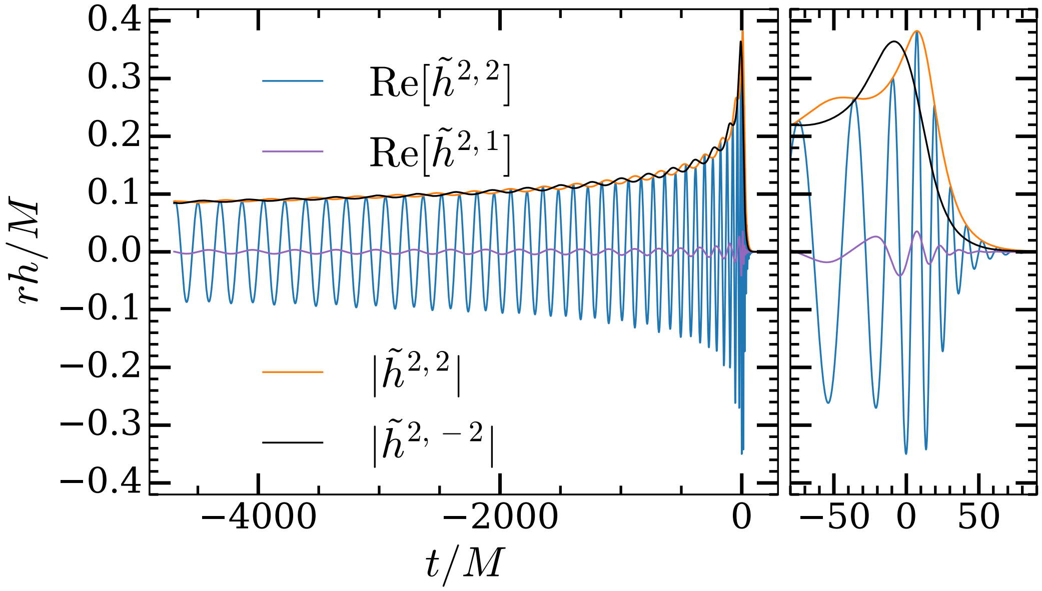

Figure 6.8: Waveform modes in the coprecessing frame for SXS:BBH:0338. The mode power hierarchy is now the same as for a non-precessing waveform, with the $(2, \pm 2)$ modes dominating, but small effects of precession are still present in the mode amplitudes and phases. The amplitudes of the $(2, \pm 2)$ modes have small nearly opposite oscillations.

that takes an arbitrary waveform $H'(t)$ and rotates it by an arbitrary unit quaternion $q(t)$, then T_Q is the left inverse of T_C, that is, $T_Q(T_C(H)) = H$. However, an arbitrary waveform $H'(t)$ and an arbitrary unit quaternion $q(t)$ do not necessarily represent the decomposition of any inertial-frame waveform H into a coprecessing frame. Therefore, for arbitrary $H'(t)$ and $q(t)$ we have in general $T_C(T_Q(H', q)) \neq (H', q)$. This property will be important in § 6.5 below.

The unit quaternion $q(t)$ has four components shown as solid lines in Fig. 6.9. However, the minimally-rotating coprecessing frame constrains $q(t)$ so as to minimize the magnitude of the frame's instantaneous angular velocity (the "minimal rotation condition") [167]. This condition, combined with the unit norm, imply that $q(t)$ has only two independent components.

Therefore, we will further decompose $q(t)$ into these two independent components, so that we have only two functions to model in order to describe the rotation. To do this, consider first the relative instantaneous rotation of the frame

$$dq(t) = q^{-1}(t)q(t + dt) = 1 + 2\vec{\omega}(t)dt + O(dt^2). \tag{6.27}$$

The minimal rotation condition says that $\omega_z = O(dt^2)$, while ω_x and ω_y are $O(1)$, so in the limit $dt \to 0$ we find that $\vec{\omega}(t)$ has only two independent components. The precession angular frequency $\omega_p(t) = |\vec{\omega}(t)|$ describes the velocity of the path on the unit sphere traced out by the z-axis of the coprecessing frame.

We approximate $dq(t)$ using finite differences:

$$\delta q(t) = q^{-1}(t)q(t+\delta t) = s(t) + \delta t\, \vec{u}(t), \tag{6.28}$$

where the scalar component $s(t)$ is $1 + O(\delta t^2)$. Thus, for a given δt, Eq. (6.28) defines $\vec{u}(t)$ in terms of $q(t)$, and furthermore, $\vec{u}(t)$ approaches $\frac{1}{2}\vec{\omega}(t)$ as $\delta t \to 0$. We find that if we use $\delta t = 0.1M$, the $\vec{u}(t)$ we obtain is sufficiently close to this limit that the error we make is negligible compared to other errors; this error is included in the *decomposition error* discussed in § 6.7. Finally, instead of using $\vec{u}(t)$ directly as independent components of $q(t)$, we define $\tilde{\omega}_p = \frac{1}{2}|\vec{u}(t)|$ and

$$\varphi_p(t) = \delta t \sum_{\tau<t} \tilde{\omega}_p(\tau) \tag{6.29}$$

$$\varphi_d(t) = \arg\left(u_x(t) + i u_y(t)\right) . \tag{6.30}$$

The length of the path on the unit sphere traced out by the z-axis of the coprecessing frame is given by $\varphi_p(t)$. In a frame instantaneously aligned with the coprecessing frame, φ_d is the phase of the projection of $\vec{u}(t)$ into the xy-plane.

We have thus decomposed the quaternion $q(t)$ into two functions $\varphi_p(t)$ and $\varphi_d(t)$. These are the two functions we will model in constructing the surrogate. We denote this transformation by

$$T_\varphi : q \to (\varphi_p, \varphi_d). \tag{6.31}$$

To perform the inverse transformation, that is, to compute $q(t)$ from φ_p and φ_d, we compute

$$\tilde{\omega}_p(t) = \frac{\varphi_p(t+dt) - \varphi_p(t)}{\delta t} \tag{6.32}$$

$$u_x(t) = 2\tilde{\omega}_p(t)\cos(\varphi_d(t)) \tag{6.33}$$

$$u_y(t) = 2\tilde{\omega}_p(t)\sin(\varphi_d(t)) \tag{6.34}$$

$$u_z(t) = 0 \tag{6.35}$$

$$s(t) = \sqrt{1 - \left(2\tilde{\omega}_p(t)\delta t\right)^2} \tag{6.36}$$

$$\delta q(t) = s(t) + \vec{u}(t)\delta t\, . \tag{6.37}$$

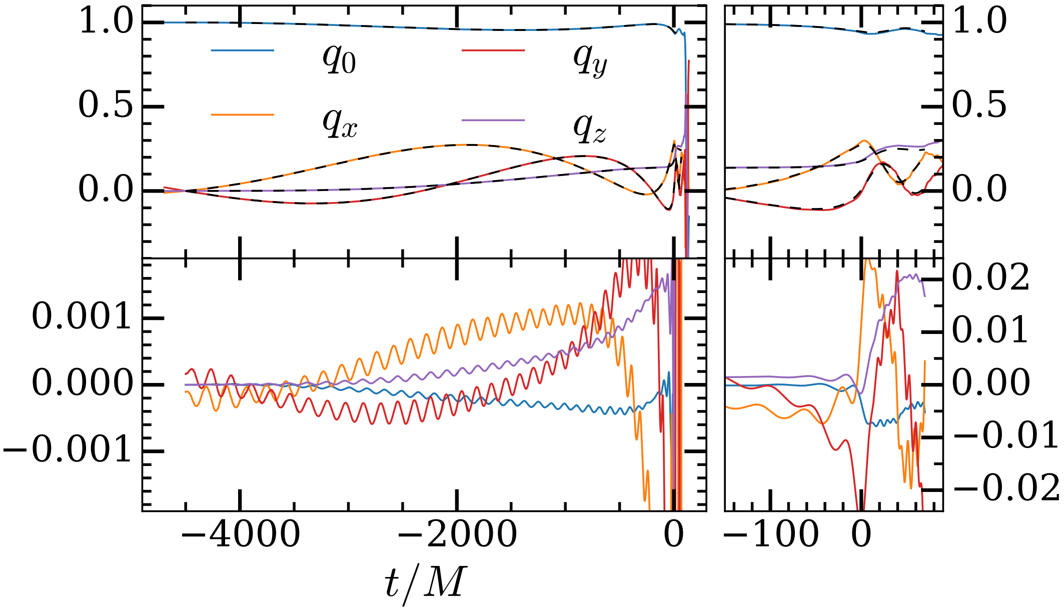

Figure 6.9: Top: Quaternion q representing the time-dependent rotation from the coprecessing frame to the inertial frame (solid lines) and the filtered quaternion $q_{\text{min-filt}}$ (dashed lines) for SXS:BBH:0338. Bottom: Differences between the filtered and unfiltered quaternions. This difference results in an error when reconstructing the waveform in the inertial frame, contributing to a “decomposition" error in the surrogate model.

We include the $(\delta t)^2$ term in Eq. (6.36) so that the reconstructed $q(t)$ has unit norm. Because we assume $\delta q_z = 0$, the δq we compute in Eq. (6.37) is not exactly the δq we started with in Eq. (6.28); however, the error we make is only $O(\delta t^3)$. Given $q(t)$ and δq, we can then compute $q(t + \delta t)$ using

$$q(t + \delta t) = q(t)\delta q(t), \tag{6.38}$$

which results in an $O(\delta t^2)$ error in $q(t + \delta t)$. Because we have $q(t_0) = 1$ at the alignment time t_0, we can use the recurrence relation Eq. (6.38) to construct $q(t)$ at all times, given $\varphi_p(t)$ and $\varphi_d(t)$.

A “filtered" coprecessing frame

The quaternion $q(t)$ representing the coprecessing frame oscillates mostly on the slow precession timescale, which makes it easier to model. However, it also has small oscillations on the much faster orbital timescale, as shown by the purple curve in the bottom plot of Fig. 6.10. These oscillations are due to the nutation of the rotation axis of the coprecessing frame, relative to the inertial frame. These small oscillations can make it more difficult to fit φ_d across parameter space. Since the effect of the nutation on the inertial frame waveform is small, we filter out the nutation in the coprecessing frame. We use a Gaussian filter with a width of π radians of the orbital phase, which is computed from the angular velocity of the waveform as described in [289]. Near the edges of the domain, we truncate the filter on both sides to keep the filter centered. Specifically, if the (monotonic) orbital phase is given by $\varphi_{\rm orb}(t)$, then we can invert the relationship to find $t(\varphi_{\rm orb})$. For a given time τ with corresponding orbital phase $\varphi^* = \varphi_{\rm orb}(\tau)$ we then compute

$$\varphi_{\rm min} = \varphi_{\rm orb}(t_{\rm min}) \tag{6.39}$$

$$\varphi_{\rm max} = \varphi_{\rm orb}(t_{\rm max}) \tag{6.40}$$

$$\Delta\varphi = \min(4\pi, |\varphi^* - \varphi_{\rm min}|, |\varphi^* - \varphi_{\rm max}|) \tag{6.41}$$

$$\varphi_\pm = \varphi^* \pm \Delta\varphi \tag{6.42}$$

$$G(\varphi) = \exp\left[-\left(\frac{\varphi - \varphi^*}{\pi}\right)^2\right] \tag{6.43}$$

$$q_{\rm init-filt}(\tau) = \frac{\int_{\varphi_-}^{\varphi_+} q(t(\varphi))G(\varphi)d\varphi}{\int_{\varphi_-}^{\varphi_+} G(\varphi)d\varphi} \tag{6.44}$$

$$q_{\rm filt}(\tau) = \frac{q_{\rm init-filt}(\tau)}{|q_{\rm init-filt}(\tau)|}\,. \tag{6.45}$$

This filtered frame corresponding to $q_{\rm filt}$ is no longer minimally rotating, but we can compute

$$\tilde{H}', q_{\rm min-filt} = T_C(T_Q(\tilde{H}, q_{\rm filt})) \tag{6.46}$$

and use the frame corresponding to $q_{\rm min-filt}$, which is minimally rotating and has much less nutation than the frame corresponding to q. The components of the filtered quaternion $q_{\rm min-filt}$ are shown in Fig. 6.9 as dashed lines. We use $\tilde{H}$, and not $\tilde{H}'$, as the filtered coprecessing waveform, because $\tilde{H}'$ is not as slowly-varying as $\tilde{H}$ and is therefore slightly more difficult to fit. We have verified that the error in the final model caused by choosing $\tilde{H}$ instead of $\tilde{H}'$ is small compared to other errors. Note that even if we choose $\tilde{H}'$, introducing a filter produces some information loss, and therefore results in some error in the final surrogate model. This *decomposition error* is discussed in section 6.7 and is plotted in Fig. 6.12 and 6.13. We thus denote the filtering transformation by

$$T_{\rm filt} : (\tilde{H}, q) \to (\tilde{H}, q_{\rm min-filt}). \tag{6.47}$$

Applying T_φ to $q_{\rm min-filt}$ results in less oscillatory behavior in φ_d than when T_φ is applied to q, as seen in Fig. 6.10. When evaluating the surrogate and reconstructing

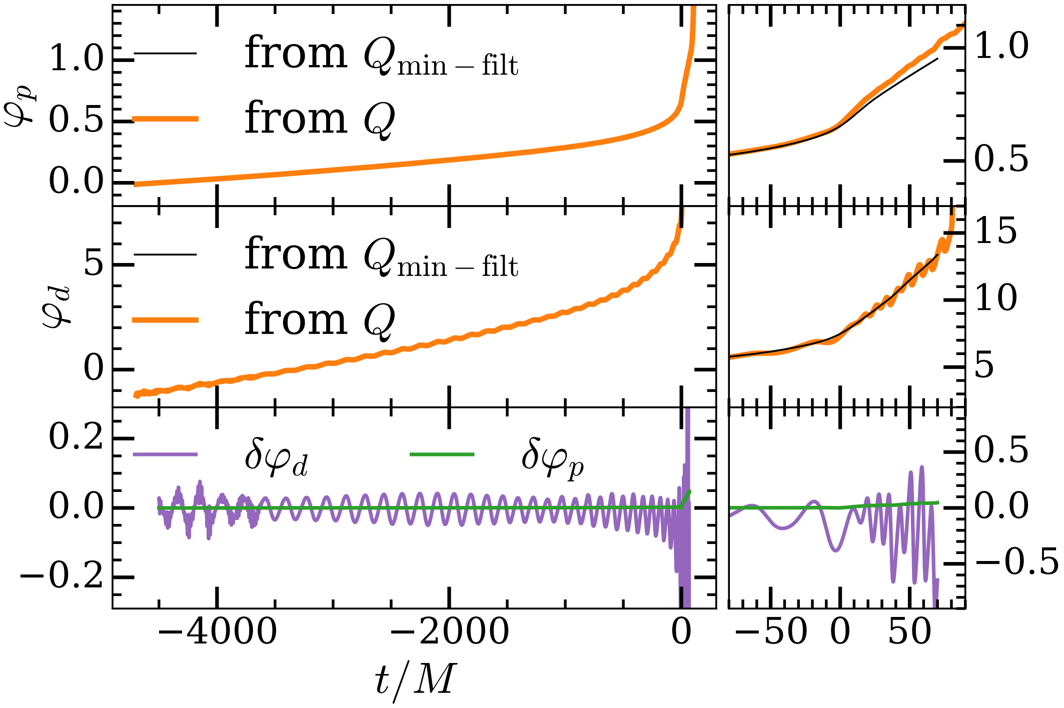

Figure 6.10: Phases φ_p (top) and φ_d (middle) for SXS:BBH:0338. These phases represent the total amount of precession and the instantaneous direction of precession respectively. Shown are phases computed from the unfiltered coprecessing quaternion (thick orange lines) and the filtered quaternion (thin black lines). The orbital timescale oscillation in φ_d is suppressed after filtering. Bottom: Differences between the filtered and unfiltered phases.

the inertial frame waveform, we do not attempt to invert T_{filt}, which contributes to the decomposition errors shown in Fig. 6.13.

Decomposition of coprecessing-frame waveforms

Once we have computed waveform modes $\tilde{h}^{\ell,m}(t)$ in the coprecessing frame, we decompose each of these modes (except for the $m = 0$ modes, which are discussed separately below) into an amplitude and a phase. However, these amplitudes and phases are difficult to model because they contain oscillations on the orbital timescale. These oscillations are due to asymmetries of waveforms from precessing systems and cannot be completely removed with a different choice of frame [286]. Fig. 6.8 shows an example of these oscillations. To better model the amplitudes and phases of $\tilde{h}^{\ell,m}(t)$, we seek to further decompose them into simpler slowly-varying functions. To do this, first note that the amplitudes of $\tilde{h}^{2,2}(t)$ and $\tilde{h}^{2,-2}(t)$ shown in Fig. 6.8 oscillate in opposite directions. The same is true for the phases, although it is not apparent in the figure, and it is also true for some (but not all) higher-order modes. This motivates the use of symmetric and antisymmetric amplitudes and phases

$$A_{\pm}^{\ell,m}(t) = \frac{1}{2}\left(|\tilde{h}^{\ell,m}(t)| \pm |\tilde{h}^{\ell,-m}(t)|\right) \tag{6.48}$$

$$\varphi_{\pm}^{\ell,m}(t) = \frac{1}{2}\left(\varphi(\tilde{h}^{\ell,m}(t)) \pm \varphi(\tilde{h}^{\ell,-m}(t))\right) \tag{6.49}$$

for $m > 0$ where $\varphi(x(t)) = \arg(x(t))$. The symmetric amplitude $A_{+}^{2,\pm2}(t)$ and the antisymmetric phase $\varphi_{-}^{2,\pm2}(t)$ contain almost no oscillations and are slowly-varying, so we use these as waveform data pieces. However, the antisymmetric amplitude $A_{-}^{2,\pm2}(t)$ and the symmetric phase $\varphi_{+}^{2,\pm2}(t)$ of the $(2,\pm2)$ mode are small oscillatory real functions, so to model them we taper each of these functions in the intervals $[t_{\min}, t_0]$ and $[t_f, t_{\max}]$ with a Planck window [220] and take a Hilbert transform, thereby producing an amplitude and phase for each of these functions; these amplitudes and phases are slowly-varying, so we use these as our waveform data pieces.

For subdominant modes, we treat $\varphi_{+}^{\ell,m}$ differently than for the $(2,\pm2)$ modes. We model $\varphi_{+}^{\ell,m}$ directly instead of using a Hilbert transform, because for these modes the Hilbert transform does not improve the model's accuracy. Fortunately, errors in $\varphi_{+}^{\ell,m}$ for $\ell > 2$ contribute very little to the overall error of the final model waveform, as seen in Table 6.3 below.

An additional difficulty is that subdominant modes can vanish at certain points in parameter space, and this makes phases ill-defined. Consider a system with $q = 1$, $|\vec{\chi}_1| = 0$, and some χ_2^z. For $\chi_2^z = 0$, the $(2,1)$ mode vanishes. For small $\chi_2^z \neq 0$, switching the sign of χ_2^z will switch the sign of the $(2,1)$ mode, meaning that the phase of the $(2,1)$ mode has a discontinuity of π as χ_2^z passes through 0. We wish to avoid such discontinuities when building surrogate models. In this particular example, the discontinuity can be avoided by defining the amplitude of the $(2,1)$ mode to be negative and the phase to be increased by π when $\chi_2^z \leq 0$.

Now consider the general case with arbitrary $\vec{\chi}_1$. At the alignment time t_0, the orbits of all NR waveforms are aligned. Because of this, at time t_0 the phase of a given (ℓ,m) mode with $m > 0$ and even will be approximately equal for all NR waveforms, i.e. for all choices of parameters. Similarly, at time t_0 the phase of a given (ℓ,m) mode with m odd will either be approximately equal or will differ by approximately π for all choices of parameters. Therefore at t_0, the phases of each non-vanishing (ℓ,m) mode, for all choices of parameters, are clustered around either one or two values, depending on the mode. Furthermore, when the phases of a given (ℓ, m) mode are clustered around two values instead of one, the clusters are separated by π and the phases of the corresponding $(\ell, -m)$ mode are also clustered around two values and not one. For modes (ℓ, m) with phases that are are clustered around one value, there is no discontinuity in phase as a function of parameters, and nothing more needs to be done. But for modes (ℓ, m) with phases clustered around two values, we remove the discontinuity. To do this, we arbitrarily choose one of the two values as the reference phase $\varphi_0^{\ell,m}$, and then compute the initial phase deviations $\delta\varphi^{\ell,m} = |\varphi(\tilde{h}^{\ell,m}(t_0)) - \varphi_0^{\ell,m}|$. Whenever $\delta\varphi^{\ell,m} + \delta\varphi^{\ell,-m} > \pi$ we take the amplitudes of the $(\ell, \pm m)$ modes to be negative and increase the phases of these modes by π. This causes the initial phase of either the $+m$ or $-m$ mode to be $> \pi$, so we subtract 2π from that phase. These transformations preserve the complex waveform mode $\tilde{h}^{\ell,m}$ but transform $A_\pm^{\ell,m} \to -A_\pm^{\ell,m}$ and $\varphi_-^{\ell,m} \to \varphi_-^{\ell,m} + \pi$, leaving $\varphi_+^{\ell,m}$ unmodified.

Now we discuss modes $\tilde{h}^{\ell,0}(t)$, with $m = 0$. As seen in Fig. 6.11, the $(2,0)$ mode has a non-oscillatory real part during the inspiral, while the imaginary part is small but oscillatory. The $(3,0)$ mode is similar, with the roles of the real and imaginary parts reversed. Therefore, we do not decompose $\tilde{h}^{\ell,0}(t)$ according to Eq. (6.48) and (6.49). Instead, we model the non-oscillatory component directly, and we take a Hilbert transform of the oscillatory component to obtain an amplitude and phase, after tapering that component in the intervals $[t_{\min}, t_0]$ and $[t_f, t_{\max}]$.

The decomposition of the NR waveforms is summarized in Fig. 6.6. The NR waveforms begin at the top of the diagram and are processed going downwards. Each blue endpoint represents one of the slowly-varying waveform data pieces that we fit as a function of parameters $\boldsymbol{\lambda}$ at each of the empirical time nodes. To evaluate the surrogate, the fits and empirical interpolants are evaluated for each of the blue endpoints, and the waveform is reconstructed by going upwards in the diagram and undoing each decomposition, eventually yielding $h^{\ell,m}(t)$.

Removing the dependence on ϕ_χ

As discussed in § 6.4, we attempt to start all NR simulations so that at the reference time $t = t_0$ we have $\phi_\chi = 0$, where ϕ_χ is the azimuthal angle of the spin of the larger black hole, as shown in Fig. 6.3. However, in practice we obtain NR simulations with nonzero values of ϕ_χ at $t = t_0$. In this section we describe how we analytically approximate the dependence of the waveform on ϕ_χ. The surrogate model is then built assuming $\phi_\chi = 0$, so that when the surrogate model predicts waveforms with

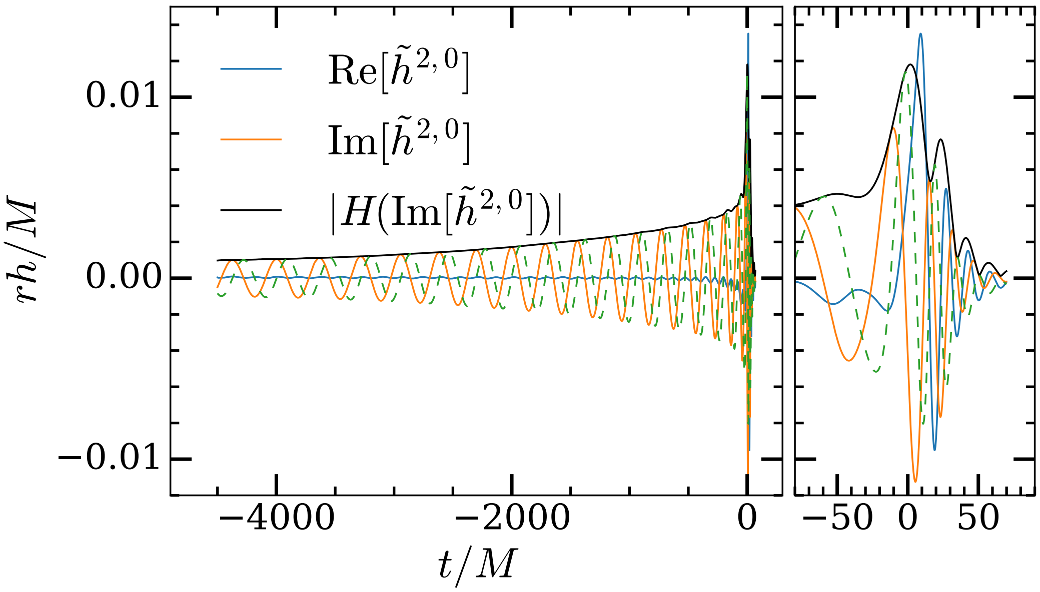

Figure 6.11: For the real-valued oscillatory components X such as $\mathrm{Im}[\tilde{h}^{2,0}]$, we perform a hilbert transform to obtain a complex signal $H(X)$ and extract an amplitude and phase. The dashed green line shows the imaginary part of $H(X)$.

$\phi_\chi \neq 0$, the ϕ_χ dependence is described fully by this analytical approximation. For an orbit-averaged PN waveform of any order that is decomposed into waveform data pieces as described above, it turns out that one can show from the equations (e.g. as written in [57]) that none of the waveform data pieces depend on the parameter ϕ_χ except for the phase $\varphi_d(t)$. This phase has a particularly simple dependence:

$$\varphi_d(t; \boldsymbol{\lambda}, \phi_\chi) = \varphi_d(t; \boldsymbol{\lambda}, 0) + \phi_\chi, \tag{6.50}$$

where $\boldsymbol{\lambda}$ describes all of the parameters except ϕ_χ. So we will make the approximation that Eq. (6.50) applies not only to orbit-averaged PN waveforms, but also to NR waveforms lying within the 5d parameter space. In addition, we find empirically for NR waveforms that the phases of the Hilbert transforms of $A_-^{\ell,m}$ and $\varphi_+^{\ell,m}$ also obey Eq. (6.50), but with the opposite sign on the last term.

Therefore, given a point $\boldsymbol{\lambda}$ in 5d parameter space, we first decompose $h_{\mathrm{NR}}(t; \boldsymbol{\lambda})$ into waveform data pieces, and we then subtract ϕ_χ from φ_d and add ϕ_χ to the phases of the Hilbert transforms of $A_-^{\ell,m}$ and $\varphi_+^{\ell,m}$. We then consider the waveform data pieces as functions of only the 4 parameters (q, $|\vec{\chi}_1|$, χ_2^z, and θ_χ), and we build a 4d model of each of these waveform data pieces. When evaluating the surrogate model waveform at a point $\boldsymbol{\lambda}$ in the full 5d parameter space, we first evaluate the 4d surrogate model expressions for the waveform data pieces at the parameters (q, $|\vec{\chi}_1|$,

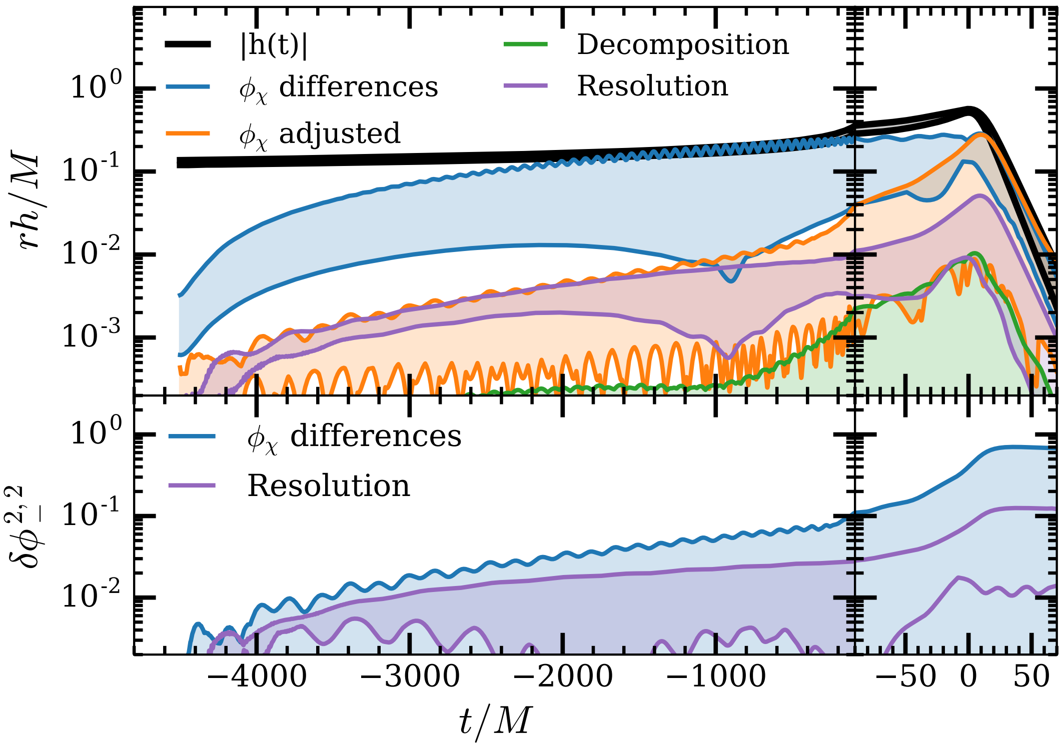

Figure 6.12: Top: waveform differences $\delta h(t)$ investigating the removal of the ϕ_χ dependence on the waveform. Each colored band includes waveforms compared to SXS:BBH:0346 and SXS:BBH:0346 for several different values of ϕ_χ. Before making any adjustment, the errors (ϕ_χ differences) are large. After adjusting, the errors (ϕ_χ adjusted) are comparable to resolution errors during the inspiral but grow large at merger. The decomposition errors are negligible. Bottom: differences in $\varphi_-^{2,2}$. Our analytic approximation to remove the effect of ϕ_χ on the waveform does not affect $\varphi_-^{2,2}$, but here we see that the orbital phase at merger can vary by nearly a radian for different values of ϕ_χ, which is the most significant contribution to the ϕ_χ adjusted errors in the top figure.

χ_2^z, and θ_χ), we add ϕ_χ to φ_d, and we subtract ϕ_χ from the phases of the Hilbert transforms of $A_-^{\ell,m}$ and $\varphi_+^{\ell,m}$. Then we combine the waveform data pieces to yield the model waveform $h_{\mathrm{sur}}(t;\boldsymbol{\lambda})$.

To verify how well this procedure removes the dependence on ϕ_χ, we performed additional SpEC simulations with parameters identical to cases SXS:BBH:0346 and SXS:BBH:0346 but with different values of ϕ_χ. We then analytically remove the ϕ_χ dependence from all these waveforms, as described above, thereby generating $\phi_\chi = 0$ versions of these waveforms, which we compare with each other. The agreement (or lack thereof) of these $\phi_\chi = 0$ waveforms is a measure of the effectiveness of our analytical procedure for removing the ϕ_χ dependence. We find that while the dependence on ϕ_χ is removed well during the inspiral, $\varphi_-^{2,2}(t)$ varies by nearly a radian during the merger as we vary ϕ_χ, which leads to errors significantly larger than the SpEC resolution errors as shown in Fig. 6.12. Incidentally, we note that for two waveforms for which ϕ_χ originally differs by π, the corresponding $\phi_\chi = 0$ waveforms are nearly identical. Before removing ϕ_χ, the largest difference in the waveforms used in this test is $\mathcal{E} = 0.0285$, while after removing ϕ_χ, the largest difference is $\mathcal{E} = 0.00684$. While our ϕ_χ-removal procedure successfully accounts for most of the effect of ϕ_χ, the error associated with this procedure is larger than the median surrogate error (see Fig. 6.13 and Sec. 6.7) and indicates this approximation could be the dominant source of error in the surrogate model.

Handling undefined phases

Our waveform decomposition scheme results in many phases, which become undefined when their corresponding amplitudes vanish. For example, φ_d is undefined for non-precessing systems, as are the phases of the Hilbert transforms of nutating quantities. Additionally, the amplitudes of subdominant modes in the coprecessing frame can briefly become 0, making the corresponding $\varphi_\pm^{\ell,m}$ quantities undefined. Since the NR waveforms contain numerical noise, in practice the phases become poorly resolved when the corresponding amplitude becomes comparable to the noise level.

When decomposing each NR waveform into waveform data pieces, if one of the amplitudes $A(t)$ falls below some threshold at any time t before the merger, then the corresponding phase $\varphi(t)$ is omitted from the model for that NR waveform. This means that when building empirical interpolants or fitting across parameter space at empirical nodes, we use fewer than our entire set of 288 waveforms to fit that particular $\varphi(t)$. The thresholds are described in table 6.1.

6.6 Building a surrogate model from decomposed waveforms

We have decomposed each NR waveform into many functions $X(t;\boldsymbol{\lambda})$ that are smoothly varying as a function of parameters $\boldsymbol{\lambda}$. Here, X represents one of the many decomposed waveform data pieces such as φ_p or $A_+^{2,2}$. Note that while different waveform data pieces X will have different linear basis sizes, empirical time nodes,

Data	Quantity used	Tol	N pass	N reject
φ_d	$\lvert\vec{\chi}_1\rvert\sin(\theta_\chi)$	10^{-3}	192	96
$\varphi[H[A_-^{2,2}]]$	$\lvert H[A_-^{2,2}]]\rvert$	10^{-6}	192	96
$\varphi[H[\varphi_+^{2,2}]]$	$\lvert H[\varphi_+^{2,2}]]\rvert$	10^{-4}	169	119
$\varphi_-^{2,1}$	$A_+^{2,1}$	10^{-4}	260	28
$\varphi_+^{2,1}$	$A_+^{2,1}$	∞	0	288
$\varphi[H[A_-^{2,1}]]$	$\lvert H[A_-^{2,1}]]\rvert$	3×10^{-6}	97	191
$\varphi[H[\mathrm{Im}[\tilde{h}^{2,0}]]]$	$\lvert H[\mathrm{Im}[\tilde{h}^{2,0}]]\rvert$	2×10^{-6}	190	98
$\varphi_\pm^{3,3}$	$A_+^{3,3}$	10^{-3}	210	78
$\varphi[H[A_-^{3,3}]]$	$\lvert H[A_-^{3,3}]]\rvert$	3×10^{-6}	166	122
$\varphi[H[A_-^{3,2}]]$	$\lvert H[A_-^{3,2}]]\rvert$	10^{-6}	140	148
$\varphi_\pm^{3,1}$	$A_+^{3,1}$	10^{-4}	137	151
$\varphi[H[A_-^{3,1}]]$	$\lvert H[A_-^{3,1}]]\rvert$	2×10^{-6}	135	153
$\varphi[H[\mathrm{Im}[\tilde{h}^{3,0}]]]$	$\lvert H[\mathrm{Im}[\tilde{h}^{3,0}]]\rvert$	2×10^{-6}	86	202

Table 6.1: Tolerances used to omit poorly resolved phases. Other than the tolerance for φ_d, which is based on the amount of in-plane spin, the tolerances are based on the minimum value of some amplitude before $t = 0$. If a tolerance is not listed for a particular phase parameter, for example $\varphi_\pm^{\ell,2}$, then that phase parameter is always included in the surrogate. The columns N pass and N reject describe the number of waveforms for which a phase is included in the surrogate, and the number for which it is not. Note that we have a total of 288 waveforms but only 276 NR simulations, because a few of the NR simulations allow us to compute waveforms for more than one set of parameters because of symmetry considerations (cf. § 6.4).

empirical interpolants, and parameter space fits, we will not always label the explicit X dependence of these quantities. For each X we have several NR solutions with different parameters $\{X(t;\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in G^X\}$ where $G^X \subset G = \{\boldsymbol{\Lambda}_i\}_{i=1}^N$. We note that the only reason we might not have $G^X = G$ is due to omitting cases with undefined phases discussed in Sec. 6.5. The next step is to model each of those functions X with its own surrogate model X_S by building an empirical interpolant and fitting the empirical nodes across the parameter space $\mathcal{T}$. The surrogate model for the waveform $h_S^{\ell,m}(t;\boldsymbol{\lambda})$ will then evaluate $X_S(t;\boldsymbol{\lambda})$ for each waveform data piece, from which the inertial frame waveform modes $\{h_S^{\ell,m}(t;\boldsymbol{\lambda})\}$ will be reconstructed. These stages are discussed below.

Empirical Interpolation

For each waveform data piece X, we build an empirical interpolant using the available solutions $\{X(t;\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in G^X\}$. Here we address modifications to the standard

Data	Tol	Data	Tol	Data	Tol	Data	Tol
φ_p	0.005	$\varphi_+^{3,3}$	10.0	$\varphi[H[A_-^{2,2}]]$	0.3	$\varphi_-^{2,2}$	0.15
φ_d	0.03	$\varphi_+^{3,2}$	10.0	$\varphi[H[A_-^{2,1}]]$	1.0	$\varphi_-^{2,1}$	1.0
$\varphi[H[\tilde{h}^{2,0}]]$	0.5	$\varphi_+^{3,1}$	10.0	$\varphi[H[A_-^{3,3}]]$	10.0	$\varphi_-^{3,3}$	0.3
$\varphi[H[\tilde{h}^{3,0}]]$	0.5	$A_+^{2,2}$	0.001	$\varphi[H[A_-^{3,2}]]$	10.0	$\varphi_-^{3,2}$	0.3
$\|H[\varphi_+^{2,2}]\|$	0.15	$A_+^{2,1}$	0.001	$\varphi[H[A_-^{3,1}]]$	1.0	$\varphi_-^{3,1}$	10
$\varphi[H[\varphi_+^{2,2}]]$	10.0	$A_+^{3,2}$	0.0003				

Table 6.2: Projection error RMS tolerances for each basis. Unlisted quantities have a default tolerance of 0.003 for amplitudes and 0.03 for phases.

empirical interpolation method discussed in Sec. 6.3.

We require an orthonormal basis B spanning the space of solutions $\{X(t;\boldsymbol{\lambda}) : \lambda \in \mathcal{T}\}$. While the standard method is to use a reduced basis that was previously constructed when determining the greedy parameters G, in our case we used PN waveforms to find the greedy parameters and have not yet built a basis for NR solutions of X. Greedy and singular value decomposition (SVD) algorithms have been used within the gravitational wave surrogate modelling community [76, 239, 265, 269], and will both provide an accurate basis provided any $X(t,\boldsymbol{\lambda})$ can be accurately approximated in the span of $\{X(t,\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in G^X\}$. A short discussion, including advantages and disadvantages of SVD and greedy algorithms in the context of surrogate waveform modeling, is given in Appendix 6.12. Despite using a greedy sampling strategy to identify the set of greedy parameters, we use a SVD basis for the NR solutions, primarily for its ability to average out uncorrelated noise sources (see Appendix 6.12).

We truncate the orthonormal basis and use the first n singular values and vectors such that all projection errors are below the tolerances given in Table 6.2. We note that n will be different for different waveform data pieces. We then proceed according to Sec. 6.3, finding empirical time nodes $\{T_j\}_{j=1}^n$ and building an empirical interpolant I_n. If we are given X_S at the empirical nodes T_j, we can now determine

$$X_S(t;\boldsymbol{\lambda}) = I_n[X_S](t) \tag{6.51}$$

for all times $t \in [t_{\min}, t_{\max}]$.

Parametric Fits

The next step is to model the dependence on $\boldsymbol{\lambda}$ of the waveform data pieces at the empirical nodes

$$X_j(\boldsymbol{\lambda}) = X(T_j;\boldsymbol{\lambda})\,. \tag{6.52}$$

We build an approximate model for X_j denoted by X_{jS} by fitting it to the available data $\{X_j(\boldsymbol{\lambda}) \,:\, \boldsymbol{\lambda} \in G^X\}$. We do so using a forward-stepwise least-squares fit [290] described in Appendix 6.11, using products of univariate basis functions in q, $|\vec{\chi}_1|$, θ_χ and χ_2^z as the fit features. For each fit, the number of fit coefficients is determined through a cross validation study using 50 trials, each of which uses $N_v = 5$ randomly chosen validation points. The number of fit coefficients used is the one minimizing the sum in quadrature over the error in each trial, which is the maximum fit residual for the validation points.

Complete Surrogate Waveform Model in inertial coordinates

Given parameters $\boldsymbol{\lambda}_5 = (q, |\vec{\chi}_1|, \theta_\chi, \phi_\chi, \chi_2^z)$, we extract $\boldsymbol{\lambda} = (q, |\vec{\chi}_1|, \theta_\chi, \chi_2^z)$ and evaluate the fits and empirical interpolants of each waveform data piece X, obtaining

$$X_S(t;\boldsymbol{\lambda}) = \sum_{j=1}^{n} X_{jS}(\boldsymbol{\lambda}) b^j(t)\,. \tag{6.53}$$

We then obtain the inertial frame waveform $h_S^{\ell,m}(t)$ by combining the waveform data pieces and flowing upwards in Fig. 6.6. Explicitly,

$$\begin{aligned}
Q(t) &= T_\varphi^{-1}\left(\varphi_d(t;\boldsymbol{\lambda}) + \phi_\chi, \varphi_p(t;\boldsymbol{\lambda})\right) &(6.54)\\
\varphi_{\mathcal{I}}^{2,0}(t) &= \varphi(H(\mathcal{I}\tilde{h}^{2,0}))(t;\boldsymbol{\lambda}) - \phi_\chi &(6.55)\\
\mathcal{I}\tilde{h}^{2,0}(t) &= A(H(\mathcal{I}\tilde{h}^{2,0}))(t;\boldsymbol{\lambda})\cos(\varphi_{\mathcal{I}}^{2,0}(t)) &(6.56)\\
\tilde{h}^{2,0}(t) &= \mathcal{R}\tilde{h}^{2,0}(t;\boldsymbol{\lambda}) + i\mathcal{I}\tilde{h}^{2,0}(t) &(6.57)\\
\varphi_{\mathcal{R}}^{3,0}(t) &= \varphi(H(\mathcal{R}\tilde{h}^{3,0}))(t;\boldsymbol{\lambda}) - \phi_\chi &(6.58)\\
\mathcal{R}\tilde{h}^{3,0}(t) &= A(H(\mathcal{R}\tilde{h}^{3,0}))(t;\boldsymbol{\lambda})\cos(\varphi_{\mathcal{R}}^{3,0}(t)) &(6.59)\\
\tilde{h}^{3,0}(t) &= \mathcal{R}\tilde{h}^{3,0}(t) + i\mathcal{I}\tilde{h}^{3,0}(t;\boldsymbol{\lambda}) &(6.60)\\
\varphi^{\ell,\pm m}(t) &= \varphi_+^{\ell,m}(t;\boldsymbol{\lambda}) \pm \varphi_-^{\ell,m}(t;\boldsymbol{\lambda}),\;\; m > 0 &(6.61)\\
A^{\ell,\pm m}(t) &= A_+^{\ell,m}(t;\boldsymbol{\lambda}) \pm A_-^{\ell,m}(t;\boldsymbol{\lambda}),\;\; m > 0 &(6.62)\\
\tilde{h}^{\ell,\pm m}(t) &= A^{\ell,\pm m}(t)\cos(\varphi^{\ell,\pm m}(t)),\;\; m > 0 &(6.63)\\
\{h^{\ell,m}(t)\} &= T_Q(\{\tilde{h}^{\ell,m}(t)\}, Q(t))\,, &(6.64)
\end{aligned}$$

where we have included the dependence on $\boldsymbol{\lambda}$ explicitly for surrogate evaluations of waveform data pieces X_S. The full NRSur4d2s surrogate evaluation producing all $2 \leq \ell \leq 4$ modes for an array of times between $t_{\min}$ and $t_{\max}$ with spacing $\delta t = 0.1$ takes $\sim 1s$ on a single modern processor. Roughly half of this time is spent computing the transformation T_Q from the coprecessing frame to the inertial frame, Eq. (6.26).

6.7 Assessing the Model Errors

Time Domain Errors

To determine how well the output of the NRSur4d2s surrogate matches a NR waveform with the same parameters, we compute

$$\mathcal{E}[h_{\mathrm{NR}}, h_{\mathrm{Sur}}] = \frac{1}{2}\frac{\delta h}{\|h_{\mathrm{NR}}\|^2}, \tag{6.65}$$

where h_{NR} and h_{Sur} are the NR and surrogate waveforms, and δh is given by Eq. (6.19). This quantifies the surrogate error as a whole at one point in parameter space. For NR waveforms that were used to build the surrogate, we call Eq. (6.65) the *training error*. For NR waveforms that were not used to build the surrogate, but are used to test the accuracy of the surrogate model versus NR, we call Eq. (6.65) the *validation error*. Because we decompose each waveform into a set of slowly-varying functions that are modeled independently (i.e., the waveform data pieces of § 6.5), it is useful to consider the contribution to the surrogate error that arises from modeling a single waveform data piece. If X denotes the waveform data piece in question, then we compute this error contribution by decomposing the NR waveform h_{NR} into waveform data pieces, we replace the NR version of X with the surrogate model for X while leaving all waveform data pieces other than X untouched, and we recombine the waveform data pieces, thus producing a waveform we call h_{X}. The error contribution from X is then $\mathcal{E}_X \equiv \mathcal{E}[h_{\mathrm{NR}}, h_{\mathrm{X}}]$. Values of $\mathcal{E}_X$ for various waveform data pieces X are listed in Table 6.3. Note that if we decompose h_{NR} into waveform data pieces and then recompose the waveform data pieces, we do not recover h_{NR} exactly, but instead we get a different waveform $h_{\varnothing}$ because there is error associated with the decomposition. This error, $\mathcal{E}_{\varnothing} \equiv \mathcal{E}[h_{\mathrm{NR}}, h_{\varnothing}]$, is also shown in Table 6.3.

A first test is to verify that the NRSur4d2s surrogate can reproduce the set of NR waveforms from which it was built. The errors for those parameters are shown as the solid blue curve in Fig. 6.13. These errors are significantly larger than the NR

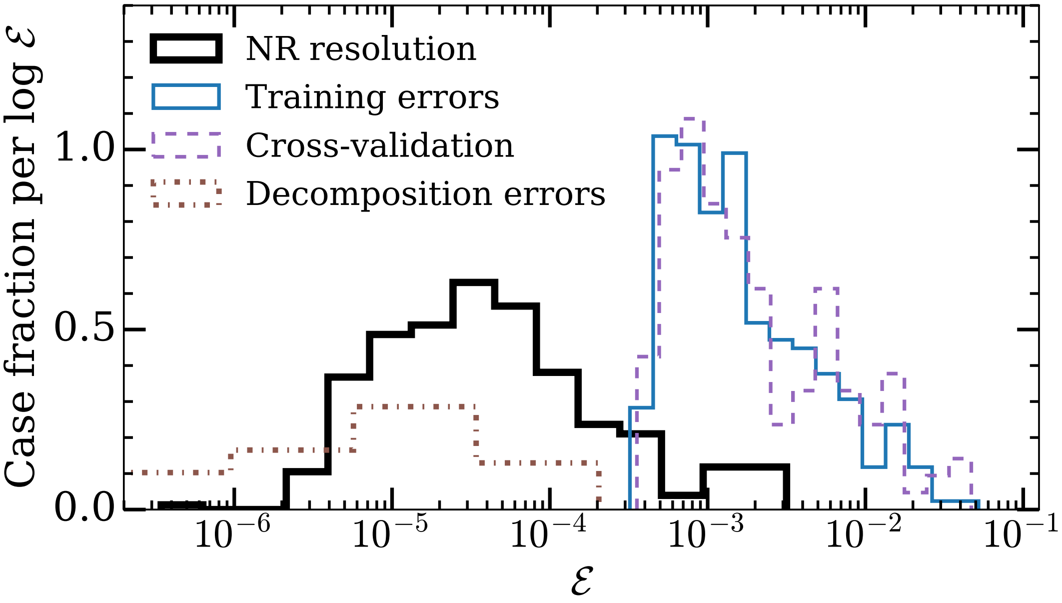

Figure 6.13: Histograms of time domain waveform errors $\mathcal{E}$ relevant to the surrogate. Equal areas under the curves correspond to equal numbers of cases, and the curves are normalized such that the total area under each curve when integrated over $\log_{10}(\mathcal{E})$ is 1. Solid black: The resolution error comparing the highest and second highest resolution NR waveforms. Dotted brown: The error intrinsic to the surrogate's waveform decomposition. Filtering out nutation in the quaternions and neglecting the small but non-zero δq_z due to discrete time sampling leads to errors in the reconstructed waveforms. These errors are nearly zero for non-precessing cases, and even for precessing cases they are smaller than the resolution errors. Thin solid blue: The errors when the full surrogate attempts to reproduce the set of waveforms from which it was built. Dashed purple: The errors when trial surrogates attempt to reproduce NR waveforms that were omitted during the surrogate construction.

resolution errors (cyan curve), which compare the highest and second highest NR resolutions. This indicates either that including additional NR waveforms when building the surrogate model would reduce the training error, or that the error is dominated by approximations made when building the model, such as the analytic treatment of ϕ_χ. The median training error is 0.00136, and in Sec. 6.5 we found that our approximation for the waveform's dependence on ϕ_χ resulted in errors up to 0.00684, indicating the model errors could be dominated by the error in this approximation. While the maximum training error is 0.05212, we only investigated the dependence on ϕ_χ for three cases and only for a few values of ϕ_χ. The parametric dependence of the training errors is illustrated in Fig. 6.14. Perhaps unsurprisingly, the largest errors occur at larger mass ratios and spin magnitudes, and for precessing spin directions.

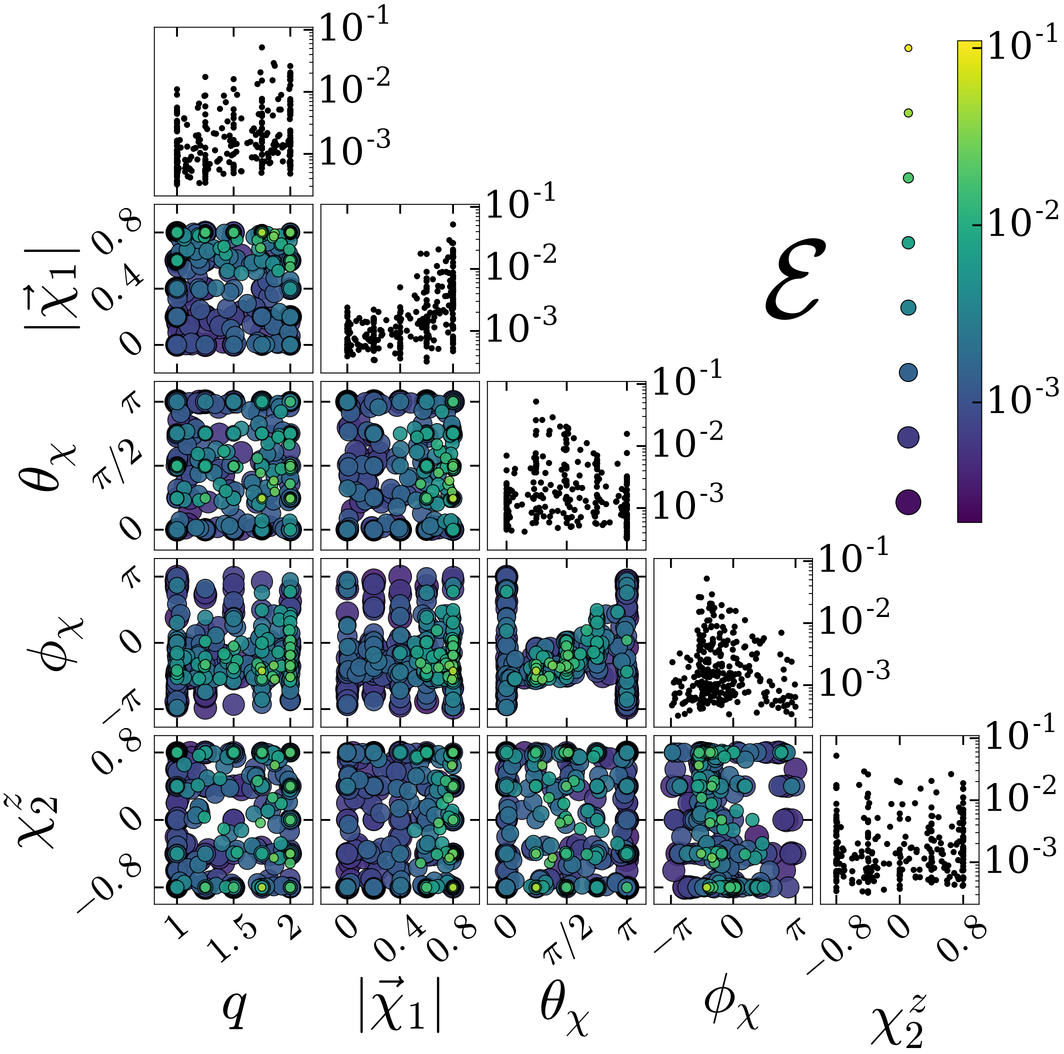

Figure 6.14: Parameter dependence of the error $\mathcal{E}[h, h_S]$ when reproducing the set of NR waveforms with the surrogate. Diagonal: For each parameter plot, the black dots label the (parameter value, $\mathcal{E}[h, h_S]$) pairs. Off-diagonal: For each pair of parameters, we show the $2d$ projection of parameters as in Fig. 6.4 while varying the color and size of the point based on the error $\mathcal{E}[h, h_S]$. Points are placed in order of increasing error, to ensure the small yellow points with large errors are visible. Larger spin magnitudes, especially for precessing spin configurations, correlate with larger errors.

X	$\mathcal{E}_X^0$	$\mathcal{E}_X^{\text{max}}$	$\mathcal{E}_X^{\text{median}}$	X	$\mathcal{E}_X^0$	$\mathcal{E}_X^{\text{max}}$	$\mathcal{E}_X^{\text{median}}$
$\varnothing$	0.0006	0.0006	0.0003	q	0.2450	0.0089	0.0004
h	0.5	0.0521	0.0014	φ_p	0.2450	0.0095	0.0004
$\tilde{h}$	0.5	0.0478	0.0013	φ_d	0.4171	0.0008	0.0003
$\tilde{h}^{2,0}$	0.0006	0.0006	0.0003	$\tilde{h}^{2,\pm 2}$	0.4999	0.0461	0.0011
$\tilde{h}^{2,\pm 1}$	0.0044	0.0016	0.0004	$A_+^{2,2}$	0.4999	0.0007	0.0003
$\tilde{h}^{3,0}$	0.0006	0.0006	0.0003	$A_-^{2,2}$	0.0018	0.0010	0.0003
$\tilde{h}^{3,\pm 1}$	0.0006	0.0006	0.0003	$\varphi_+^{2,2}$	0.0027	0.0049	0.0004
$\tilde{h}^{3,\pm 2}$	0.0008	0.0007	0.0003	$\varphi_-^{2,2}$	0.9959	0.0446	0.0009
$\tilde{h}^{3,\pm 3}$	0.0043	0.0020	0.0004				

Table 6.3: Maximum and median errors when attempting to reproduce the set of NR waveforms when a single waveform data piece is replaced X with its surrogate evaluation X_S and the waveforms are reconstructed. This can be compared with $\mathcal{E}_X^0$, which is the maximum error when replacing X with 0 (or the identity quaternion when $X = q$) instead of with X_S. When $X = \varnothing$ we replace no waveform data piece, but there is still decomposition error due to the lack of $\ell > 3$ modes in the surrogate waveforms, filtering, and neglecting q_z. Note that the errors for $\tilde{h}^{\ell,\pm m}$ include replacing both the (ℓ, m) and $(\ell, -m)$ coprecessing modes. Some components X (such as $X = \tilde{h}^{3,0}$) have $\mathcal{E}_X^0 \sim \mathcal{E}_\varnothing^{\text{max}}$, indicating the error associated with replacing X with 0 is similar to or smaller than the decomposition errors. $\varphi_-^{2,2}$ is the biggest source of error in the surrogate, although φ_p also contributes significantly.

To test the interpolation accuracy of the surrogate, we perform a cross-validation study. For each of ten trials, we randomly select $N_\text{v} = 10$ waveforms which we call *validation waveforms*, and we build a trial surrogate using the remaining $N_\text{t} = N - N_\text{v}$ waveforms. The trial surrogate is evaluated at the N_v validation parameters, and the results are compared to the validation NR waveforms. These validation errors are shown as the purple dashed curve in Fig. 6.13. The validation errors are quite similar to the training errors, indicating we are not overfitting the data.

The maximum and median values of the training errors $\mathcal{E}_X$ are listed in Table 6.3. The decomposition errors $\mathcal{E}_\varnothing$, also shown as the dotted brown curve in Fig. 6.13, are similar or smaller to the NR resolution errors and are therefore negligible. All component errors $\mathcal{E}_X$ include the decomposition errors by construction, and we see that $X = \tilde{h}^{\ell,m}$ leads to negligible errors except for the $(2,2)$, $(2,1)$, and $(3,3)$ modes. The $(2,2)$ mode is the dominant contribution to the error, and its error is dominated by the error in $\varphi_-^{2,2}$. The precession phase φ_p is the dominant precession error, and is the next most significant contribution to the total error in h. Fig. 6.15 shows histograms of the dominant sources of error, and Fig. 6.16 shows the time-dependent errors of these components for the case with the largest training error.

Reference Case	Ecc	$\mathcal{E}$
SXS:BBH:0534	0.000375	0.000007
SXS:BBH:0534	0.002272	0.000162
SXS:BBH:0546	0.000316	0.000004
SXS:BBH:0546	0.000381	0.000005
SXS:BBH:0546	0.002389	0.000106

Table 6.4: Errors $\mathcal{E}[h_0, h_{\rm ecc}]$ where h_0 is the waveform from a reference case used to build the surrogate and $h_{\rm ecc}$ is a waveform from a NR simulation with nearly identical parameters but with a larger eccentricity. For SXS:BBH:0534, h_0 has an eccentricity of 0.000027, and for SXS:BBH:0534, h_0 has an eccentricity of 0.000055.

We have constructed the surrogate models and computed $\mathcal{E}$ assuming zero orbital eccentricity. However, it is not possible to construct NR simulations with exactly zero eccentricity, and the simulations used to build the surrogate have eccentricities of up to 0.00085. To estimate the effect that the eccentricity of the NR waveforms has on our surrogate, we repeated two of our NR simulations changing nothing except the eccentricity. The errors we found are listed in Table 6.4. The largest eccentricities in these additional simulations are several times larger than the maximum eccentricity in the NR simulations used to build the surrogate, yet the resulting waveform errors are smaller than the surrogate errors and comparable to the NR resolution errors. This suggests that the small eccentricities present in the NR waveforms used to build the surrogate are negligible compared to the NR resolution errors.

Frequency-domain comparisons

In this section we compute mismatches in the frequency domain between surrogate waveforms and NR waveforms. To ascertain the significance of these mismatches, we also compute mismatches between two NR waveforms with the same parameters but different resolutions. For comparison, we also compute mismatches between NR waveforms and the phenomenological inspiral-merger-ringdown waveform model IMRPhenomPv2 (which follows the procedure outlined in [66] with IMRPhenomD [65] as the aligned-spin model) and between the effective-one-body model SEOBNRv3 [152], both of which include the effects of precession.

We minimize the frequency domain mismatches over time and polarization angle shifts analytically as described in Appendix 6.14, and also minimize them over orbital phase shifts numerically. When we compare two waveforms, we choose one waveform as the *reference* waveform with fixed parameters, and optimize over

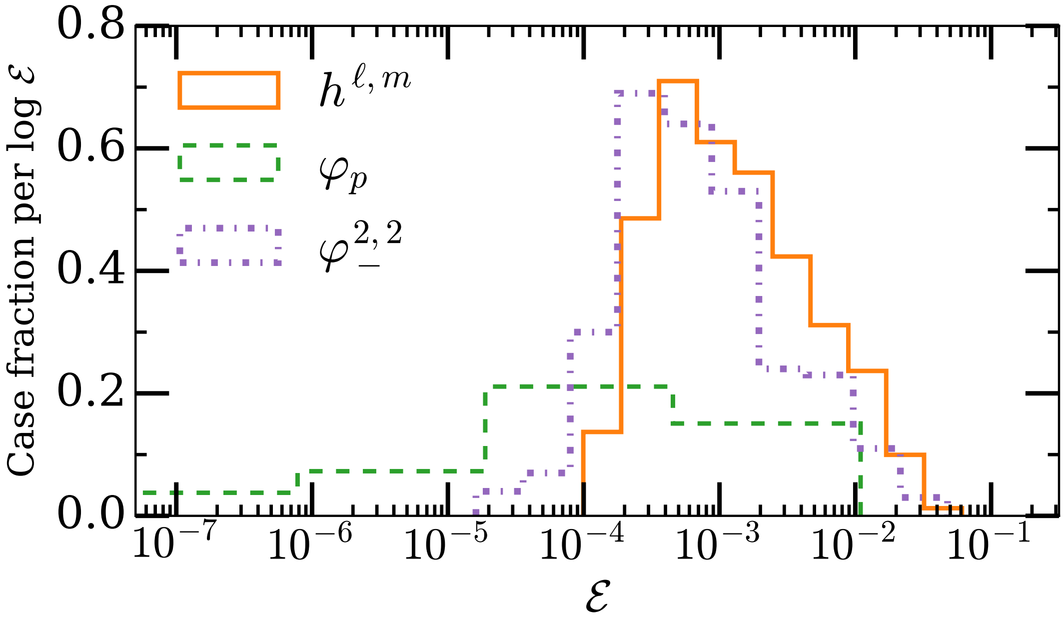

Figure 6.15: Errors $\mathcal{E}_X$ showing the error contribution of a single surrogate component X.

the parameters of the other waveform. When comparing two NR waveforms, the reference waveform is the one with the highest resolution; when comparing NR with some model waveform, the NR waveform is chosen as the reference.

The SEOBNRv3 and IMRPhenomPv2 waveforms are generated with the lalsimulation package [182]. Each SEOBNRv3 waveform is generated in the time domain; the spin directions are specified at the start of the waveform, which is determined by specifying a minimum frequency. We ensure the spin directions are consistent with those of the NR waveforms by varying the minimum frequency in order to obtain a waveform with a peak amplitude occuring $4500M$ after its initial time. The IMRPhenomPv2 waveforms are generated in the frequency domain, and the spin directions are specified at a reference frequency f_{ref} that can be freely chosen. For IMRPhenomPv2 it is not straightforward to determine f_{ref} such that the spin directions are specified at a time of $4500M$ before the peak amplitude. Therefore, we instead choose f_{ref} differently: we minimize the mismatches by varying f_{ref}, with an initial guess of twice the initial orbital frequency of the NR waveform.

To transform the time domain waveforms into the frequency domain, we first taper them using Planck windows[220], rolling on for $t \in [t_0, t_0 + 1000M]$ and rolling off for $t \in [50M, 70M]$ where $t_0 = -4500M$ is the time at which the parameters are measured, and t=0 is the time of peak waveform amplitude. We then pad them with

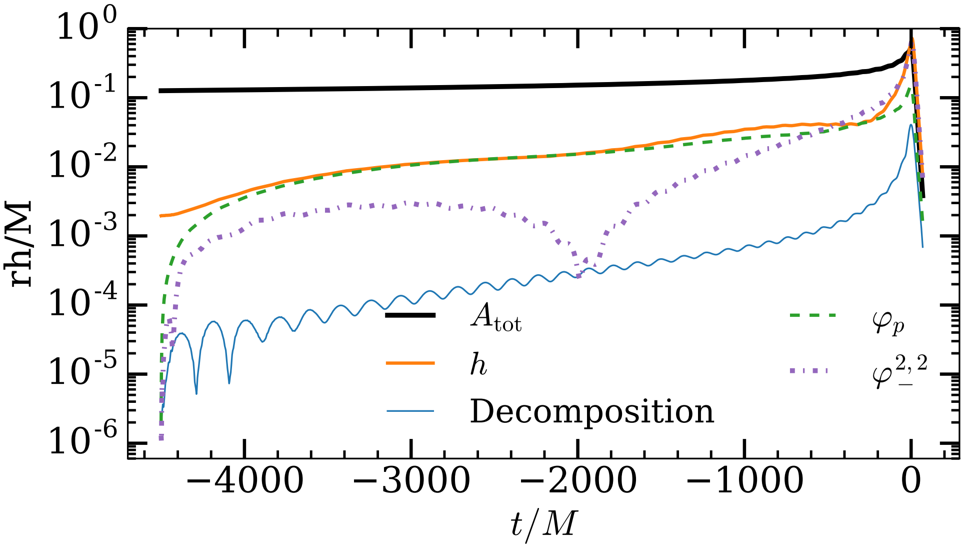

Figure 6.16: Error contributions $\delta h(t)$ of those waveform data pieces X that have the largest error $\mathcal{E}[h, h_X]$ for a selected simulation: ID 79. To compute the error, the NR waveform is decomposed into the surrogate components, and component X is replaced with its surrogate evaluation. The waveform is then reconstructed, and $\delta h(t)$ is computed from Eq. (6.16). The solid black curve is given by Eq. (6.24). The dashed curve is the error in φ_p, which is the dominant error in modeling the precession, and the dominant error source during the inspiral. The dotted curve is the error in a quantity similar to twice the orbital phase, and becomes the dominant error source during the merger and ringdown. The contribution from errors in the other waveform data pieces is smaller, as shown in Table 6.3.

zeros and compute the frequency domain waveforms via the fast Fourier transform (FFT). For the reference NR waveform, we obtain 30 random samples of the direction of gravitational wave propagation (θ, ϕ) from a distribution uniform in $\cos\theta$ and in ϕ, and we uniformly sample the polarization angle ψ between $[0, \pi]$ to obtain

$$h_\psi(t) = h_+(t)\cos(2\psi) + h_\times(t)\sin(2\psi). \tag{6.66}$$

For the non-reference waveform, we use the same parameters except we add an additional initial azimuthal rotation angle ϕ, a polarization angle ψ, and a time offset, and we optimize over these three new parameters to yield a minimum mismatch. Because the waveform models do not intrinsically depend on the total mass, we first use a flat noise curve to evaluate the overlap integrals; this provides a raw comparison between models. We evaluate Eq. 6.23 with $f_{\rm min}$ being twice the orbital frequency of the NR waveform at $t = -3500M$.

The mismatches using a flat noise curve are shown in the top panel of Figure 6.17.

We find that both the IMRPhenomPv2 (green dot-dashed curve) and SEOBNRv3 (solid curve) models have median mismatches of $\sim 10^{-2}$ with the NR waveforms. The mismatches between our surrogate model and the NR waveforms are given by the "Training" (solid blue) and "Validation" (dashed purple) curves and have median mismatches of $\sim 10^{-3}$ with the NR waveforms; see § 6.7 for a discussion of training and validation errors. Finally, NR waveforms of different resolution have median mismatches (solid black curve) of $\sim 10^{-5}$. In the middle and bottom panels, we repeat this study while restricting which *coprecessing-frame* modes are used. IMRPhenomPv2 contains only the $(2, \pm 2)$ modes, while SEOBNRv3 also contains the $(2, \pm 1)$ modes. Obtaining larger mismatches in the top panel when comparing against all NR modes indicates these waveform models would benefit from additional modes. We find that our surrogate performs roughly an order of magnitude better than the other waveform models in its range of validity, but still has mismatches two orders of magnitude larger than the intrinsic resolution error of the NR waveforms. This suggests that the surrogate could be improved with additional waveforms and/or improved model choices. However, we also note that neither IMRPhenomPv2 nor SEOBNRv3 have been calibrated to precessing NR simulations.

Since a realistic noise curve will affect mismatches, we also compute mismatches for total masses M between $20M_\odot$ and $320M_\odot$ using the advanced LIGO design sensitivity [217]. In Fig. 7.4, the lower and upper curves for each waveform model denote the median mismatch and 95th percentile mismatch. We note that for $M < 114_\odot$, some NR and surrogate waveforms begin at $f_{\mathrm{min}} > 10\,\mathrm{Hz}$ and the noise-weighted inner products will not cover the whole advanced LIGO design sensitivity band. The surrogate model errors increase with total mass, indicating a larger amount of error in the merger phase and less error in the inspiral phase. Note that our largest systematic source of error, the approximate treatment of the waveform's dependence on the angle ϕ_χ, is much larger during the merger than during the inspiral, as discussed in § 6.5 and plotted in Fig. 6.12. This error source arises from our attempt to model a 5d parameter space with a 4d surrogate model, so it will not be relevant for a full 7d surrogate model. Even with this error, our surrogate model performs better than the other waveform models up to $320M_\odot$ within the surrogate parameter space.

To determine if the discrepancy between the surrogate errors and NR resolution errors is due to an insufficient number of NR waveforms in the surrogate, we study how the errors depend on the number of waveforms used to build the surrogate. We construct trial surrogates using the first N_{train} NR waveforms for $N_{\mathrm{train}} \in [30, 200]$;

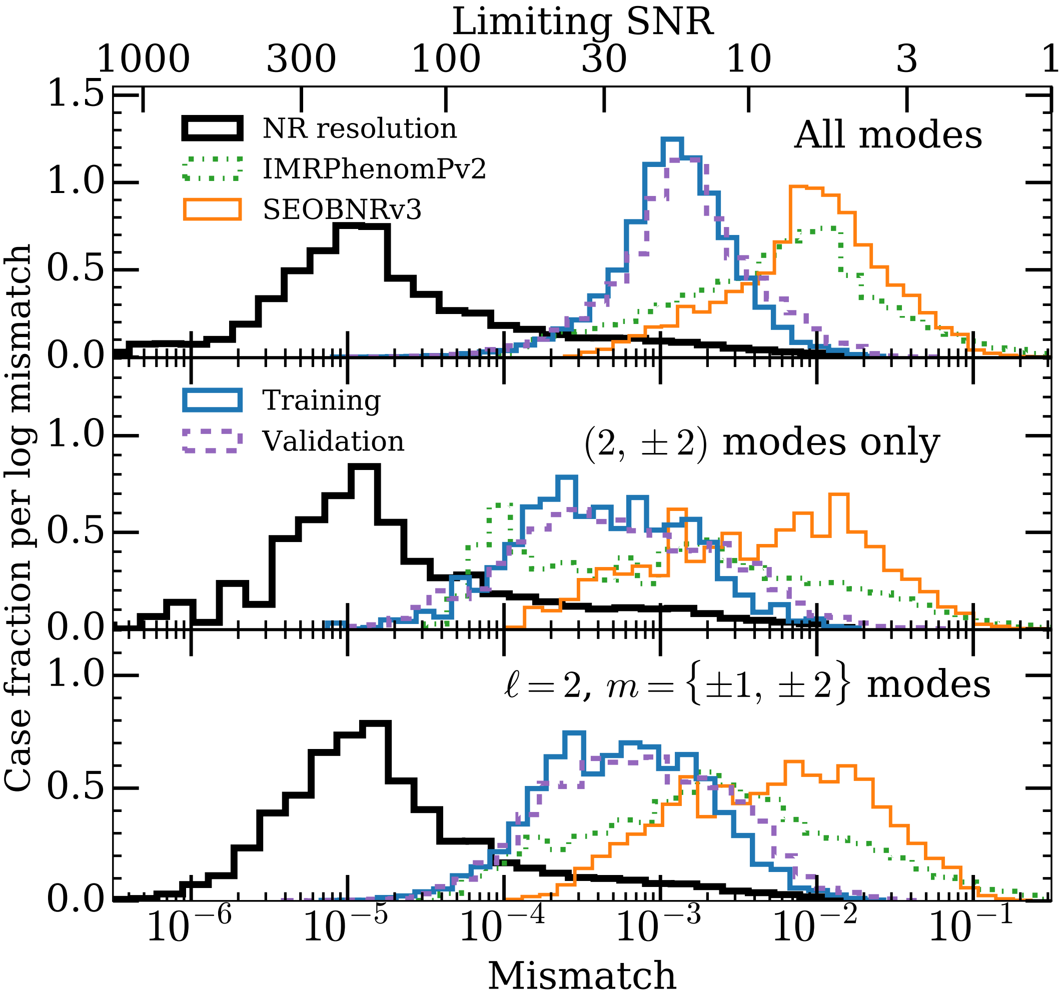

Figure 6.17: Mismatches, computed using a flat noise curve, versus the highest resolution NR waveforms. Histograms are normalized to show the error fraction per log-mismatch, such that the area under each curve is the same. A sufficient but not necessary condition for a mismatch to have a negligible effect is that the signal-to-noise ratio (SNR) lies below the limiting SNR $\rho_* = 1/\sqrt{2\text{Mismatch}}$ given on the top axis [147]. Top: All modes available to each waveform model are included, and the NR waveforms use all $\ell \leq 5$ modes. Middle: All *coprecessing-frame* modes other than $(2, \pm 2)$ are set to zero in all waveforms. Bottom: All *coprecessing-frame* modes other than $(2, \pm 1)$ and $(2, \pm 2)$ are set to zero in all waveforms. These restricted mode studies are done to compare more directly with IMRPhenomPv2 and SEOBNRv3, which retain the coprecessing-frame modes of the middle and bottom panels respectively.

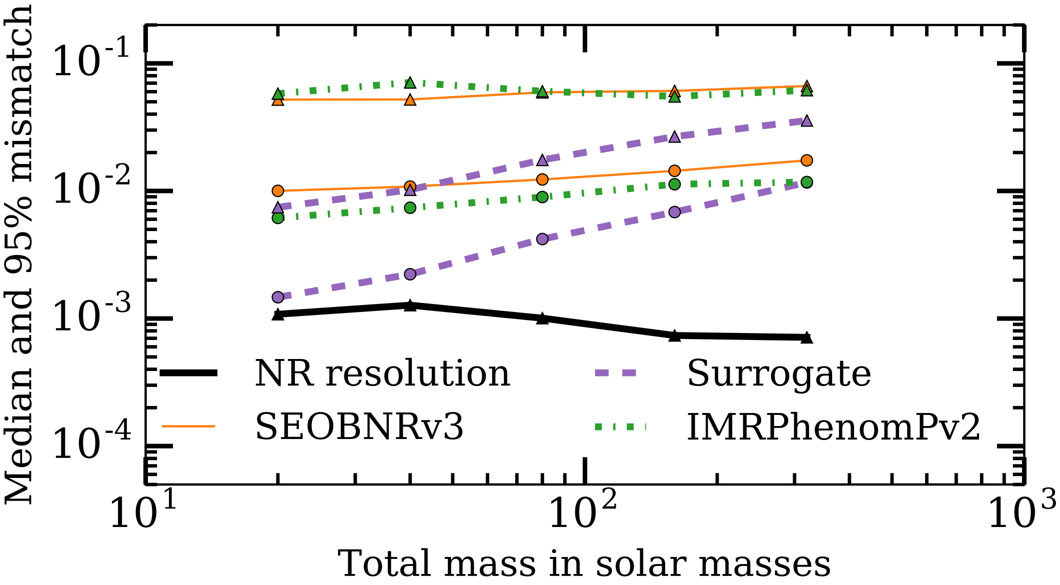

Figure 6.18: Median (lower curves, circles) and 95th percentile (upper curves, triangles) mismatches for various total masses M using the advanced LIGO design sensitivity. The median NR resolution mismatches are all below 2×10^{-5}. The "Surrogate" mismatches shown here are "Validation" errors described in § 6.7.

for validating the surrogate, we use the $N - 200$ waveforms that are not used to build any of these trial surrogates. By using the same $N - 200$ validation waveforms for all choices of N_{train}, we ensure that any changes in the error distribution resulting from changes in N_{train} are due to changes in the surrogate model and not in the set of validation waveforms. The validation errors, shown in Fig. 6.19, decrease quite slowly with additional waveforms when $N_{\text{train}} > 100$, suggesting that the number of NR waveforms would have to increase dramatically to have a noticeable affect on the predictive ability of the surrogate.

Representing arbitrary spin directions

One of the limitations of the NRSur4d2s surrogate model is that it only produces waveforms for binaries with a restricted spin direction on the smaller black hole. However, it is possible to make use of effective spin parameters to create a parameter mapping

$$f : (q, \vec{\chi}_1, \vec{\chi}_2) \to \vec{x}_{\text{model}} \tag{6.67}$$

from the $7d$ space of binaries with arbitrary spin directions to a lower-dimensional parameter subspace [199, 200, 291]. The use of a model with such a parameter space mapping in gravitational wave source parameter estimation leads to equivalence

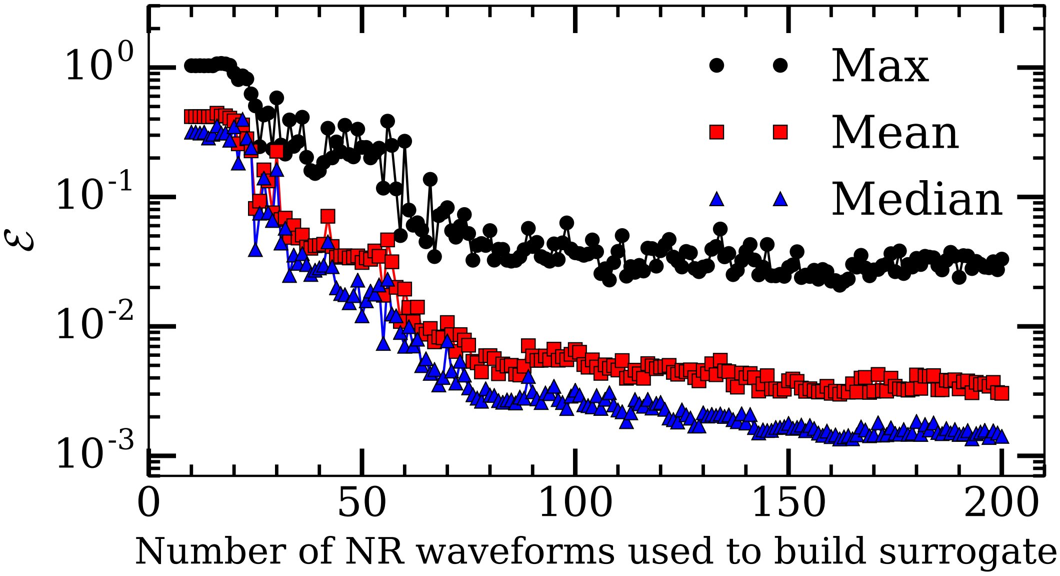

Figure 6.19: Max, mean, and medians of the distributions of $\mathcal{E}$ when building a surrogate using the first N waveforms and a validation set consisting the remaining $200 - N$ waveforms.

classes

$$\{(q, \vec{\chi}_1, \vec{\chi}_2) : f(q, \vec{\chi}_1, \vec{\chi}_2) = \vec{x}_{\rm model}\}\,, \tag{6.68}$$

where multiple values of the 7d parameters map to the same lower-dimensional parameter vector $\vec{x}_{\rm model}$. For parameter estimation, all members of the equivalence class have the same likelihood, so distinguishing parameters within one equivalence class can be done only using knowledge of the prior.

Here we investigate several possible mappings from the full $7d$ parameter space to the $5d$ subspace covered by the NRSur4d2s surrogate model, and we investigate the accuracy of these mappings using 3 SpEC simulations with parameters outside the $5d$ subspace. In our case, $\vec{x}_{\rm model}$ is the vector $(q, \vec{\chi}_1, \chi_2^z)$ at $t = t_0$. To construct a parameter space mapping from $(q, \vec{\chi}_1, \vec{\chi}_2)$ to $\vec{x}_{\rm model}$, we use the values of $\vec{\chi}_1$ and $\vec{\chi}_2$ at $t = t_0$ to form an effective spin $\vec{\chi}_{\rm eff}$, and then construct $\vec{x}_{\rm model}$ using $\vec{\chi}_{\rm eff}$ instead of $\vec{\chi}_1$. This preserves the values of q and χ_2^z, while reducing the other 5 spin components to 3.

The most simple mapping would be to ignore the x and y components of $\vec{\chi}_2$ at $t = t_0$ and take

$$\vec{\chi}_{\rm eff}^{\rm Drop} = \vec{\chi}_1. \tag{6.69}$$

A second possibility would be to use a similar parameter mapping as is used in

SSX:BBH:ID	q	$\vec{\chi}_1$	$\vec{\chi}_2$
0607	1.5	$(0.067, -0.199, 0.212)$	$(0.139, -0.374, 0.202)$
0608	1.7	$(0.053, -0.085, 0.001)$	$(0.494, 0.337, 0.113)$
0609	1.9	$(0.094, -0.145, 0.099)$	$(-0.398, 0.576, 0.001)$

Table 6.5: Parameters for three additional SpEC simulations with unrestricted spin directions. The spins are measured at $t = t_0$.

IMRPhenomP [66] with an effective precessing spin χ_p [291] and take

$$B_1 = \left(2 + \frac{3}{2q}\right)\left(\frac{q}{1+q}\right)^2, \tag{6.70}$$

$$B_2 = \left(2 + \frac{3q}{2}\right)\left(\frac{1}{1+q}\right)^2, \tag{6.71}$$

$$i^* = \underset{i=1,2}{\arg\max}\, B_i \|\vec{\chi}_i^\perp\|, \tag{6.72}$$

$$\vec{\chi}_\mathrm{eff}^{\chi_\mathrm{p}} = \frac{B_{i^*}}{B_1}\vec{\chi}_{i^*}^\perp + \chi_1^z \hat{z}, \tag{6.73}$$

where $\vec{\chi}_i^\perp$ is the part of $\vec{\chi}_i$ orthogonal to the Newtonian orbital angular momentum, which is $(\chi_i^x, \chi_i^y, 0)$ at $t = t_0$. This mapping uses the in-plane spin components of whichever spin contributes the most to precession at leading PN order, scaled appropriately and placed on the heavier black hole. This mapping is particularly effective when the in-plane spins of the smaller BH are negligible, i.e., for high mass ratios, and for long duration GWs. However, it has also been shown to prove sufficient for binaries similar to GW150914 [20, 263].

In our case, we have a couple precession cycles at most, and we might consider adding the effects of the in-plane components of the two spins. A further motivation to add the spins is that for nearly equal masses, the precession rates of the two spins will be nearly equal [279, 292]. When adding the dimensionless spins, we can either do so directly,

$$\vec{\chi}_\mathrm{eff}^\mathrm{Add} = \vec{\chi}_1 + \frac{1}{q^2}\vec{\chi}_2^\perp, \tag{6.74}$$

or again using the leading order PN contribution to precession,

$$\vec{\chi}_\mathrm{eff}^\mathrm{PN} = \vec{\chi}_1 + \frac{B_2}{B_1}\vec{\chi}_2^\perp. \tag{6.75}$$

We do a brief investigation of the quality of these parameter space mappings using three additional SpEC simulations. The waveforms are aligned as described in Sec. 6.4, and their parameters at $t = t_0$ are measured and listed in Table 6.5. For each

Map	$\mathcal{E}$			Median Mismatch		
	0607	0608	0609	0607	0608	0609
Drop (Eq. 6.69)	**0.016**	0.007	0.008	**0.0054**	0.0026	0.0031
Add (Eq. 6.74)	0.007	0.009	**0.013**	0.0046	0.0051	**0.0076**
PN (Eq. 6.75)	0.008	0.019	**0.021**	0.0041	0.0075	**0.0109**
χ_{p} (Eq. 6.73)	0.014	0.018	**0.044**	0.0050	0.0074	**0.0161**

Table 6.6: Errors between the three NR waveforms and the surrogate evaluation for a given parameter space mapping. Mismatches are optimized over time, polarization angle, and orbital phase shifts. For each mapping, the largest error is in bold.

case and each parameter space mapping, we compute the mapped parameters and compare the surrogate evaluation with the mapped parameters to the NR waveform. The time-dependent waveform errors are shown in Fig. 6.20 and $\mathcal{E}$ values as well as mismatches are given in Table 6.6. ~ 0.01, which is larger than the median surrogate errors but well within the possible range of surrogate errors, so we cannot rule out that these errors are dominated by surrogate error. The “Drop" parameter space mapping performs reasonably well since the cases investigated are far enough away from equal mass that the spin of the smaller black hole has a small effect on the waveform.

6.8 Building the Frequency Domain Surrogate

Evaluating the NRSur4d2s surrogate takes $\sim 1s$ on a single modern processor. Evaluating all coprecessing modes takes $\sim 0.21s$, evaluating the frame quaternions $q(t)$ takes $\sim 0.38s$ and is dominated by evaluating Eq. 6.38 sequentially for all times, and rotating the modes into the inertial frame with the transformation T_Q takes $\sim 0.41s$. Gravitational wave parameter estimation is typically done using Markov-chain Monte Carlo [187] and can require $\mathcal{O}(10^8)$ waveform evaluations; this motivates us to build a faster surrogate model. We also wish the faster surrogate model to be in the frequency domain, where most parameter estimation is currently done. Accelerated frequency-domain surrogates have been built in $3d$ [239, 265] using cubic tensor-spline interpolation of the waveform amplitudes and phases at some sparsely sampled frequency points.

To build the frequency-domain NRSur4d2s_FDROM surrogate, we first choose a uniformly spaced grid of $N = N_q \times \ldots \times N_{\chi_2^z}$ points in our $5d$ parameter space and evaluate the NRSur4d2s surrogate model at each point on the grid. We taper the waveforms with Planck windows [220], rolling on for $t \in [-4500M, -3500M]$ and

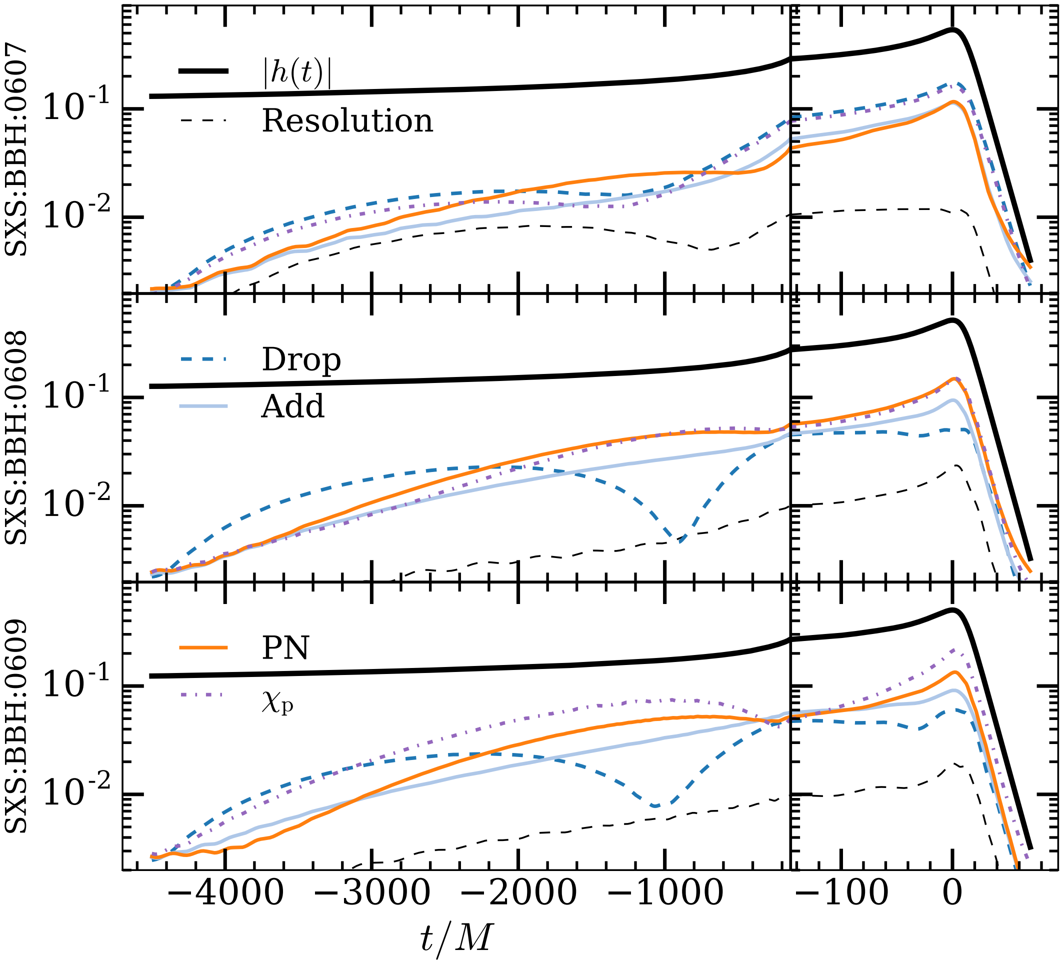

Figure 6.20: Comparing surrogate evaluations to three NR waveforms (top, middle and bottom plots) with spins outside the $5d$ parameter subspace. For each case and each of the four parameter mappings, the surrogate model waveform error is shown. In all cases, the ‘PN’ mapping performs well at the very start of the waveform and the ‘Drop’ mapping performs poorly, but there is no clear overall best mapping. Surrogate modeling errors contribute to the difficulty in assessing the quality of the mappings.

Grid label	N_q	$N_{\lvert\chi_1\rvert}$	N_{θ_χ}	N_{ϕ_χ}	$N_{\chi_2^z}$
5	5	4	7	4	6
6	6	4	8	4	7
7	7	5	9	4	8
8	8	6	11	4	9
9	9	6	13	5	11
10	10	7	14	6	12
11	11	8	15	7	14
12	12	9	17	8	16
13	13	10	19	9	19

Table 6.7: Grid sizes for tensor-spline interpolation in the frequency-domain surrogate. The size in each dimension is chosen such that surrogates for $1d$ slices in all dimensions have comparable interpolation errors.

rolling off for $t \in [50M, 70M]$. We then pad the waveform modes with zeros and perform a fast Fourier transform to obtain the frequency domain modes $\tilde{h}^{\ell,m}(f)$. We then downsample the frequency domain waveforms to a non-uniformly spaced set of frequencies, which are chosen to be the same for all waveforms and to be uniformly spaced in gravitational-wave phase for an equal-mass zero-spin binary. This significantly reduces the cost of evaluating the model, with a negligible loss in accuracy. For each mode $\tilde{h}^{\ell,m}(f)$, we build an empirical interpolant in frequency using all N waveforms, and we keep the first 100 basis vectors. At each empirical frequency node, we fit the real and imagniary parts of each mode across parameter space using a cubic tensor-product spline; we use “not-a-knot” boundary conditions that have a constant third derivative across the first and last knots [293]. Finding the spline coefficients involves solving a sparse linear system of size $(N_q+2)\times\ldots\times(N_{\chi_2^z}+2)$, for which we used Suitesparse [294, 295] and/or SuperLU DIST [296, 297]. The advantage of using a spline is that the evaluation cost is nearly independent of the grid size N, and requires only $4^{d=5}$ coefficients and basis functions to be evaluated.

Implementing the NRSur4d2s_FDROM surrogate model in both C and Python, we find it takes 50ms to evaluate a single waveform in either case. Empirical interpolation accounts for roughly 10% of the cost, and the remaining 90% comes from to the 2400 spline evaluations. Assembling the waveform at a desired sky direction from the modes and interpolating onto the desired frequencies have negligible cost.

To ensure that the empirical interpolants and parameter space splines are sufficiently accurate, we construct many frequency-domain surrogates for increasingly large parameter space grids. We monitor the differences between the frequency domain

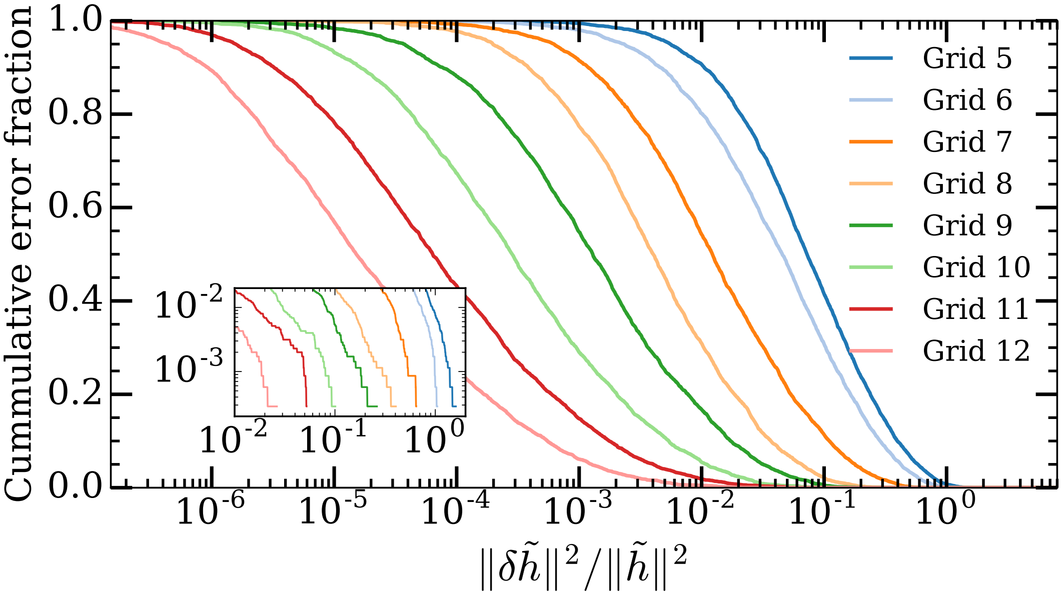

Figure 6.21: Cumulative error distributions of the frequency domain NRSur4d2s_FDROM surrogate waveforms compared to the time domain NRSur4d2s surrogate waveforms transformed to the frequency domain, evaluated for randomly chosen uniformly distributed parameters. The curves indicate the fraction of errors at least as large as the indicated error. The NRSur4d2s_FDROM output converges to the FFT of the NRSur4d2s output as the grid size is increased.

surrogate waveforms and the FFT of the tapered NRSur4d2s waveforms, and we demand that these differences decrease with increasing grid size. We use a different number of grid points in each parameter-space dimension, since the waveforms vary more in some dimensions than others. To determine the number of grid points to use, we construct frequency-domain surrogates for $1d$ slices of the parameter space, where the other parameters are fixed at a single intermediate value. We then arbitrarily choose a value of N_q, the number of grid points covering the dimension of mass ratio, and we determine the maximum error of the $1d$ surrogate in which only the mass ratio q is varied. Call this error E_q. Then we find the number of points $N_{|\chi_1|}$ for which the $1d$ surrogate for $|\chi_1|$ has an error of approximately E_q, and similarly for the other parameters. The resulting grid sizes are listed in table 6.7. In Fig. 6.21, we see that the errors converge as the grid size increases.

6.9 Discussion

We have built the first NR surrogate model of BBH waveforms that covers a multi-dimensional portion of the BBH parameter space. This extends the work in [269], where a 1-dimensional (i.e. zero spin) NR surrogate served as a proof of principle that surrogate models of NR waveforms can be made highly accurate. The non-spinning surrogate model is inappropriate for use in GW parameter estimation, as neglecting all spin effects could lead to large parameter biases. Extending the parameter space to include both aligned spin components and one precessing component makes the new model presented here the first NR surrogate suitable for gravitational wave parameter estimation. While two of the in-plane spin components are still neglected by the NRSur4d2s surrogate model, IMRPhenomPv2 neglects similar information but obtains parameters for GW150914 that are compatible with those obtained using SEOBNRv3, which includes all spin components [263]. We note, however, that for edge-on systems otherwise similar to GW150914 IMRPhenomPv2 can obtain biased parameter estimates [268].

To reduce computational cost, the simulations used to build the NRSur4d2s surrogate were restricted to mass ratios $q \leq 2$ and spin magnitudes $|\vec{\chi}_i| \leq 0.8$. This limits the range of GW events for which the surrogate model could be used. GW150914 is within this range, while the mass ratio posterior of GW151226 extends well beyond $q = 2$. Ultimately, a NR surrogate model covering the fully precessing $7d$ parameter space up to large mass ratios and spin magnitudes will be needed.

Use of the NRSur4d2s surrogate is also limited by the length (i.e. number of orbits) of the waveforms used to build it. GW151226 enters the sensitive LIGO band approximately 55 cycles before merger [251], while the NRSur4d2s surrogate produces waveforms with between 30 and 40 cycles before merger. Since these waveforms are tapered before building the faster NRSur4d2s_FDROM surrogate, the latter includes only 25 to 35 cycles before merger. There are a few ways to build an NR surrogate with longer waveforms, so that the surrogate is applicable to GW events of lower total mass. First, one could build a surrogate model using longer NR waveforms. A less computational expensive option would be to hybridize [215, 298–300] the NR waveforms with PN or EOB waveforms before building a surrogate model. A final option would be to use a time domain surrogate which produces waveforms of moderate length as done here, to hybridize the surrogate output with PN or EOB waveforms before transforming them into the frequency domain, and finally to build a frequency domain surrogate for the hybrid waveforms.

Phenomenological and semi-analytic waveform modeling approaches have already led to precessing waveform models suitable for GW parameter estimation from a large class of GW events. These models have an underlying structure, and are calibrated by tuning a set of numerical coefficients such that the model waveforms have good agreement with NR waveforms. NR surrogate models provide an independent approach. NR surrogate models make no assumptions about the waveform structure, although knowledge of the waveform structure may lead to a better decomposition and smaller errors for a given number of input NR waveforms. We find our NRSur4d2s surrogate model to have better agreement with NR waveforms than other leading waveform models within the range of validity of the surrogate, although we again note that these other models have not been calibrated to precessing NR simulations. As gravitational wave detector sensitivities improve, this increased waveform accuracy will become important for unbiased measurements of the parameters from the loudest GW events, as well as when making astrophysical statements using many GW events.

Since we have not performed Cauchy characteristic extraction [138, 241, 242, 301], but instead have extracted waveforms from the simulations at a series of finite radii and then extrapolated them to infinite radius [216], the $(2,0)$ modes of the numerical waveforms in the coprecessing frame may not be accurate [213]. In particular, we do not see the expected gravitational wave memory in the real part of the $(2,0)$ mode [302, 303]. This should lead to negligible errors for most LIGO purposes, since the memory signal is low frequency and has very little contribution within the LIGO band. However, NRSur4d2s would not be suitable to detect a memory signal with a method requiring templates that include memory. A direct measurement of the memory signal using the method proposed in [319], however, could make use of waveforms from NRSur4d2s, as they have the $(2,\pm 1)$ and $(3,\pm 3)$ modes in the coprecessing frame necessary to determine the sign of the memory.

The errors in the NRSur4d2s surrogate are significantly larger than the resolution of the NR waveforms used in its construction. An incomplete treatment of the spin angle φ_χ (see Fig. 6.3) is one large source of error, and a complete $7d$ NR surrogate model would not suffer from this issue. Aligning the rotation of the waveforms (see § 6.4) closer to merger might reduce the errors, since $\varphi_-^{2,2}$ at the empirical nodes would have less variation across parameter space. Since the parameters of the NR simulations were chosen such that $\vec{\chi}_2$ is aligned with the orbital angular momentum $4500M$ before merger, it would be non-trivial to build a surrogate model from these NR waveforms if the rotation alignment were performed at some other time. This is another issue which will be resolved by including all 7 dimensions of parameter space.

Incorporating additional NR waveforms into the NRSur4d2s surrogate should also reduce the surrogate errors, although Fig. 6.19 indicates that with the current surrogate choices a very large number of additional NR waveforms would be needed for a significant reduction. Alternative methods of fitting empirical nodes could also help. The training and validation errors in Fig. 6.17 and 6.13 are nearly identical, while in [269] the validation errors were roughly a factor of 2 larger than the training errors. This suggests we may be under fitting the data and could use tighter parameter space fit tolerances.

In addition to model cross-validation, there is a variety of informative diagnostics we could monitor to diagnose sources of surrogate error. Failing to meet one of these diagnostics would indicate an unexpected source of surrogate error that could be improved:

- *Decay of the temporal basis error*. Smooth models are expected to have an exponentially decaying basis projection error and empirical interpolation error. Numerical noise in the NR waveforms means the exponential decay will not continue to arbitrarily small errors, but if the error curves do not display a region of exponential decay there is reason to suspect the basis is not accurate enough.

- *Decay of the parametric fitting error*. It is known that expanding (with orthogonal projection) a smooth function with polynomials results in an exponentially decaying approximation error. We believe the waveform data pieces evaluated at empirical nodes can be described by a smooth function plus (relatively small) noise. Thus, just as in the case of the basis projection error, the fitting error is expected to decay exponentially before the noise sources dominate the approximation. This can be seen in Fig. 6.22, where the exponential decay only lasts for approximately 10 coefficients before noise sources cause the validation errors to flatten and then slowly rise.

- *Robustness to noise*. We could build surrogates from waveforms with different NR resolutions. In our case, since the surrogate errors are larger than the NR resolution errors, we expect to obtain a surrogate of comparable quality using slightly lower resolution NR waveforms. If we use really low resolution NR waveforms, we would expect the surrogate errors to rise accordingly. In other cases where we do achieve surrogate errors similar to the NR resolution errors, comparing surrogates built from NR waveforms of different resolutions should yield similar differences to comparing the NR waveforms themselves.

- *Residual structure*. We could examine the parametric fit residuals and cross-validation residuals as a function of parameters. If the surrogate model captures the dominant features of NR waveforms then these residuals should appear random. From Fig. 6.14 we see that the largest errors occur at large values of $|\vec{\chi}_1|$ and for intermediate values of θ_χ, where precession has the largest effect. This indicates additional highly-precessing NR simulations may help significantly in reducing the surrogate errors.

6.10 Acknowledgements

We thank Michael Boyle, Alessandra Buonanno, Kipp Cannon, Maria Okounkova, Richard O'Shaughnessy, Christian Ott, Harald Pfeiffer, Michael Pürrer, and Saul Teukolsky for many useful discussions throughout this project. We also thank Andy Bohn, Nick Demos, Alyssa Garcia, Matt Giesler, Maria Okounkova, and Vijay Varma for helping to carry out the SpEC simulations used in this work. This work was supported in part by the Sherman Fairchild Foundation and NSF grant PHY-1404569 at Caltech. J.B. gratefully acknowledges support from NSERC of Canada. Computations were performed on NSF/NCSA Blue Waters under allocation PRAC ACI-1440083; on the NSF XSEDE network under allocation TG-PHY990007; on the Zwicky cluster at Caltech, which is supported by the Sherman Fairchild Foundation and by NSF award PHY-0960291; and on the ORCA cluster at California State University at Fullerton, which is supported by NSF grant PHY-1429873, the Research Corporation for Science Advancement, and California State University at Fullerton.

6.11 Forward-stepwise greedy fit algorithm

Here we describe in more detail the algorithm we use in section 6.6 used to fit the waveform data pieces evaluated at the empirical time nodes. Given N numerical reltivity simulations at parameters $\boldsymbol{\lambda}_{\mathrm{N}R} = \{\boldsymbol{\lambda}_i\}_{i=1}^N$, where $\boldsymbol{\lambda} = (q, |\chi_1|, \theta_\chi, \chi_2^z) = (\lambda^1, \lambda^2, \lambda^3, \lambda^4)$, we obtain each waveform data piece $X = \{X(t, \boldsymbol{\lambda}_i)\}_{i=1}^N$. Evaluating the surrogate model requires predicting $X_m(\boldsymbol{\lambda}) = X(T_m, \boldsymbol{\lambda})$ for each empirical time node T_m and for $\boldsymbol{\lambda} \notin \boldsymbol{\lambda}_{\mathrm{NR}}$. Denoting the model prediction as $X_{mS}(\boldsymbol{\lambda})$, we need not restrict to an interpolation scheme where $X_{mS}(\boldsymbol{\lambda}_i) = X_m(\boldsymbol{\lambda}_i)$ because the data contain numerical noise. Instead, we use linear fits such that

$$X_{mS}(\boldsymbol{\lambda}) = \sum_{i=1}^{M} c_i B^i(\boldsymbol{\lambda}) \tag{6.76}$$

for some set of basis functions $\{B^i\}_{i=1}^M$.

For simplicity, we choose all multivariate basis functions to be products of one-dimensional basis functions; that is, we choose $B^i \in \{B^{\vec{\alpha}}\}$, where

$$B^{\vec{\alpha}}(\boldsymbol{\lambda}) = \prod_{l=1}^{d} B_l^{\alpha^l}(\lambda^l). \tag{6.77}$$

Here $d = 4$ is the dimension of the parameter space, $\vec{\alpha} = (\alpha^1, \ldots, \alpha^d)$ labels which univariate basis functions enter the product, and we choose

- $B_1^k(q) = T_k(2q - 3)$
- $B_2^k(|\chi_1|) = \left(\frac{|\chi_1|}{0.8}\right)^k$
- $B_3^k(\theta_\chi) = \cos(k\theta_\chi)$
- $B_4^k(\chi_2^z) = T_k(\frac{\chi_2^z}{0.8})$,

where the T_k are Chebyshev polynomials of the first kind. We restrict the maximum order of the basis functions so that $\alpha^l \leq k_{\mathrm{max}}^l$ where $\vec{k}_{\mathrm{max}} = (5,6,6,4)$. We also restrict $\alpha^3 \leq \alpha^2$ to ensure θ_χ does not affect the surrogate output when $|\chi_1| = 0$.

The above choices are made for all waveform data pieces X except for $X = \varphi_p$. If the waveform data piece is φ_p we do the same as above except we instead choose

$$B_3^k(\theta_\chi) = \sin((k+1)\theta_\chi), \tag{6.78}$$

and we restrict $1 \leq \alpha^2 \leq 6$ and allow all $0 \leq \alpha^3 \leq 6$. We treat φ_p differently because the amount of precession is approximately proportional to the spin component orthogonal to the orbital angular momentum, while other waveform data pieces depend more strongly on the parallel component.

The above choices yield 1008 possible basis functions (1512 for φ_p), which is more than $N \leq 300$, so we will use only a subset of the possible basis functions. We determine elements $B^i \in \{B^{\vec{\alpha}}\}$ of this subset in a greedy manner with a forward-stepwise least-squares fit [290]. We proceed by iteratively updating two quantities:

r_j^n, which is the jth fit residual at the nth iteration, and $b_j^{\vec{\alpha},n}$, which is the orthogonal component of the basis function $B^{\vec{\alpha}}$ at the nth iteration evaluated at parameters $\boldsymbol{\lambda}_j$. For the zeroth iteration we begin with

$$r_j^0 = X_m(\boldsymbol{\lambda}_j) \tag{6.79}$$

$$b_j^{\vec{\alpha},0} = B^{\vec{\alpha}}(\boldsymbol{\lambda}_j). \tag{6.80}$$

At the nth iteration, we compute the inner product of the residuals with the basis functions

$$d_n^{\vec{\alpha}} = \sum_j r_j^n b_j^{\vec{\alpha},n}. \tag{6.81}$$

We then select the next most relevant basis function as the one with the largest magnitude inner product with the residuals

$$\vec{\alpha}_n^* = \arg\max_{\vec{\alpha}} |d_n^{\vec{\alpha}}| \tag{6.82}$$

and choose $B^n = B^{\vec{\alpha}_n^*}$. We compute the new residuals by subtracting the projection onto the newly chosen basis function

$$r_j^{n+1} = r_j^n - d_n^{\vec{\alpha}_n^*} b_j^{\vec{\alpha}_n^*,n} \tag{6.83}$$

and also orthogonalize the basis functions with respect to the new basis function

$$b_j^{\vec{\alpha},n+1} = b_j^{\vec{\alpha},n} - e^{\vec{\alpha},n} b_j^{\vec{\alpha}_n^*,n} \tag{6.84}$$

$$e^{\vec{\alpha},n} = \sum_j b_j^{\vec{\alpha},n} b_j^{\vec{\alpha}_n^*,n}. \tag{6.85}$$

We continue until we have performed $m \leq N$ iterations. We can then perform a least-squares fit using the m selected basis functions to find the coefficients c_i. In practice this is done during the greedy iteration by keeping track of the matrix of transformations relating $B^{\vec{\alpha}}(\vec{x}_j)$ and $b_j^{\vec{\alpha},n}$ as well as the coefficients $d_n^{\vec{\alpha}_n^*}$.

This procedure does not indicate which value of m (the number of fit coefficients) to use. Using N fit coefficients would be overfitting the data, and setting individual fit tolerances by hand for each empirical node of each data component would be time consuming and error prone. So instead, we repeat the above procedure for different values of m, we perform cross-validation studies on the resulting fits, we find the value of m that leads to the smallest validation errors (call this value m^*), and we choose $m = m^*$. For each trial $k = 1, \ldots, K = 50$ of this cross-validation procedure, we randomly divide the N data points into $N_v = 5$ validation points and

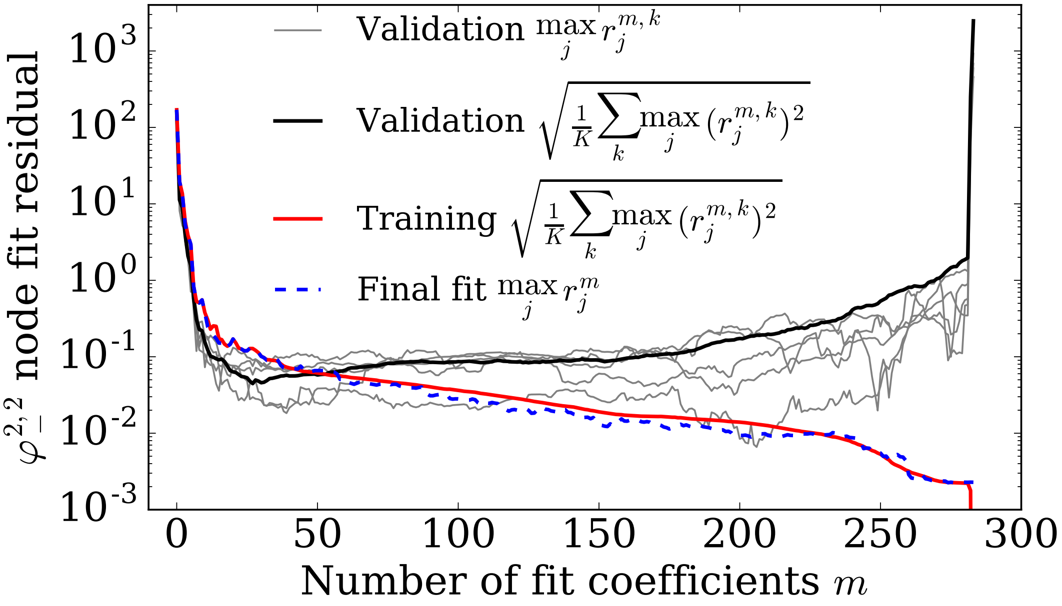

Figure 6.22: Fit residuals for the second empirical node of $\varphi_-^{2,2}$ at $t = -806.5M$. Blue dashed: The maximum fit residual using all data. Thin grey lines: Maximum validation residual for individual trials. Thick black line: The RMS of the validation residuals for $K = 50$ trials. It takes its minimum value at $m = 30$, which determines the number of fit coefficients to use for this node in the model. Red: The RMS of the training residuals for $K = 50$ trials.

$N_t = N - N_v$ training points. Using only the training data, we perform the above greedy forward-stepwise fitting procedure. For values of $m \in [0, N_t]$, we obtain a least-squares fit with m coefficients using the training data and evaluate the fit residuals $r_j^{m,k}$ for the validation data. We choose

$$m^* = \arg\min_m \sum_{k=1}^{K} \max_{j=1}^{N_v} \left(r_j^{m,k}\right)^2 . \tag{6.86}$$

We use the maximum over j because we seek to minimize the largest fit residuals, and we sum in quadrature over k rather than maximize to account for cases where data points with large errors or corner cases are selected as validation points, which can lead to large fit residuals. The dependence of the residuals on m for one case is shown in Fig. 6.22.

6.12 Comparing reduced basis constructions

We compare two commonly used methods to generate a reduced basis in gravitational waveform reduced-order modeling. The first uses a singular value decomposition

(SVD) of a data set whose output consists of a set of basis vectors ranked by their "singular values", which are eigenvalues when the input data is square. The SVD reduced basis follows by truncating the output basis beyond a selected singular value. The resulting basis is accurate up to that singular value as measured in a root-mean-square norm. The second method uses a greedy algorithm, which is iterative and nested, to expose the most relevant elements of the input (or *training*) data set [162, 304]. The greedy algorithm selects the element with the largest current projection error (as measured by a specified norm), orthonormalizes the selected element with respect to the current basis, and adds this orthonormalized element to the set of basis vectors. In practice, one uses an iterated, modified Gram-Schmidt process [305] for orthonormalization, which is robust to the accumulation of numerical round-off effects from subtraction until very large basis sizes. The algorithm ends when the largest projection error is below a specified tolerance; it also ends if a previously-selected training data element is selected again, which, if it were allowed to occur, would introduce a linearly dependent element to the basis. The output includes a (greedy) reduced basis and a set of parameters or labels that indicate the most relevant elements of the training data from which the basis is built.

Both SVD and greedy methods output a reduced basis that accurately represents the training data to the requested singular value or tolerance. The output of the SVD algorithm depends only on the training data. The greedy algorithm, on the other hand, begins by choosing one of the training data elements as the first basis vector, so its output depends also on that choice. How that choice is made is often arbitrary and may depend on the application. For example, one may seed the greedy algorithm with an arbitrary element from the training set or choose the element that has the largest absolute value or norm. However, it has been shown that the choice of seed is largely irrelevant as the greedy algorithm seeks to minimize the maximum projection error across the entire training set, no matter what the seed. The resulting variations in the size of the greedy reduced basis due to arbitrary seed choices are marginal and typically span a few percent about the mean size [75, 282, 306].

Practical implementations of the SVD algorithm can be found rather easily because of its broad use across many disciplines. Therefore, building an SVD reduced basis for a training set of waveforms is as straightforward as calling the appropriate programmed function. However, if the training data contains N waveforms with L time or frequency samples then the SVD algorithm is $O(N^2L)$, which can be intensive in both time and physical memory. For this reason, the authors in [307] divide the full training space into narrow strips in one direction of the parameter space. Dividing the training space into smaller subsets results in a direct product of reduced bases, one basis for each subset. Unfortunately, the total number of the basis elements tends to be larger than if one had performed a SVD on the full training data (if it can be done). Consequently, the reduction of the data is not maximized.

One often has considerable flexibility in designing a greedy algorithm for a specific application. If the training set remains fixed throughout the course of the greedy algorithm (see [248] for an example where this is not the case) then each iteration step can be performed in constant time so that the totality scales as $O(nN)$ if n is the number of reduced basis elements needed to reach the specified tolerance. Typically, $n \ll N$ so that greedy algorithms tend to terminate more quickly than an application of SVD on the same training data, though there is some additional influence from implementation details. The greedy algorithm can be parallelized to break up the computation of expensive integrals across different processes [308]. In addition, the size and memory requirements of a very large training set pose little problem for greedy algorithms. The training space can be divided into subsets so that a reduced basis is built for each with a tolerance up to numerical round-off as measured in the L_∞ norm (to have point-wise accuracy for the data). Then, one may apply a second greedy algorithm on the full training data by using instead the basis data on each subset to represent the original data of each subset. In this way, one can generate a reduced basis that spans all the subsets and maximizes the reduction of the full training set [309]. Combining this two-step greedy algorithm with the parallelization of the projection integrals discussed above provides a viable and practical strategy for building a reduced basis for training sets of virtually any size. Another strategy is to randomly repopulate the training set at each iteration of the greedy algorithm [248, 310]. This approach requires that the training data can be generated at will for any parameter values but also avoids storing prohibitively large amounts of data at any step in the greedy algorithm.

Finally, greedy algorithms allow one to use any measure for determining the projection errors. This includes choosing among L_2, L_∞, and L_n error norms or any combination thereof. In addition, computing the integrals for projecting the training data onto the basis can be achieved with any quadrature rule one wishes. However, implementations of the SVD algorithm are restricted to the L_2 measure and the reduced basis will depend on how the training data is sampled in time or frequency.

Let us next investigate a toy problem to facilitate a comparison of the outputs of a

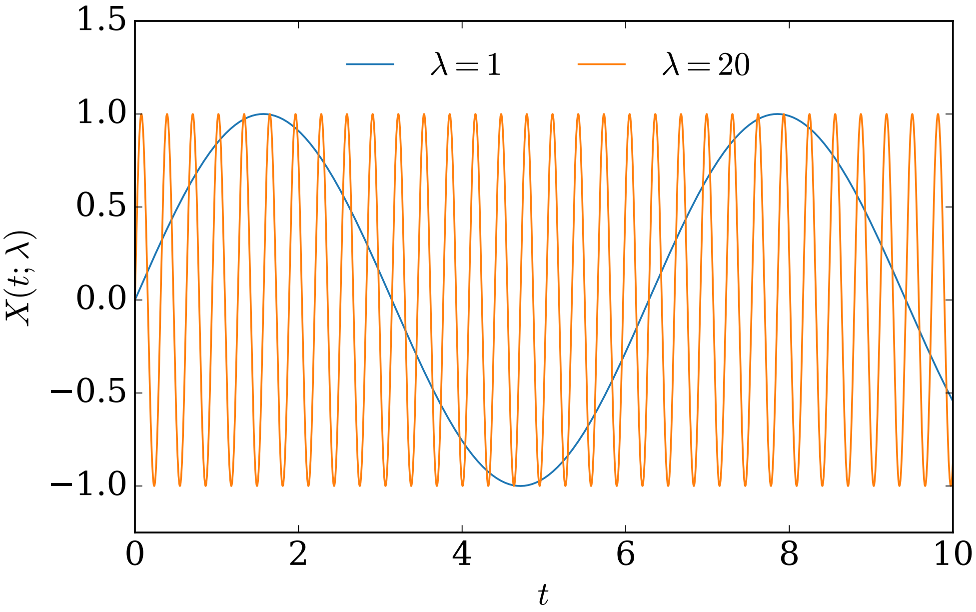

Figure 6.23: Plots of $X(t;\lambda)$ for our toy problem evaluated at the smallest and largest values of λ in the training set.

basic greedy algorithm and SVD. We consider a function

$$X(t;\lambda) = \sin(\lambda t) + 10^{-5}\sin(10\lambda t) + 10^{-10}\xi(t)\,, \tag{6.87}$$

where $t \in [0,10]$ with a parameter $\lambda \in [1,20]$. There is a relatively high frequency component with an amplitude of 10^{-5}. The quantity $\xi(t)$ is a random variable drawn from a normal distribution with zero mean and variance of one. This stochastic term has an amplitude of only 10^{-10}.

Our training set will consist of $N = 1000$ uniformly spaced values of λ. Figure 6.23 shows training data for the smallest and largest parameter values considered here. We sample the function in (6.87) at 10,000 uniformly spaced times.

We construct three reduced bases. The first is built from an SVD on the training data. The second uses a greedy algorithm to generate a reduced basis and a corresponding set of parameters; here we use the L_2 norm to measure the difference between each training set element and its projection onto the basis. The third is built in the same greedy manner as the second but uses the L_∞ norm to measure the projection error. Recall that the L_2 error constitutes a kind of average as it involves an integration in time whereas the L_∞ error measures the largest, point-wise, absolute difference and is thus more stringent. Figure 6.24 shows the maximum projection errors, as

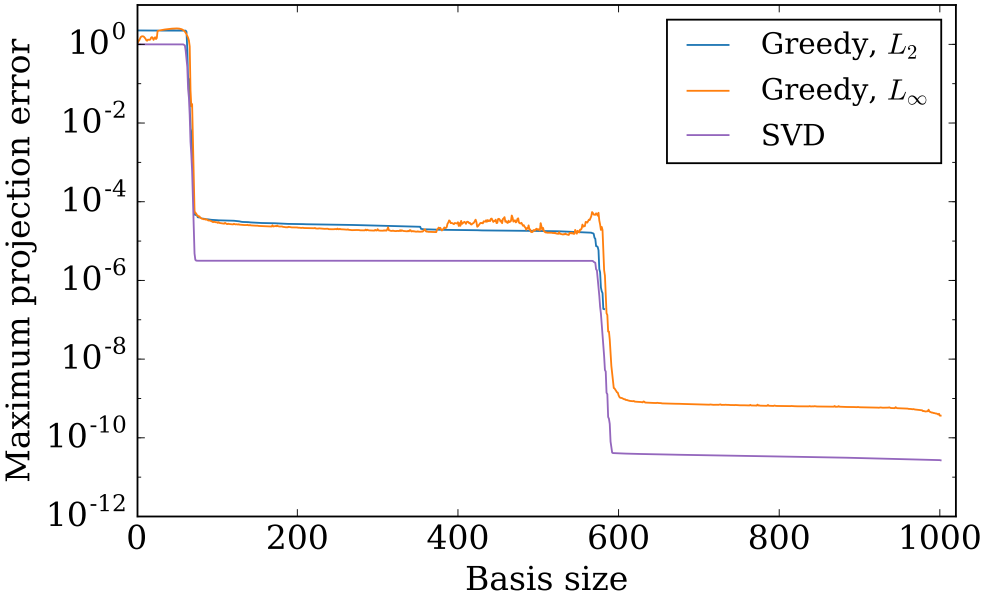

Figure 6.24: Maximum projection errors of all three reduced bases (see text for a description) versus the size of the basis.

measured with their respective norms, associated with these three methods as a function of the size of the basis. The absolute tolerance on the greedy algorithm bases is 10^{-14} while the smallest singular value kept is 10^{-14} relative to the largest. We observe three plateaus for each of the cases, which can be attributed to each algorithm trying to resolve the features at the $O(1)$, $O(10^{-5})$, and $O(10^{-10})$ scales in the data; see (6.87). In fact, none of the algorithms are able to completely resolve the very low-amplitude stochastic features until the training set has been exhausted and all data has been used to build the reduced bases. Notice that the error curve is somewhat noisy for the L_∞ case while the other two are smooth. Also, the maximum projection error for the L_2 case ends at about 10^{-7} due to a parameter being selected a second time.

Figure 6.25 shows the projection errors (as measured in the L_2 norm) onto each of the three reduced bases for test data generated by randomly selecting 1000 values of λ in the training interval $[1,20]$. The errors for “Greedy, L_∞” and “SVD” lie nearly on top of each other while those for “Greedy, L_2” are relatively large because the effective greedy algorithm tolerance for this basis is only 10^{-7} as discussed above. In all cases, the small-amplitude stochastic noise in the data prevents the projection errors of the test data from being less than a few times 10^{-10}; see (6.87).

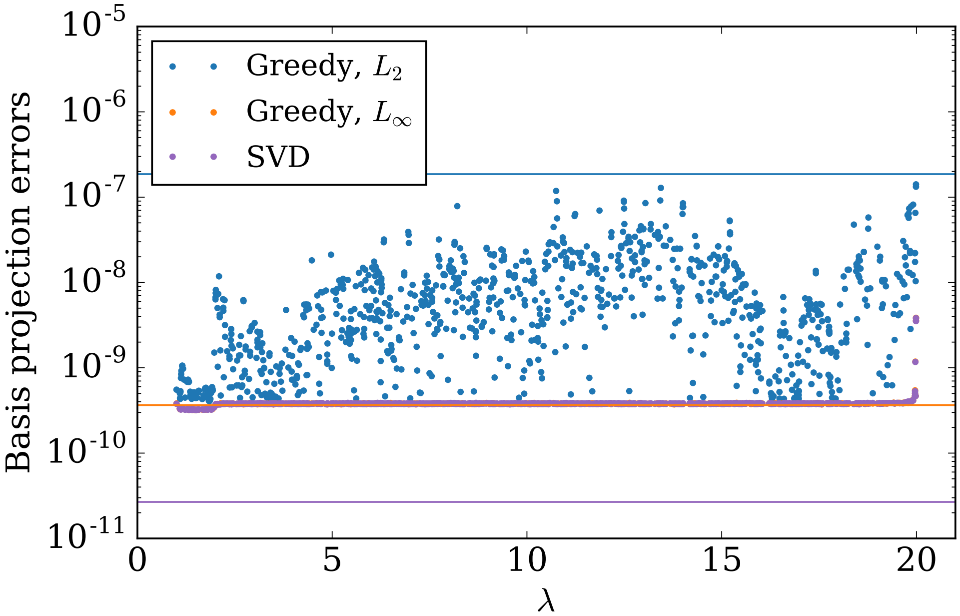

Figure 6.25: Projection errors, measured in the L_2 norm for the three reduced bases described in the text, computed for test data generated from 1000 randomly selected parameters λ in $[1, 20]$. The corresponding colored lines indicate the smallest projection errors on the training sets shown in Fig. 6.24. The errors for "Greedy, L_∞" and "SVD" lie nearly on top of each other. However, the maximum projection error implied by SVD (purple line) underestimates the true errors (dots) by an order of magnitude.

Finally, the SVD method is able to produce a reduced basis with elements that smooth many uncorrelated features manifest in the training data. Such smoothing is useful for surrogate model building because the resulting basis elements tend to exhibit smoother variation in time or frequency; this translates into smoother variations across parameters, thereby yielding more accurate fits for the parametric variation at the empirical interpolation nodes. The reduced bases produced by greedy methods tend to not to share this smoothing ability of the SVD method.

To demonstrate SVD's smoothing abilities, we replace the function in (6.87) with a smooth oscillating term plus a stochastic term with amplitude of 10% of the first so that the noise is visible to the naked eye,

$$X(t;\lambda) = \sin(\lambda t) + 0.1\,\xi(t). \tag{6.88}$$

We build three reduced bases on the corresponding training sets (with the same t and λ intervals and samples) using the same methods as before. Figure 6.26 shows the tenth basis element as a function of t for each of the three reduced basis building strategies. The two bases built from a greedy method exhibit the noise found in the training data. However, the SVD basis element in the bottom panel reveals a smooth function with very low amplitude noise, much lower than appears in the training data amplitudes.

In the case of the NRSur4d2s surrogate discussed here, note that data from each of the NR simulations contains spurious oscillations on the orbital timescale; these oscillations are caused by residual orbital eccentricity and by nutation effects that we have not filtered out (§ 6.5), and because these oscillations are uncorrelated from one simulation to another, they appear as stochastic noise. To smooth this noise, we therefore use the SVD method to obtain basis vectors for emperical interpolation when building NRSur4d2s (§ 6.6). This smoothing significantly improves the accuracy of our fits of the waveform quantities at the empirical interpolation nodes. However, note also that for NRSur4d2s we use the greedy method to expose the BBH parameters for performing expensive NR simulations (§ 6.4). Therefore, we use the benefits of both the greedy and SVD methods in building NRSur4d2s.

6.13 Motivating the use of $\mathcal{E}$

A commonly used measure of the difference between waveforms $h_1(t, \theta_1, \phi_1; \boldsymbol{\lambda}_1)$ and $h_2(t, \theta_2, \phi_2; \boldsymbol{\lambda}_2)$ is the overlap error

$$1 - \mathcal{O} = 1 - \frac{\langle h_1, h_2 \rangle}{\sqrt{\langle h_1, h_1 \rangle \langle h_2, h_2 \rangle}}, \tag{6.89}$$

where $\langle \cdot, \cdot \rangle$ is often chosen to be the frequency domain noise-weighted inner product [287]

$$\langle a, b \rangle_f = 4\mathrm{Re} \int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df. \tag{6.90}$$

Here $S_n(f)$ is the power spectral density of noise in a gravitational wave detector and tildes are used to represent a Fourier transform.

If we use a flat (frequency-independent) power spectral density, we may instead perform the integration in the time domain and use

$$\langle a, b \rangle_t = \mathrm{Re} \int_{t_{\mathrm{min}}}^{t_{\mathrm{max}}} a(t) b^*(t) dt \tag{6.91}$$

to obtain the same overlap error. While a completely flat power spectral density is unphysical, the design sensitivity of aLIGO [311] varies only by a factor of ~ 2 between 50Hz and 1000Hz. Putting rigorous limits on weighted frequency domain

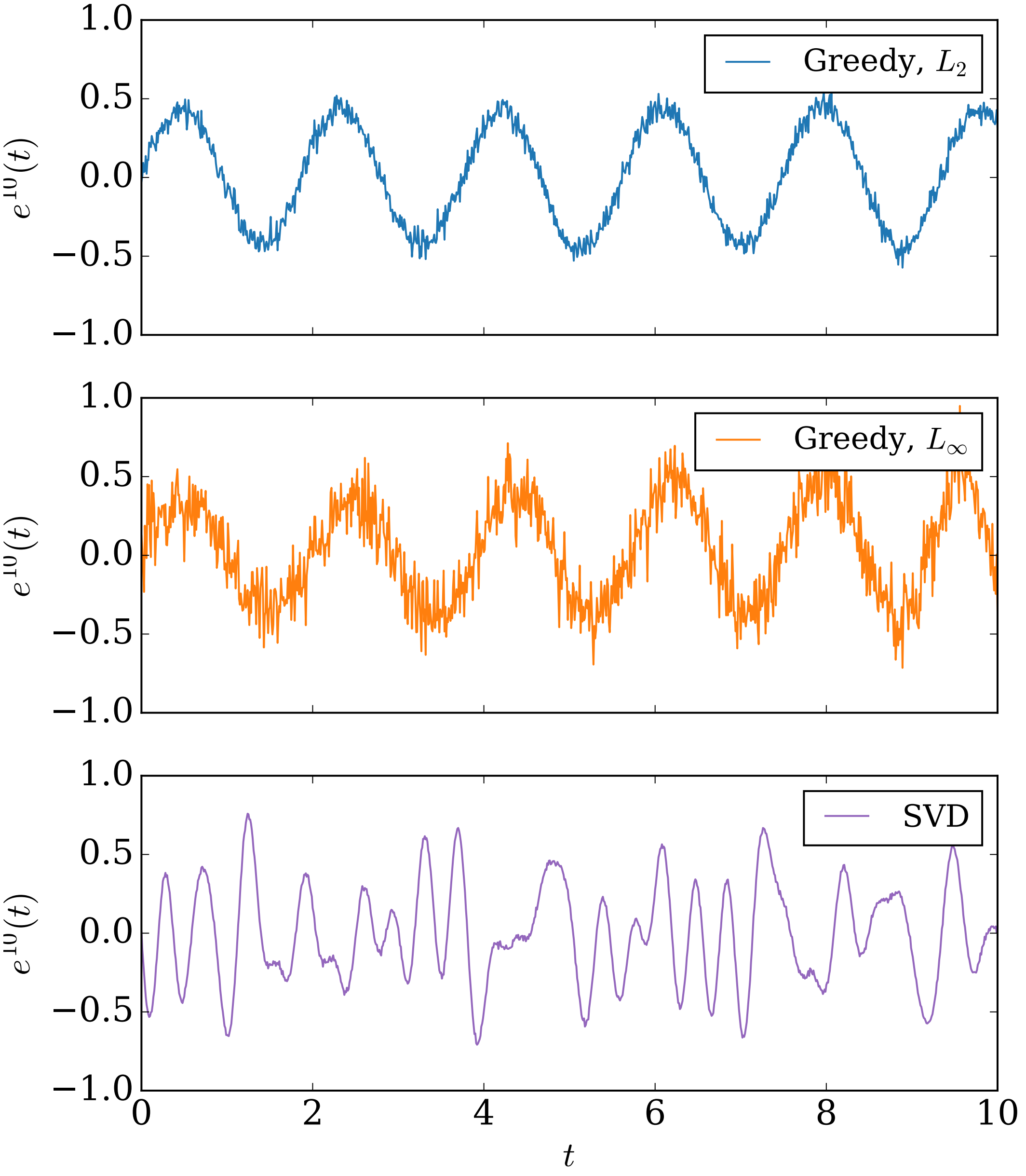

Figure 6.26: The tenth basis element as a function of t from the three reduced bases elements described in the text. The training data used is given by the parameterized function in (6.88) and exhibits relatively large amplitude fluctuations. Whereas the top two plots show significant noise in the basis element, the SVD method smooths away almost completely the uncorrelated stochastic features to generate a basis element that is smooth in t.

errors based on unweighted time domain errors is not straightforward [147, 312], but the time domain errors are computationally cheap to compute, useful for quantifying time domain waveform models, and (like NR waveforms and our surrogate model NRSur4d2s) independent of the total binary mass M.

We can relate the time domain overlap error to δh by performing a weighted average over the sphere and using

$$\int_{S^2} a(\theta,\phi) b^*(\theta,\phi) d\Omega = \sum_{\ell,m} a^{\ell,m} b^{\ell,m*} \tag{6.92}$$

due to the orthonormality of the SWSHs. Using $||a||_t^2 = \langle a, a \rangle_t$, we have

$$\begin{aligned} \delta h^2 &= \frac{1}{T} \sum_{\ell,m} ||\delta h^{\ell,m}||_t^2 && (6.93) \\ &= \frac{1}{T} \int_{S^2} ||h_1(t,\theta,\phi;\boldsymbol{\lambda}_1) - h_2(t,\theta,\phi;\boldsymbol{\lambda}_2)||_t^2 d\Omega && (6.94) \\ &= \frac{1}{T} \int_{S^2} \left(||h_1||_t^2 + ||h_2||_t^2 - 2\langle h_1 h_2 \rangle_t \right) d\Omega\,, && (6.95) \end{aligned}$$

where in the last line we have omitted arguments to h_1 and h_2. If $||h_1(t,\theta,\phi;\boldsymbol{\lambda}_1)||_t = ||h_2(t,\theta,\phi;\boldsymbol{\lambda}_2)||_t$ for all θ, ϕ then we would have

$$\frac{\delta h^2}{\sum_{\ell,m} ||h_1^{\ell,m}||_t^2} = \frac{2 \int_{S^2} w(\theta,\phi)(1 - \mathcal{O}(\theta,\phi)) d\Omega}{\int_{S^2} w(\theta,\phi) d\Omega}\,, \tag{6.96}$$

where $w(\theta,\phi) = ||h_i(t,\theta,\phi;\boldsymbol{\lambda}_i)||_t^2$. Denoting $||h||^2 \equiv \sum_{\ell,m} ||h^{\ell,m}||_t^2$, this motivates the use of the relative error measure

$$\mathcal{E} \equiv \frac{1}{2} \frac{\delta h^2}{||h_1||^2} \tag{6.97}$$

as it is similar to a sphere-weighted average of overlap errors, where the weighting emphasizes directions with a larger amount of gravitational wave emission. We note, however, that while the overlap error vanishes if h_1 and h_2 are identical except for normalization, $\mathcal{E}$ does not and vanishes only when h_1 and h_2 are identical. This is important as a different normalization will lead to a bias when measuring the distance to the source of a gravitational wave.

6.14 Mismatches optimized over time and polarization shifts

Given gravitational waveform polarization signals $h_+(t)$ and $h_\times(t)$, each gravitational wave detector in a detector network will observe a linear combination of $h_+(t)$ and

$h_\times(t)$ depending on their orientation with respect to the direction of propagation and polarization axes. For the purposes of building gravitational wave models, we are interested in the best case scenario when both polarizations are measured. Including “blind spots" in the detector network could lead to artificially large relative errors, so we assume a network of two detectors where one measures $h_+(t)$ and the other measures $h_\times(t)$. Given model predictions $h_+^m(t)$ and $h_\times^m(t)$ for the two polarizations, we compute the two-detector overlap

$$\mathcal{O} = \frac{\langle h_+, h_+^m\rangle + \langle h_\times, h_\times^m\rangle}{\sqrt{\left(\langle h_+, h_+\rangle + \langle h_\times, h_\times\rangle\right)\left(\langle h_+^m, h_+^m\rangle + \langle h_\times^m, h_\times^m\rangle\right)}}$$

with a real inner product given by

$$\langle a, b\rangle = \mathrm{Re}\left[\langle a, b\rangle_C\right] \tag{6.98}$$

$$\langle a, b\rangle_C = \int \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(|f|)} df. \tag{6.99}$$

As in Eq. 6.23, a tilde denotes a frequency domain signal, which is computed by using an FFT after tapering the ends of the time domain signal. In this case, the complex inner product $\langle \cdot, \cdot \rangle_C$ is integrated over the negative and positive frequency intervals $[-f_{\mathrm{max}}, -f_{\mathrm{min}}]$ and $[f_{\mathrm{min}}, f_{\mathrm{max}}]$ for some positive f_{min} and f_{max}. Note that for any two real functions $a(t)$ and $b(t)$, we have

$$\tilde{a}(-f)\tilde{b}^*(-f) = \left(\tilde{a}(f)\tilde{b}^*(f)\right)^* \tag{6.100}$$

and so $\langle a, b\rangle_C$ is real.

Defining complex gravitational wave signals

$$h(t) = h_+(t) - i h_\times(t) \tag{6.101}$$

$$h^m(t) = h_+^m(t) - i h_\times^m(t), \tag{6.102}$$

we can compute a complex overlap

$$\begin{aligned}\mathcal{O}_C &= \frac{\langle h, h^m\rangle_C}{\sqrt{\langle h, h\rangle_C \langle h^m, h^m\rangle_C}} \\ &= \frac{\langle h_+, h_+^m\rangle + \langle h_\times, h_\times^m\rangle + i\left(\langle h_+, h_\times^m\rangle - \langle h_\times, h_+^m\rangle\right)}{\sqrt{\left(\langle h_+, h_+\rangle + \langle h_\times, h_\times\rangle\right)\left(\langle h_+^m, h_+^m\rangle + \langle h_\times^m, h_\times^m\rangle\right)}}.\end{aligned}$$

Since the time domain polarization signals are all real, we have

$$\mathcal{O} = \mathrm{Re}[\mathcal{O}_C]. \tag{6.103}$$

A polarization angle shift of ψ and time shift of δt in the model waveform results in the transformations

$$h^m(t) \to h^m_T(t) = h^m(t+\delta t)e^{2i\psi}, \tag{6.104}$$

$$\tilde{h}^m(f) \to \tilde{h}^m_T(f) = \tilde{h}^m(f)e^{2i\psi}e^{2\pi i\delta t}, \tag{6.105}$$

where h^m_T is the transformed model waveform. The overlap of the signal waveform with the transformed model waveform is then

$$\begin{aligned}\mathcal{O}(\psi,\delta t) &= \mathrm{Re}\left[\frac{\langle h, h^m_T\rangle_C}{\sqrt{\langle h,h\rangle_C\langle h^m_T, h^m_T\rangle_C}}\right]\\ &= \mathrm{Re}\left[\frac{e^{-2i\psi}}{\sqrt{\langle h,h\rangle_C\langle h^m, h^m\rangle_C}}\int \frac{\tilde{h}(f)\tilde{h}^{m*}(f)}{S_n(|f|)}e^{-2i\pi\delta t}\right].\end{aligned}$$

The above integral can be evaluated efficiently for many values of δt using an FFT. We can then compute the mismatch

$$\text{mismatch} = 1 - \max_{\psi,\delta t}\mathcal{O}(\psi,\delta t) \tag{6.106}$$

by taking the absolute value of the complex overlap for each δt to maximize over ψ, and taking the maximum over all available values of δt. In practice, the true maximum over δt will lie between available samples, so we fit the overlap peak to a quadratic function in δt using the largest overlap sample and the neighboring value on either side. We also pad with zeros before taking the FFT to obtain a finer sampling in δt.

6.15 Post-Newtonian surrogate waveform decomposition

The second greedy algorithm described in Sec. 6.4 makes use of surrogate models of Post-Newtonian (PN) waveforms. At each greedy step, a new PN surrogate model is built from PN waveforms evaluated at the currently known greedy parameters G. This surrogate is evaluated for each training point $\boldsymbol{\lambda} \in \mathcal{T}^i_{\mathrm{TS}}$ and the surrogate waveform is compared to the actual PN waveform. Here, we describe the differences between how the PN surrogates were built compared to the NR surrogate NRSur4d2s described in the main body.

PN waveforms do not contain a merger phase, so we cannot use the peak amplitude to align the waveforms in time. We instead choose $t = 0$ to correspond to an orbital

Data	Tol	Data	Tol	Data	Tol	Data	Tol
φ_p	0.01	φ_d	0.1	$\varphi_-^{2,2}$	0.01	$\varphi_-^{2,1}$	0.1
$\varphi[H[X]]$	0.1	$\lvert H[\varphi_+^{2,2}]\rvert$	0.0001	$\lvert H[\varphi_+^{2,1}]\rvert$	0.0001		

Table 6.8: Fit tolerance for the empirical node parametric fits of PN surrogates. Fit coefficients were added until the maximum fit residual fell below the tolerance. A tolerance of 0.001 was used for unlisted waveform data pieces.

angular frequency of 0.09. This frequency is computed from the waveform [289]. We choose $t_{\rm min} = -5000M$, $t_0 = -4500M$, $t_f = -100M$, and $t_{\rm max} = 0$. The PN waveforms used to build the PN surrogate then have domain $t \in [-5000M, 0]$, and the PN surrogate waveforms have domain $t \in [-4500M, -100M]$. The parameters of the PN waveforms are given at $t = t_0$. The rotation alignment at $t = t_0$ is the same as for the NR waveforms, described in Sec. 6.4.

The waveform decomposition used for the PN surrogates was slightly different from the one described in Sec. 6.5. We limited the PN waveforms to contain only the $\ell = 2$ modes (with all 5 values of m). Additionally, since we were able to obtain the desired values of ϕ_χ at $t = t_0$ with PN waveforms, there was no need to make any transformations related to ϕ_χ.

The number of coefficients used in the parametric fits of the empirical nodes was determined differently for PN surrogates than for NRSur4d2s. Instead of the cross-validation method described in Appendix 6.11, coefficients were added until the fit residuals fell below a specified tolerance, given in Table 6.8. To prevent overfitting, the number of fit coefficients was also limited to be at most 75% of the number of data points used in the fit. The basis functions in $|\chi_1|$ used for the fits were also different, with $B_2^k(|\chi_1|) = T_k(2.5|\chi_1| - 1)$. For the PN surrogates, we did not make the restriction $\alpha^3 \leq \alpha^2$ so θ_χ affected the PN surrogate output when $|\vec{\chi}_1| = 0$.

7 A Numerical Relativity Waveform Surrogate Model for Generically Precessing Binary Black Hole Mergers

7.1 Background and Contributions

This chapter [15] has been submitted to Physical Review D for publication. A preprint can be found at `https://arxiv.org/abs/1705.07089`.

This chapter presents the first surrogate model of numerical relativity (NR) waveforms that includes all 7 intrinsic parameters. The surrogate model covers mass ratios up to $q = 2$ and dimensionless spin magnitudes up to 0.8, and is nearly as accurate as performing new NR simulations within this parameter range.

I helped select the parameters for the new NR simulations, and performed some of them. I determined the methods to use, implemented them, and built the surrogate models. New methods include an ODE evolution of the spin, orbital, and precesion dynamics, extending spins through merger and ringdown with post-Newtonian equations, and using the spins at empirical time nodes rather than the initial spins for improved parametric fit accuracy. I also performed the error studies. I generated the figures and the first draft of the manuscript, and assisted in editing the manuscript.

A Numerical Relativity Waveform Surrogate Model for Generically Precessing Binary Black Hole Mergers

Jonathan Blackman,[1] Scott E. Field,[2, 3] Mark A. Scheel,[1] Chad R. Galley,[1]
and Christian D. Ott[1, 4]

[1] *TAPIR, Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, CA 91125, USA*
[2] *Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853, USA*
[3] *Mathematics Department, University of Massachusetts Dartmouth, Dartmouth, MA 02747, USA*
[4] *Center for Gravitational Physics and International Research Unit of Advanced Future Studies, Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto, Japan*

A generic, non-eccentric binary black hole (BBH) system emits gravitational waves (GWs) that are completely described by 7 intrinsic parameters: the black hole spin vectors and the ratio of their masses. Simulating a BBH coalescence by solving Einstein's equations numerically is computationally expensive, requiring days to months of computing resources for a single set of parameter values. Since theoretical predictions of the GWs are often needed for many different source parameters, a fast and accurate model is essential. We present the first surrogate model for GWs from the coalescence of BBHs including all 7 dimensions of the intrinsic non-eccentric parameter space. The surrogate model, which we call NRSur7dq2, is built from the results of 744 numerical relativity simulations. NRSur7dq2 covers spin magnitudes up to 0.8 and mass ratios up to 2, includes all $\ell \leq 4$ modes, begins about 20 orbits before merger, and can be evaluated in $\sim$ 50 ms. We find the largest NRSur7dq2 errors to be comparable to the largest errors in the numerical relativity simulations, and more than an order of magnitude smaller than the errors of other waveform models. Our model, and more broadly the methods developed here, will enable studies that require millions of numerical relativity waveforms, such as parameter inference and tests of general relativity with GW observations.

7.2 Introduction

With LIGO's two confident detections of gravitational waves (GWs) from binary black hole (BBH) systems [250, 251], we have entered the exciting new era of GW astronomy. The source black hole (BH) masses and spins can be determined by comparing the signal to waveforms predicted by general relativity (GR) [20], and new strong-field tests of GR can be performed [252]. These measurements and tests require GW models that are both accurate and fast to evaluate. The total mass of the system M can be scaled out of the problem, leaving a 7-dimensional intrinsic parameter space over which the waveform must be modeled, consisting of the mass ratio and two BH spin vectors.

Numerical relativity (NR) simulations of BBH mergers [191, 209, 253–257] solve the full Einstein equations and produce the most accurate waveforms. These simulations are computationally expensive, requiring weeks to months on dozens of CPU cores for a waveform beginning ~ 20 orbits before the merger. Analytic and semi-analytic waveform models [65, 66, 70, 152, 227, 261, 262, 264, 265] are quick to evaluate, but they make approximations that can introduce differences with respect to the true waveform predicted by GR. These differences could lead to parameter biases or inaccurate tests of GR for some high signal-to-noise ratio detections that could be made in the near future [147].

A *surrogate waveform model* [14, 76, 239, 265, 269] is a model that takes a set of precomputed waveforms that were generated by some other model (e.g., NR or a semianalytic model), and interpolates in parameter space between these waveforms to quickly produce a waveform for any desired parameters. A surrogate waveform can be evaluated much more quickly than the underlying model, and can be made as accurate as the underlying model given a sufficiently large set of precomputed waveforms that cover the parameter space. Previous surrogate models based on NR waveforms were built for non-spinning BBH systems [269] and for a 4-dimensional ($4d$) parameter subspace containing precession [14]. Here, we present the first NR surrogate model including all 7 dimensions of the parameter space. The model, which we call NRSur7dq2, produces waveforms nearly as accurate as those from NR simulations, but can be evaluated in $\sim$ 50 ms on a single CPU core for a speedup of more than 8 orders of magnitude compared to NR. Our method enables performing high accuracy GW data analysis, including parameter inference for astrophysics and tests of GR.

7.3 Numerical Relativity Data

The NR simulations used to build the surrogate model are performed using the Spectral Einstein Code (SpEC) [209, 211, 270–274]. The simulations begin at a coordinate time $\tau = 0$, where we specify the BH mass ratio $q = m_1/m_2 \geq 1$ and initial dimensionless spin vectors

$$\vec{\chi}_i(\tau = 0) = \vec{S}_i(\tau = 0)/m_i^2\,,\ i \in \{1,2\}\,. \tag{7.1}$$

The system is evolved through merger and ringdown, and the GW is extracted at multiple finite radii from the source. These are extrapolated to future null infinity [216] using quadratic polynomials in $1/r$, where r is a radial coordinate. The effects of any drifts in the center of mass that are linear in time are removed from the waveform [140, 286, 289, 313]. The waveforms at future null infinity use a time coordinate $\tilde{\tau}$, which is different from the simulation time τ, and begins approximately at $\tilde{\tau} = 0$. The spins $\vec{\chi}_i$ are also measured at each simulation time τ. To compare spin and waveform features, we identify τ with $\tilde{\tau}$. While this identification is not gauge-independent, the spin directions are already gauge-dependent. We note that the spin and orbital angular momentum vectors in the damped harmonic gauge used by SpEC agree quite well with the corresponding vectors in post-Newtonian (PN) theory [57].

Once we have the spins $\vec{\chi}_i(\tau)$ and spin-weighted spherical harmonic modes of the waveform $h^{\ell,m}(\tau)$, we perform the same alignment discussed in Sec. III.D of Ref. [14]. Briefly, we first determine the time τ_{peak} which maximizes the total amplitude of the waveform

$$A_{\mathrm{tot}}(\tau) = \sqrt{\sum_{\ell,m} |h^{\ell,m}(\tau)|^2}\,. \tag{7.2}$$

We choose a new time coordinate

$$t = \tau - \tau_{\mathrm{peak}}\,, \tag{7.3}$$

which maximizes A_{tot} at $t = 0$. We then rotate the waveform modes such that at our reference time of $t = t_0 = -4500M$, $\hat{z}$ is the principal eigenvector of the angular momentum operator [167] and the phases of $h^{2,2}(t_0)$ and $h^{2,-2}(t_0)$ are equal. We sample the waveform and spins in steps of $\delta t = 0.1M$, from t_0 to $t_f = 100M$.

We first include all 276 NR simulations used in the NRSur4d2s surrogate model and the 9 additional simulations used in Sec. IV.D and Table V of Ref. [14]. We perform 459 additional NR simulations. The first 361 of these are chosen based on sparse grids [314] and include combinations of extremal parameter values (such as $q \in \{1,2\}$) and intermediate values as detailed in Appendix 7.10. The parameters for the remaining 98 simulations are chosen as follows. We randomly sample 1000 points in parameter space uniformly in mass ratio, spin magnitude, and spin direction on the sphere. We compute the distance between points a and b using

$$ds^2 = (0.3\,(q_a - q_b))^2 + \sum_{i\in\{1,2\}} \|\vec{\chi}_{ia} - \vec{\chi}_{ib}\|^2 \,. \tag{7.4}$$

For each sampled parameter we compute the minimum distance to all previously chosen parameters. We then choose the sampled parameter maximizing this minimum distance. We then resample the 1000 parameters for the next of the 98 iterations. This results in a total of 744 NR simulations. For simulations with equal masses and unequal spins, we use the results twice by reversing the labeling of the BHs and rotating the waveform accordingly. There are 142 such simulations, leading to 886 NR waveforms.

7.4 Waveform Decomposition

The goal of a surrogate model is to take a precomputed set of waveform modes $\{h_i^{\ell,m}(t)\}$ at a fixed set of points in parameter space $\{\vec{\lambda}_i\}$, and to produce waveform modes $\{h^{\ell,m}(t)\}$ at new desired parameter values. Because $h^{\ell,m}(t)$ is highly oscillatory and changes in a complicated way as one varies the masses and spins, it is not feasible to directly interpolate $\{h_i^{\ell,m}(t)\}$ in parameter space with only $886^{1/7} \approx 2.64$ available points per dimension. Instead, we decompose each waveform $h(t)$ into many *waveform data pieces*. Each waveform data piece is a simpler function that varies slowly over parameters. Once we have interpolated each waveform data piece to a desired point in parameter space, we recombine them to form $h(t)$. Our decomposition is similar to but improves upon the one used in Ref. [14].

We first determine the unit quaternions $\hat{q}(t)$ that define the coprecessing frame, and we determine the waveform modes $\{h_C^{\ell,m}(t)\}$ in this frame. This is done using the transformation T_C given by Eq. 27 of Ref. [14]. The spins $\vec{\chi}_i(t)$ are also transformed to the coprecessing frame using

$$\vec{\chi}_i^{\mathrm{copr}}(t) = \hat{q}^{-1}(t)\vec{\chi}_i(t)\hat{q}(t)\,. \tag{7.5}$$

Note that quaternion multiplication is used here, and vectors are treated as quaternions with zero scalar component. We note that unlike in Ref. [14], here we do not filter $\hat{q}(t)$ and, as described below, the spins play an important role in the surrogate's construction.

The orbital phase

$$\varphi(t) = \frac{1}{4}\left(\arg\left[h_C^{2,-2}(t)\right] - \arg\left[h_C^{2,2}(t)\right]\right) \tag{7.6}$$

is computed from the coprecessing waveform modes. This is expected to be superior to computing the orbital phase from the BH trajectories because unlike the coordinate-dependent trajectories, the waveform can be made gauge invariant up to Bondi-Metzner-Sachs transformations [140]. We then transform the spins and waveform modes to a coorbital frame, in which the BHs are nearly on the x axis. The coorbital frame is just the coprecessing frame rotated by $\varphi(t)$ about the z axis. Specifically, we have

$$\hat{q}_{\mathrm{r}}(t) = \cos\left(\frac{\varphi(t)}{2}\right) + \hat{z}\sin\left(\frac{\varphi(t)}{2}\right), \tag{7.7}$$

$$\vec{\chi}_i^{\mathrm{coorb}}(t) = \hat{q}_{\mathrm{r}}^{-1}(t)\,\vec{\chi}_i^{\mathrm{copr}}(t)\hat{q}_{\mathrm{r}}(t), \tag{7.8}$$

$$h_{\mathrm{coorb}}^{\ell,m}(t) = h_C^{\ell,m}(t)e^{im\varphi(t)}, \tag{7.9}$$

where $\hat{q}_{\mathrm{r}}(t)$ is a unit quaternion representing a rotation about the $\hat{z}$ axis by φ. Finally, using 4th order finite differences, we compute the orbital frequency

$$\omega(t) = \frac{d}{dt}\varphi(t) \tag{7.10}$$

and the spin time derivatives in the coprecessing frame, which we then transform to the coorbital frame

$$\dot{\vec{\chi}}_i^{\mathrm{coorb}}(t) = \hat{q}_{\mathrm{r}}^{-1}(t)\,\dot{\vec{\chi}}_i^{\mathrm{copr}}(t)\hat{q}_{\mathrm{r}}(t), \tag{7.11}$$

where a dot means d/dt. For the precession dynamics, we compute the angular velocity of the coprecessing frame

$$\frac{1}{2}\vec{\Omega}^{\mathrm{copr}}(t) = \lim_{dt\to 0}\frac{1}{dt}\left(\hat{q}^{-1}(t)\hat{q}(t+dt) - 1\right), \tag{7.12}$$

$$= s(t)\dot{\vec{v}}(t) - \dot{s}(t)\vec{v}(t) - \vec{v}\times\dot{\vec{v}}(t), \tag{7.13}$$

where $s(t)$ and $\vec{v}(t)$ are the scalar and vector components of $\hat{q}(t)$. We also transform $\vec{\Omega}^{\mathrm{copr}}(t)$ to the coorbital frame to obtain $\vec{\Omega}^{\mathrm{coorb}}(t)$ as in Eq. (7.11). The minimal rotation condition of the coprecessing frame ensures

$$\Omega_z^{\mathrm{coorb}}(t) = \Omega_z^{\mathrm{copr}}(t) = 0 \tag{7.14}$$

up to finite difference errors.

Given a waveform data piece $X(t)$ evaluated at a set of parameters, one would be tempted to parameterize $X(t)$ at any fixed time t_i by the mass ratio and the initial spins, and then construct a fit to $X(t_i)$ as a function of these parameters. However, we find much better fits if we instead parameterize $X(t_i)$ by the spins *at time* t_i and the mass ratio. While this is easy to do during the inspiral where we still have two BHs with individual spins, we seek a way to extend this parameterization through the merger and ringdown, where individual BH spins are no longer available. We extend, unphysically, the spin evolution through the merger and ringdown using the PN expressions

$$\frac{d}{dt}\vec{\chi}_i = \vec{\Omega}_i^{\mathrm{Spin}} \times \vec{\chi}_i\,, \tag{7.15}$$

where $\vec{\chi}_i$ is the spin in the inertial frame, and $\vec{\Omega}_i^{\mathrm{Spin}}$ is a PN expression given by Eq. A32 of Ref. [57]. $\vec{\Omega}_i^{\mathrm{Spin}}$ is a function of the orbital angular momentum vector $\hat{l}(t)$, a vector pointing from one BH to the other $\hat{n}(t)$, and the PN parameter $x(t)$. In PN theory, these quantities are typically computed from BH trajectories, but here we instead compute them from the waveform. Evaluating $\vec{\Omega}_i^{\mathrm{Spin}}$ requires several quantities that are typically computed from BH trajectories in PN theory. Since the trajectories are also not available after the merger, we compute them from the waveform. We take $\hat{l}$ and $\hat{n}$ to be the $\hat{z}$ and $\hat{x}$ axes of the coorbital frame, and we take the PN parameter x to be $\omega^{2/3}$, where ω is defined in Eq. (7.10) (see Eq. 230 of Ref. [315]). We choose $t_{\mathrm{PN}} = -100M$ and begin the PN integrations from the spins at t_{PN}. The extended spins are somewhat robust to the choice of t_{PN} as seen in Fig. 7.1. We stress that these extended spins are not physically meaningful for $t > t_{\mathrm{PN}}$, but provide a convenient parameterization of the system that leads to accurate parametric fits.

7.5 Building the Model

In this section, we describe the quantities that are computed from the waveform data pieces and stored when building the NRSur7dq2 surrogate model. The subsequent section will then describe how the NRSur7dq2 surrogate model uses these stored quantities to generate waveforms.

We first construct surrogate models for the waveform modes in the coorbital frame $h^{\ell,m}_{\mathrm{coorb}}(t)$. For $m = 0$ modes, we model the real and imaginary components. For

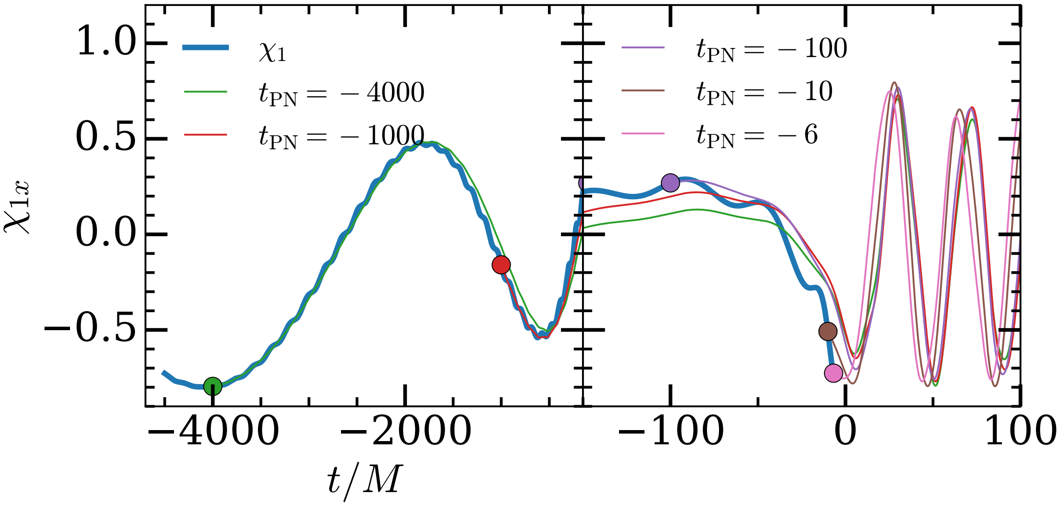

Figure 7.1: The x-component of a spin extended through merger and ringdown with PN expressions. These spins are not physically meaningful, but provide a parameterization of the system leading to accurate fits. The thick solid blue curve shows the spin from an NR simulation, and is not measured past $t = -6M$ due to the merger of the BHs. Each thin line is identical to the NR curve before some time $t_{\rm PN}$ indicated with a dot, after which the spins are evolved using Eq. (7.15). The spins during the ringdown are affected somewhat by the choice of $t_{\rm PN}$, but the overall phasing is quite similar.

$m > 0$, we compute

$$h_{\pm}^{\ell,m} = \frac{1}{2}\left(h_{\rm coorb}^{\ell,m} \pm h_{\rm coorb}^{\ell,m\,*}\right) \tag{7.16}$$

and model the real and imaginary parts of $h_{\pm}^{\ell,m}$. Each of these modeled components is considered a *waveform data piece*. We proceed according to Sec. V of Ref. [14]: For each waveform data piece, we construct a compact linear basis using singular value decomposition with a RMS tolerance of 3×10^{-4}. We then construct an empirical interpolant and determine one empirical node time T_j for each basis vector. The times T_i are chosen differently for each waveform data piece. Finally, for each T_j, we construct a *parametric fit* for the waveform data piece evaluated at T_j, which is described below. The fits are functions of the mass ratio and the coorbital spin components $\vec{\chi}_i^{\rm coorb}(t)$ evaluated at T_j. Note that the x component of a vector in the coorbital frame is roughly the component in the direction of a vector pointing from one BH to the other, the z component is along the axis of orbital angular momentum, and the y component is the remaining orthogonal direction. In addition to the resulting fit data, the empirical interpolation matrix for each of these waveform data pieces are stored in the NRSur7dq2 surrogate model.

These parametric fits use the forward-stepwise greedy fitting method described in Appendix A of Ref. [14]. We choose the basis functions to be monomials in the mass ratio and spin components. We consider up to cubic functions in the mass ratio and up to quadratic functions in the spin components. We perform $K = 20$ trials using $N_v = 50$ validation points each. The fit coefficients and the basis functions selected during the fitting procedure are stored in the NRSur7dq2 surrogate model.

We also construct parametric fits for $\omega(t)$, $\Omega^{\text{coorb}}_{\{x,y\}}(t)$, and $\dot{\chi}^{\text{coorb}}_{j\{x,y,z\}}(t)$ at selected time nodes t_i. These quantities describe the dynamics of the binary and the spins, so we call these t_i the dynamics time nodes. We attempt to choose the time nodes t_i to be approximately uniformly spaced in $\varphi(t)$ with 10 nodes per orbit. Because $\varphi(t)$ is different for different simulations, and we choose the same time nodes for all simulations, in practice our choice of 238 time nodes gives us between 8 and 15 nodes per orbit. We find that this is sufficient — including additional nodes per orbit does not improve the accuracy of the surrogate model. Our time nodes are labeled $t_0 < t_1 < \cdots < t_{234} = 100M$ plus three additional nodes $t_{\frac{1}{2}}$, $t_{\frac{3}{2}}$, and $t_{\frac{5}{2}}$, which are the midpoints of their adjacent integer time nodes. The reason for including the fractional time nodes is for Runge-Kutta time integration at the beginning of the time series, which will be made clear in the next section. In Appendix 7.11, we describe in detail the algorithm for choosing t_i, but any choice that is roughly uniformly-spaced in $\varphi(t)$ and sufficiently dense should yield a surrogate with comparable accuracy.

7.6 Evaluating the model

To evaluate the NRSur7dq2 surrogate model, we provide the mass ratio q and initial spins $\vec{\chi}_j(t_0)$ as inputs. The evaluation consists of three steps: we first integrate a coupled ODE system for the spins, the orbital phase, and the coprecessing frame, then we evaluate the coorbital waveform modes, and finally we transform the waveform back to the coorbital frame. We describe each of these steps below.

We initialize the ODE system with

- $\varphi(t_0) = 0$
- $\hat{q}(t_0) = 1$
- $\vec{\chi}^{\text{copr}}_j(t_0) = \vec{\chi}_j(t_0)$,

and we evolve these variables forward in time as follows. At a given dynamics time node t_i, we first determine $\vec{\chi}_j^{\text{coorb}}(t_i)$ by rotating the x and y components of $\vec{\chi}_j^{\text{copr}}(t_i)$ by an angle $\varphi(t_i)$ as in Eq. (7.9). We then evaluate the fits for $\omega(t_i)$, $\Omega_{\{x,y\}}^{\text{coorb}}(t_i)$, and $\dot{\chi}_{j\{x,y,z\}}^{\text{coorb}}(t_i)$ using the mass ratio q and the current coprecessing spins $\vec{\chi}_j^{\text{coorb}}(t_i)$. We set $\Omega_z^{\text{coorb}}(t_i) = 0$, and obtain $\dot{\vec{\chi}}_j^{\text{copr}}(t_i)$ and $\vec{\Omega}^{\text{copr}}(t_i)$ by rotating the x and y components of the corresponding coorbital quantities by an angle of $-\varphi(t_i)$. We evolve the coprecessing vectors instead of the coorbital vectors because the former evolve on the longer precession timescale, allowing us to take large timesteps. Finally, after computing

$$\left.\frac{d}{dt}\hat{q}(t)\right|_{t_i} = 2\hat{q}(t_i)\vec{\Omega}^{\text{copr}}(t_i)\,, \tag{7.17}$$

we obtain the time derivatives of φ, $\hat{q}$, and $\vec{\chi}_j^{\text{copr}}$ at $t = t_i$.

These time derivatives are then used to integrate φ, $\hat{q}$, and $\vec{\chi}_j^{\text{copr}}$ using an ODE solver. We desire an ODE integration method that uses few evaluations of the time derivatives to keep the computational cost of evaluating the model low. We use a fourth-order Adams-Bashforth method [316, 317] detailed in Appendix 7.12, which determines the solutions at the next node based on the time derivatives at the current and three previous nodes. This allows us to reuse fit evaluations from the previous nodes, and requires only one additional evaluation of the fits per node compared to four evaluations for a fourth-order Runge-Kutta scheme. The Adams-Bashforth integration is initialized by performing the first integration steps with fourth-order Runge-Kutta. This is why we include the three additional time nodes $t_{\frac{1}{2}}$, $t_{\frac{3}{2}}$ and $t_{\frac{5}{2}}$; they enable evaluating the midpoint increments of the initial Runge-Kutta scheme. Once we have evaluated the solutions at the time nodes t_i, we use cubic spline interpolation to determine the solutions at all times.

Now that we have φ, $\hat{q}$, and $\vec{\chi}_j^{\text{copr}}$ for all t, we then evaluate each coorbital waveform data piece. This is done by first evaluating the fits at the empirical nodes T_i using the mass ratio q and the coorbital spins at the empirical nodes $\vec{\chi}_j^{\text{coorb}}(T_i)$, and then evaluating the empirical interpolant to obtain the waveform data piece at all times. Finally, we transform the coorbital frame waveform modes back to the coprecessing frame using $\varphi(t)$ and then to the inertial frame using $\hat{q}(t)$. The NRSur7dq2 surrogate data and Python evaluation code can be found at [318].

To reduce the computational cost of transforming the coprecessing waveform modes to the inertial frame using $\hat{q}(t)$, which takes $\sim$ 1 s using all $\ell \leq 4$ modes sampled with $\delta t = 0.1M$, we reduce the number of time samples of the coorbital waveform data pieces by using non-uniform time steps. We choose 2000 time samples that are roughly uniformly spaced in the orbital phase, using the same method used to choose the dynamics time nodes described in Appendix 7.11. This is sufficiently many time samples to yield negligible errors when interpolating back to the dense uniformly-spaced time array using cubic splines on the real and imaginary parts of the waveform modes.

Integrating the ODE system takes $\sim$ 6 ms, where the numerical computations are performed by a Python module written in C. Interpolating the results of the ODE integration to the 2000 time samples described above takes $\sim$ 3 ms using cubic splines. Evaluating the coorbital waveform surrogate takes $\sim$ 5 ms, and transforming the modes to the inertial frame takes $\sim$ 25 ms, for a total of $\sim$ 40 ms. Variations in the evaluation time can increase this up to $\sim$ 50 ms. Restricting to only $\ell = 2$ modes can reduce this cost to $\sim$ 20 ms. If we wish to sample the surrogate waveform at the same time nodes as the original numerical relativity simulations, which is a uniformly-spaced time array with $\delta t = 0.1M$, the modes are interpolated to these points using cubic splines. This costs $\sim$ 10 ms per mode, for a total cost of $\sim$ 300 ms when all $\ell \leq 4$ modes are interpolated in this way. We note, however, that the original NR simulations are oversampled for typical gravitational wave data analysis purposes. For example, a sampling rate of 4096 Hz for a $M = 60M_\odot$ binary has $\delta t \approx 0.83M$, leading to an evaluation cost of $\sim$ 100 ms.

7.7 Surrogate Errors

We use two error measures to quantify the accuracy of the surrogate model. Given two sets of waveform modes h_1 and h_2, we first compute

$$\mathcal{E}[h_1, h_2] = \frac{1}{2}\frac{\sum_{\ell,m}\int_{t_0}^{t_f}|h_1^{\ell,m}(t) - h_2^{\ell,m}(t)|^2 dt}{\sum_{\ell,m}\int_{t_0}^{t_f}|h_1^{\ell,m}(t)|^2 dt}\,, \tag{7.18}$$

which is introduced in Eq. 21 of [14]. Since we have aligned all the NR waveforms at $t = t_0$ and the surrogate model reproduces this alignment, we do not perform any time or phase shifts when computing $\mathcal{E}$.

For these comparisons, we use modes $\ell \leq 5$; if a mode is not included in a particular waveform model, we assume this mode is zero for that model. Since the NRSur7dq2 model does not contain $\ell = 5$ modes, this ensures that the errors discussed below include the effect of neglecting $\ell = 5$ and higher modes.

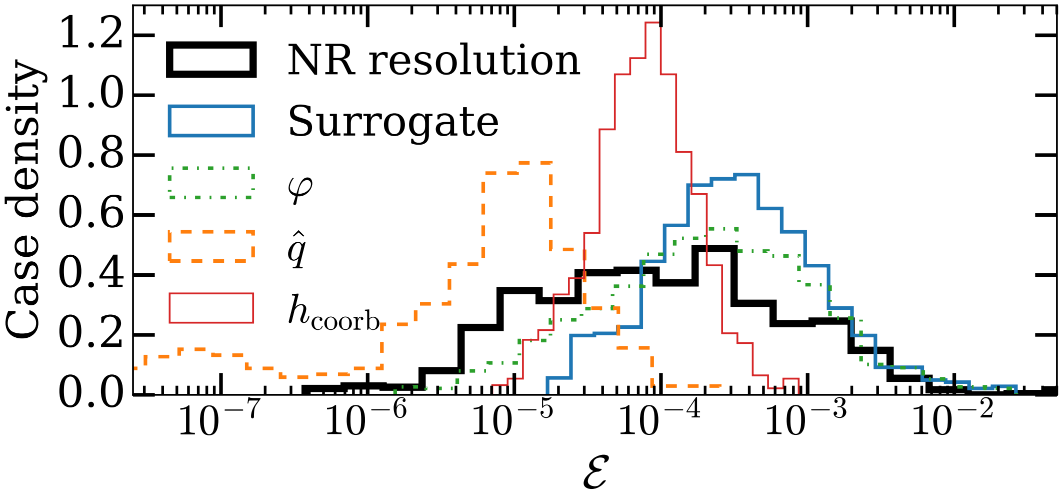

Figure 7.2: Error histograms for $\mathcal{E}$ defined in Eq. (7.18), normalized such that the area under each curve is 1 when integrated over $\log_{10}(\mathcal{E})$. The largest surrogate errors are comparable to the largest NR resolution errors, which compare high and medium resolution NR simulations to estimate the error in the NR waveforms. The error in the orbital phase φ is the dominant error contribution to the surrogate.

Histograms of $\mathcal{E}$ for all 886 NR waveforms are given in Fig. 7.2. For all curves in the figure, h_1 is the highest available resolution NR waveform. For the thick solid black curve, h_2 is the same NR waveform as h_1, except computed at a lower numerical resolution, so this curve represents an estimate of the numerical truncation error in the NR waveforms used to build the surrogate model. For the solid blue curve, h_2 is the NRSur7dq2 surrogate waveform evaluated with the same mass ratio and initial spins of h_1. Note that since the surrogate was trained using all NR waveforms, this is an *in-sample* error.

The remaining curves in Fig. 7.2 indicate the in-sample error contribution from each of the three main waveform data pieces in the surrogate waveform: the orbital phase φ (dash-dotted green curve), the quaternions $\hat{q}$ representing the precession (dashed orange curve), and the waveform modes in a coorbital frame h_{coorb} (thin solid red curve). For these curves, h_2 is computed by using the surrogate evaluation for one waveform data piece and the NR evaluation of the other pieces. The orbital phase errors give rise to the largest surrogate errors, indicating that efforts to improve the surrogate model should be focused on improving the orbital phasing.

We then compute mismatches

$$1 - \frac{\langle h_1, h_2 \rangle}{\sqrt{\langle h_1, h_1 \rangle \langle h_2, h_2 \rangle}}, \tag{7.19}$$

where $\langle \cdot, \cdot \rangle$ is a noise-weighted inner product computed in the frequency domain, as in Sec. VI.B of Ref. [14]. We use a flat power spectral density to avoid a dependence on the total mass of the system. The mismatches are minimized over timeshifts, polarization angle shifts, and shifts in the azimuthal angle of the direction of GW propagation, where the system's orbital angular momentum is initially aligned with the $\hat{z}$ axis.

Histograms of the mismatches are given in Fig. 7.3 and are comparable to the top panel of Fig. 17 in Ref. [14]. To estimate the *out-of-sample* errors of the surrogate model, we perform a 20-fold cross-validation test. This is done by first randomly dividing the 886 NR waveforms into 20 sets of 44 or 45 waveforms. For each set, we build a *trial surrogate* using the waveforms from the other 19 sets. The trial surrogate is then evaluated at the parameters corresponding to the waveforms in the chosen *validation* set, and the results are compared to the NR waveform. These cross-validation mismatches are given by the dashed purple curve. They are quite similar to the in-sample errors given by the solid blue curve, indicating that we are not overfitting the data. We also compute mismatches for a fully precessing effective one-body model (SEOBNRv3 [152]), and for a phenomenological waveform model that includes some, but not all, effects of precession (IMRPhenomPv2 [66]). These models have mismatches more than an order of magnitude larger than our NRSur7dq2 surrogate model. Both IMRPhenomPv2 and SEOBNRv3 depend on a parameter f_{ref}, which is a reference frequency at which the spin directions are specified. For SEOBNRv3, which is a time-domain model, we choose f_{ref} so that the waveform begins at $t = t_0$. For IMRPhenomPv2, which is a frequency-domain model, we minimize the mismatches over f_{ref}, using an initial guess of twice the orbital frequency of the NR waveform at $t = t_0$. While all of the mismatches can be decreased by minimizing over additional parameters such as BH masses and spins, this would result in biased parameters when measuring the source parameters of a detected GW signal.

We then compute mismatches using the advanced LIGO design sensitivity noise curve [217] using various total masses M. For each mass M, we obtain histograms as in Fig. 7.3, and we show the median and 95th percentile mismatches from these histograms in Fig. 7.4. We note that for $M \lesssim 114 M_{\odot}$ some or all waveforms begin above 10 Hz and do not cover the full design sensitivity frequency band. We find that the 95th percentile mismatches of our surrogate model are similar to the corresponding NR mismatches, except for total masses above $160 M_{\odot}$ where the NR mismatches are slightly smaller. The NRSur7dq2 surrogate yields mismatches at

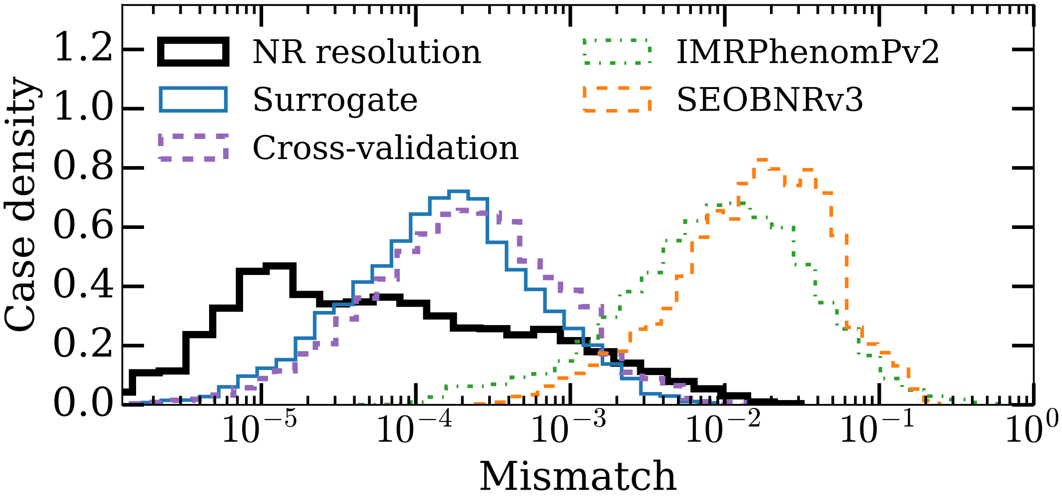

Figure 7.3: Mismatch histograms computed in the frequency domain with a flat power spectral density. The NR resolution mismatches compare waveforms from high and medium resolution NR simulations. This can be an overestimate of the error in the high resolution NR waveform, leading to some NR resolution mismatches being larger than the surrogate mismatches. We note that the IMRPhenomPv2 model does not contain all spin components.

least an order of magnitude smaller than the other waveform models for all total masses investigated.

Figure 7.5 shows the real part of $h^{2,2}(t)$ for the cases leading to the largest mismatches in Fig. 7.3. The top panel shows the case leading to the largest surrogate cross-validation mismatch, and the bottom panel shows the case leading to the largest SEOBNRv3 mismatch. The surrogate waveforms shown are evaluated using the appropriate trial surrogate, so that they were not trained on the NR waveforms they are compared with. All waveforms are aligned to have their peak amplitude at $t = 0$ and are rotated to have their orbital angular momentum aligned with the z axis at $t = t_0 = -4500M$. In the top panel, we see that both the SEOBNRv3 and surrogate waveforms have a similar phasing error around $t = -50M$. The phasing error of the surrogate does not grow significantly larger through merger and ringdown, so most of this error can be removed with a time and phase shift. For the SEOBNRv3 waveforms in both the upper and lower panels, the phasing error changes significantly during the merger; therefore this error does not decrease significantly even after performing a time and phase shift. In the top panel of Figure 7.5, the IMRPhenomPv2 waveform does as well as the surrogate; in the bottom panel, the IMRPhenomPv2 waveform has large errors in both phase and amplitude.

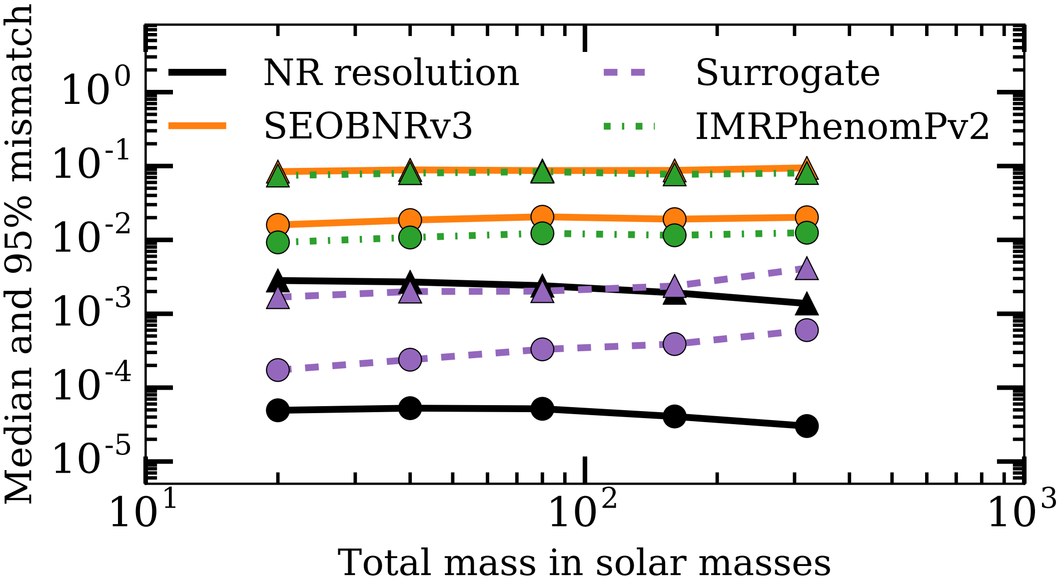

Figure 7.4: Median (circles) and 95th percentile (triangles) mismatches of all 866 cases computed with the advanced LIGO design sensitivity curve. The surrogate mismatches are computed using trial surrogates, as in the cross-validation curve of Fig. 7.3.

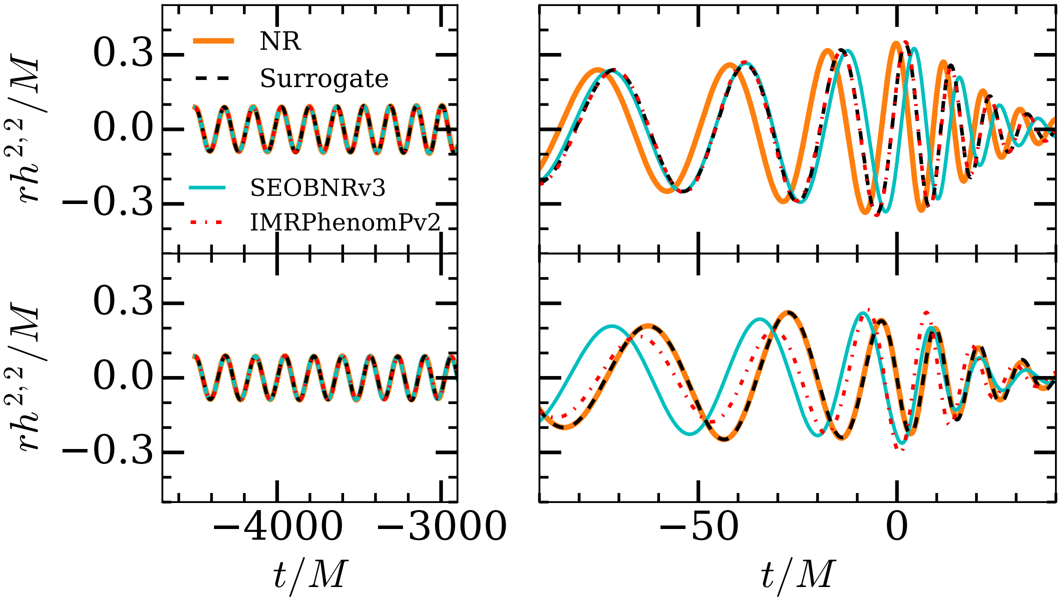

Figure 7.5: The real part of time domain waveforms for the case leading to the largest surrogate mismatch (top) and the largest SEOBNRv3 mismatch (bottom). The surrogate waveforms are evaluated using trial surrogates which were not trained with the NR waveform shown. The top panel uses SXS:BBH:0922 with $q \approx 2$, $\vec{\chi}_1(t_0) \approx 0.8\hat{z}$, and $\vec{\chi}_2(t_0) \approx -0.8\hat{y}$. The lower panel uses SXS:BBH:0900 with $q \approx 2$, $\vec{\chi}_1(t_0) \approx (0.29, -0.74, 0.02)$ and $\vec{\chi}_2(t_0) \approx (0.43, -0.34, 0.58)$.

7.8 Discussion and Conclusions

Within its range of validity, our NRSur7dq2 surrogate model is nearly as accurate as performing new NR simulations. The surrogate model takes only $\sim$ 50 ms to evaluate on a single CPU core, making it sufficiently fast for current gravitational wave data analysis applications such as parameter estimation. This evaluation time can be compared to $\mathcal{O}$(weeks) on dozens of CPU cores to perform a new NR simulation, decreasing the cost in CPU-hours by $\mathcal{O}(10^8)$. The NRSur7dq2 surrogate model data along with Python evaluation code is publicly available for download at [318].

Our surrogate model is limited to mass ratios $q \leq 2$ and spin magnitudes $|\vec{\chi}_{1,2}| \leq 0.8$. While in principle the parametric fits can be extrapolated to more extreme mass ratios and spin magnitudes, we do not expect extrapolation to yield accurate waveforms. However, these limits can be extended in future versions of our surrogate model by performing NR simulations with larger mass ratios and spins.

Additionally, the waveforms produced by NRSur7dq2 are limited in duration to $4500M$ before the peak amplitude. This covers frequencies $f \geq 20\,\mathrm{Hz}$ for all systems with $M \gtrsim 57M_\odot$. For systems with lower total masses, or for systems with $M \lesssim 114M_\odot$ when including frequencies down to 10 Hz, longer waveforms are needed. In future work, we plan to overcome this limitation by hybridizing with either PN or SEOBNRv3 [215, 298–300], either by hybridizing the NR waveforms before building the surrogate or by hybridizing the surrogate waveforms. Longer NR waveforms would then be needed to test the accuracy of the hybridization step.

7.9 Acknowledgments

We thank Matt Giesler for helping to carry out the new SpEC simulations used in this work. We thank Saul Teukolsky, Patricia Schmidt, and Rory Smith for helpful discussions. This work was supported in part by the Sherman Fairchild Foundation and by NSF grants CAREER PHY-1151197 and PHY-1404569. Computations were performed on NSF/NCSA Blue Waters under allocation PRAC ACI-1440083; on the NSF XSEDE network under allocation TG-PHY100033; and on the Zwicky cluster at Caltech, which is supported by the Sherman Fairchild Foundation and by NSF award PHY-0960291. This paper has been assigned YITP report number YITP-17-44.

7.10 Sparse grid parameters

We take the polar and azimuthal spin angles of $\vec{\chi}_i$ to be θ_i and ϕ_i respectively, for $i \in \{1,2\}$. We can then parametrize our 7-dimensional parameter space by

- $q \in [1,2]$,
- $|\vec{\chi}_i| \in [0,0.8]$,
- $\theta_i \in [0,\pi]$,
- $\phi_i \in [0,2\pi]$.

The range of each of these variables is some closed interval $[a,b]$. For a variable x with range $[a,b]$, we define a grid of N uniformly-spaced points

$$g_x^N = \{a + \frac{n}{N-1}(b-a) : n = 0, \ldots, N-1\} \tag{7.20}$$

where $N \geq 2$. We then define a sequence of grids

$$G_x \equiv G_x^0,\, G_x^1,\, \ldots \tag{7.21}$$

where

$$G_x^n = g_x^{f_x(n)} \tag{7.22}$$

for some monotonically increasing function $f_x(n)$. We call G_x^n the *level n grid for x*. We take

$$f_q(n) = f_{|\vec{\chi}_i|}(n) = 1 + 2^n, \tag{7.23}$$

$$f_{\theta_i}(n) = 1 + 2^{n+1}, \tag{7.24}$$

$$f_{\phi_i}(n) = 1 + 3 \cdot 2^n\,. \tag{7.25}$$

These choices ensure that $G_x^n \subset G_x^{n+1}$, and that the level 0 grids already give a description of the parameter space that doesn't leave out any phenomenology; the level 0 grids for θ_i contain the midpoint $\pi/2$ leading to precession, and the level 0 grids for ϕ_i contain 3 unique points (since $\phi_i = 0$ and $\phi_i = 2\pi$ lead to the same physical spin) in order to get at least some resolution of features that behave like $\sin(\phi_i + \phi + 0)$.

We have already seen that $\phi_i = 0$ and $\phi_i = 2\pi$ correspond to the same physical spin, but we will have many other scenarios where two combinations of variables lead to the same physical configuration. For example, if $|\vec{\chi}_1| = 0$, all combinations of θ_1 and ϕ_1 lead to the same physical configuration. We will neglect these degenerate combinations for now, and remove them later on.

Dense grids in parameter space could be constructed as

$$G^n_{\text{dense}} = G^n_q \times \prod_{i=1,2} G^n_{|\vec{\chi}_i|} \times G^n_{\theta_i} \times G^n_{\phi_i} \tag{7.26}$$

where $\times$ and $\prod$ denote the cartesian product. While the 1-dimensional grids grow in size as $O(2^n)$, these dense grids grow in size as $O(2^{7n})$ or as the seventh power of the size of the 1-dimensional grids. This is known as the *curse of dimensionality*; the amount of data needed often grows exponentially with the dimensionality. Sparse grids [314] overcome the curse of dimensionality by using a sparse product such that the grids grow in size as $O(2^n)$. If G_x and G_y are two sequences of grids, we define the *sparse product* of G_x with G_y to be $G_{x,y} = G_x \bullet G_y$ where

$$G^n_{x,y} = \bigcup_{k=0}^{n} G^k_x \times G^{n-k}_y \,. \tag{7.27}$$

We now define the sparse grids for our parameter space by the sequence of grids

$$G = G_q \bullet G_{|\vec{\chi}_1|} \bullet G_{\theta_1} \bullet G_{\phi_1} \bullet G_{|\vec{\chi}_2|} \bullet G_{\theta_2} \bullet G_{\phi_2} \tag{7.28}$$

such that

$$\begin{aligned} G^n = \bigcup_{\sum_{i=1}^{7} k_i = n} \; & G^{k_1}_q \times G^{k_2}_{|\vec{\chi}_1|} \times G^{k_3}_{\theta_1} \times G^{k_4}_{\phi_1} \\ & \times G^{k_5}_{|\vec{\chi}_2|} \times G^{k_6}_{\theta_2} \times G^{k_7}_{\phi_2} \end{aligned}$$

We performed 361 new NR simulations based on parameters in G^1. We removed physically identical configurations. We also removed configurations with $\vec{\chi}_2 \propto \hat{z}$, which are within the parameter space of the NRSur4d2s surrogate model which was already covered by the 276 NRSur4d2s NR simulations.

7.11 Time sampling

We wish to choose time nodes $t_0 < t_1 < \ldots < t_f$ that are roughly uniformly spaced in the orbital phase $\varphi(t)$ for all cases. Given some number N, we choose time nodes yielding roughly N nodes per orbit. Since different NR waveforms have different orbital frequencies, they will have a different number of time nodes per orbit. Our scheme for choosing the time nodes given N is based on the leading order PN expression for $\omega(t)$ during the inspiral, smoothly transitioning to a maximum value of $\omega = 2\pi/(20M)$ during the ringdown. We do this by computing a bounded time

$$\tilde{t}(t) = -1.7 + \frac{1}{2}\left((t+5) - \sqrt{(t+5)^2 + 25}\right) \tag{7.29}$$

and then choosing

$$\omega_{\mathrm{ref}}(t) = \omega_{0\mathrm{PN}}(\tilde{t}(t)) = \left(\frac{64}{5}(-\tilde{t}(t))\right)^{-\frac{3}{8}} . \tag{7.30}$$

We then use spacings between nodes $t_{j+1} - t_j = \omega_{\mathrm{ref}}(t_j)$.

7.12 Fourth-order Adams-Bashforth method

We integrate the ODE system on a non-uniformly spaced grid of time nodes $t_0 < t_1 < \ldots < t_f$ using a fourth-order Adams-Bashforth scheme [316, 317]. We denote the solution $\vec{y}(t)$, and at each time node t_i we can evaluate fits to determine

$$\frac{d\vec{y}}{dt} = \vec{f}(t;\vec{y}) . \tag{7.31}$$

We first integrate up to t_3 using a Runge-Kutta fourth-order scheme.

Once we have integrated up to t_i for $i >= 3$, we have previously evaluated

$$\vec{k}_j = \vec{f}(t_j;\vec{y}(t_j)) \tag{7.32}$$

for $0 \le j < i$, and we now evaluate $\vec{k}_i$. We approximate $\vec{g}(t) = \vec{f}(t;\vec{y}(t))$ by a cubic function

$$\vec{g}(t) \approx \vec{g}_3(t) = \vec{A} + \vec{B}(t-t_i) + \vec{C}(t-t_i)^2 + \vec{D}(t-t_i)^3 . \tag{7.33}$$

The coefficients are chosen such that $\vec{g}_3(t_j) = \vec{g}(t_j) = \vec{k}_j$ for $i-3 \le j \le i$, giving $\vec{A} = \vec{k}_i$, and

$$\begin{bmatrix} \vec{B} \\ \vec{C} \\ \vec{D} \end{bmatrix} = \begin{bmatrix} \frac{\delta_{-1,0}\delta_{-2,0}}{\Delta_1} & \frac{\delta_{-1,0}\delta_{-3,0}}{\Delta_2} & \frac{\delta_{-2,0}\delta_{-3,0}}{\Delta_3} \\ \frac{\delta_{-2,0}+\delta_{-1,0}}{\Delta_1} & \frac{\delta_{-1,0}+\delta_{-3,0}}{\Delta_2} & \frac{\delta_{-2,0}+\delta_{-3,0}}{\Delta_3} \\ \frac{1}{\Delta_1} & \frac{1}{\Delta_2} & \frac{1}{\Delta_3} \end{bmatrix} \begin{bmatrix} \vec{k}_i - \vec{k}_{i-3} \\ \vec{k}_i - \vec{k}_{i-2} \\ \vec{k}_i - \vec{k}_{i-1} \end{bmatrix} .$$

Here, $\delta_{n,m} = t_{i+m} - t_{i+n}$ and

$$\Delta_1 = \delta_{-3,-2}\delta_{-3,-1}\delta_{-3,0} , \tag{7.34}$$

$$\Delta_2 = \delta_{-3,-2}\delta_{-2,-1}\delta_{-2,0} , \tag{7.35}$$

$$\Delta_3 = \delta_{-2,-1}\delta_{-3,-1}\delta_{-1,0} . \tag{7.36}$$

Finally, we approximate

$$\begin{aligned}
\vec{y}(t_{i+1}) &= \vec{y}(t_i) + \int_{t_i}^{t_{i+1}} g(t)dt \\
&\approx \vec{y}(t_i) + \int_{t_i}^{t_{i+1}} g_3(t_j)dt \\
&= \vec{y}(t_i) + \delta_{0,1}\vec{A} + \frac{1}{2}\delta_{0,1}^2\vec{B} + \frac{1}{3}\delta_{0,1}^3\vec{C} + \frac{1}{4}\delta_{0,1}^4\vec{D} \,.
\end{aligned}$$

8 Other work

This chapter presents a brief summary of papers to which I had a minor contribution.

8.1 Detecting Gravitational-Wave Memory with LIGO: Implications of GW150914

This section discusses a paper published in Physical Review Letters on August 5 2016 [319], in issue 6 of volume 117, which can be found at `https://link.aps.org/doi/10.1103/PhysRevLett.117.061102`. A preprint can be found at `https://arxiv.org/abs/1605.01415`.

We present a method of detecting gravitational wave memory by combining the gravitational wave signals from multiple events. The memory signal is a small permanent alteration of spacetime, which will alter the distance between test particles in flat space as it passes through them. This can be seen in the top panel of Fig. 8.1, where the blue waveform settles down to a non-zero strain after $t = 0$. It is found that a confident detection of the memory could be obtained with fewer than 100 events similar to GW150914. The method involves aligning the memory pulses seen in the detectors (see the dashed curve in the inset of Fig. 8.1) in time, and adding up many such signals such that they stack coherently.

For a waveform consisting only of $(\ell, \pm 2)$ modes there is a parameter degeneracy involving a gravitational wave polarization shift and orbital phase shift of $\pi/2$, which would flip the sign of the memory signal. With N detections and no knowledge of the sign of the memory in the LIGO detectors, the signal-to-noise ratio of the memory signal would only grow as $\sqrt[4]{N}$ with an excess-power search. Determining the sign of the memory improves this to $\sqrt{N}$, and significantly decreases the number of gravitational wave detections needed for a confident memory detection. We therefore compute a degeneracy-breaking signal using odd m modes, using the numerical relativity surrogate model [269] presented in chapter 5. Rather than do

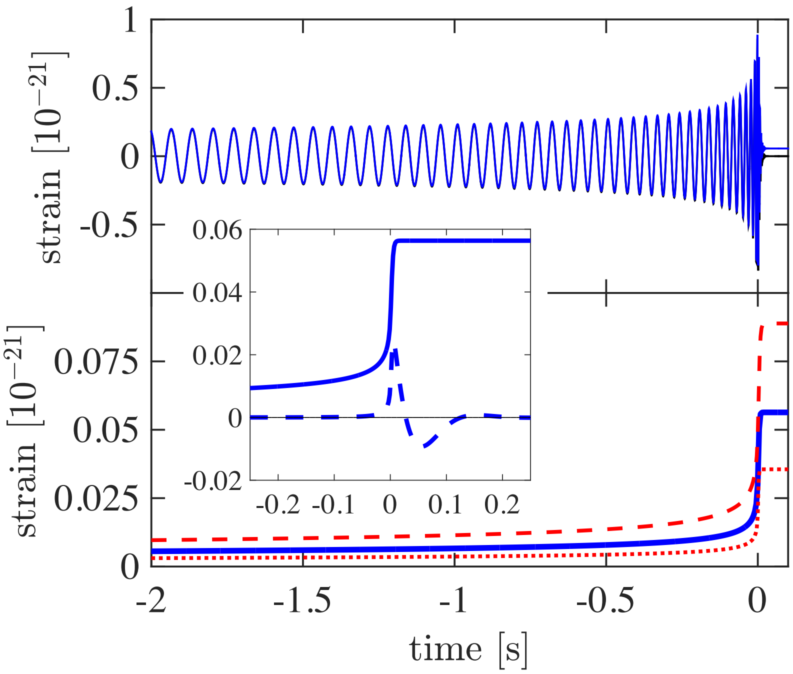

Figure 8.1: Top panel: Gravitational-wave strain for parameters consistent with GW150914 [20, 250]. The blue and black curves show the strain with and without including the memory signal, respectively. Bottom panel: The memory signal alone (blue). The red dotted and dashed curves show the memory signals for equal mass binaries with total masses of $40M_{\odot}$ and $100M_{\odot}$, compared to $65M_{\odot}$ for the blue curve. Inset: An enlarged version of the blue curve from the bottom panel, before (solid) and after (dashed) a high-pass filter to show the signal visible in aLIGO detectors.

a full parameter estimation study to determine the uncertainty of the sign of the memory for each signal, we estimate our confidence in determining the sign of the memory using the signal-to-noise ratio of this degeneracy-breaking signal.

8.2 An architecture for efficient multimodal gravitational wave parameter estimation with linear surrogate models

This section discusses a paper submitted to Classical and Quantum Gravity [17]. A preprint can be found at `https://arxiv.org/abs/1701.01137`.

Gravitational wave parameter esimation [20] is a computationally expensive task. The full 15-dimensional parameter space of extrinsic and intrinsic parameters consists of two black hole masses, two 3-dmensional dimensionless black hole spin vectors, four spacetime coordinates describing the location and time of the merger, and three angles describing the orientation of the binary. Brute-force sampling can require $\mathcal{O}(10^7)$ samples or more, even when using sophisticated Markov chain Monte-Carlo [7] or Nested sampling [8] techniques. Each sample consists of choosing a point in parameter space, evaluating the corresponding signals we would expect to see in the detectors for the given parameters using a waveform model, and comparing this to the observed data.

Surrogate models, also called reduced-order models (ROM), reduce the cost of parameter estimation by speeding up waveform models. Independently, a method was developed [9] to reduce the number of waveform evaluations by separating the parameter space into intrinsic and extrinsic parameters, and evaluating the likelihood analytically over extrinsic parameters for a given set of intrinsic parameters. This paper combines these approaches in a natural way for the case of linear surrogate models, which express waveform modes as a sum over basis vectors multiplied with analytic coefficients that depend on the intrinsic parameter space. The result is an analytic likelihood function that can be computed from a handful of inner products between basis vectors and detector data, as well as inner products between two basis vectors.

In practice, most surrogate models are built for dimensionless waveforms by factoring out the total mass M, so the basis vectors will depend on M. This means that for each M we must compute some inner products, at which point we have an analytic likelihood function valid for this particular total mass M. Additionally, the basis coefficients of linear surrogates are typically either obtained through least-squares fits or tensor-spline interpolation (as in Appendix A). In either case, while analytic, the full likelihood function is impractical to express in closed form and will typically be evaluated numerically. Nevertheless, these coefficients are fast to evaluate, which leads to very rapid parameter estimation.

8.3 Black hole spectroscopy with coherent mode stacking

This section discusses a paper published in Physical Review Letters on April 20 2017 [319], in issue 16 of volume 118, which can be found at `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.118.161101`. A preprint can be found at `https://arxiv.org/abs/1701.05808`.

This paper presents a method of measuring subdominant modes in the ringdown phase of binary black hole mergers by coherently stacking the signals from many detections. Such a measurement could be used to test General Relativity (GR), as was done with the dominant quasinormal mode (QNM) of the ringdown of GW150914 [252]. The method used is somewhat similar to that used to detect gravitational wave memory [319], in that it involves aligning multiple detections such that they can be stacked coherently.

For each detected signal, the results of parameter estimation will lead to knowledge of the remnant black hole mass and spins, which fully determines the QNM frequencies. Each signal can then be scaled in time such that the $(3,3)$ QNM (for example) has the same frequency f_0 in all cases. The much larger $(2,2)$ mode will then have a frequency $\leq \frac{2}{3} f_0$ depending on the spin of the remnant, so it can be distinguished from the $(3,3)$ mode and will not contaminate the measurement. The signals must also be phase-shifted such that the $(3,3)$ modes are in phase with each other. By investigating the waveform modes produced by the non-spinning [269] and NRSur4d2s [14] surrogate models, we found that the phase of the $(3,3)$ mode is well-determined by knowledge of the phase of the $(2,2)$ mode even when significant parameter uncertainty (for example in the black hole masses) is present. While the mass ratio and black hole spins do affect the phase difference between these modes at merger, the difference is small. We find that the $(3,3)$ QNM can be measured confidently with ~ 1 yr of aLIGO data at design sensitivity using a merger rate of $40\,\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$.

8.4 Gravitational Waves from Binary Black Hole Mergers Inside of Stars

This section discusses a paper submitted to Physical Review Letters [18]. A preprint can be found at `https://arxiv.org/abs/1704.07383`.

The Fermi satellite observed a $\gamma-$ray event [12] coincident with GW150914 [250]. Binary black hole (BBH) mergers are not expected to produce an electromagnetic counterpart [13], so it would be surprising for these to be directly related. It has been proposed that a BBH merger that occurs within a massive star undergoing gravitational collapse could lead to such an electromagnetic counterpart [320]. We rule out this scenario by evolving BBH systems within a gas of constant density, as the resulting waveforms are inconsistent with waveforms from BBH mergers in vacuum even with densities lower than those that would be realistically expected within such a star.

We performed numerical relativity simulations of two $30M_\odot$ non-spinning black holes in a gas environment with initially constant densities ρ_0. Increasingly large densities of gas lead to increasingly faster chirps of the gravitational wave (GW) signal due to converting orbital energy into kinetic and internal energy of the gas. For gas densities $\rho_0 \geq\sim 10^7\,\mathrm{gcm}^{-3}$, these GW signals are inconsistent with GWs from *any* BBH merger occuring in vacuum, even after varying the black hole masses and spins to improve the match. Typical densities in the core of a presupernova star are $\sim 10^9 - -10^{10}\,\mathrm{gcm}^{-3}$, and even at a radius of $1000\,\mathrm{km}$ densities remain above $\sim 10^7\,\mathrm{gcm}^{-3}$. Since GW150914 is consistent with a vacuum BBH GW signal, we therefore conclude that GW150914 is inconsistent with the proposed scenario. Here, we are making a mild assumption that BBH mergers within collapsing stars with *other* masses and spins will also lead to GWs inconsistent with GW150914. Since this particular system leads to GWs inconsistent with *any* vacuum GWs, and since the inconsistency is due to an overly rapid chirp due to drag forces which should be present in any BBH system within gas, we find this assumption to be justified.

A Tensor Spline Surrogates

A.1 Introduction

When building surrogate models of waveforms from numerical relativity simulations as done in chapters 5, 6 and 7, we often have very few waveforms per dimension of the parameter space. For example, the surrogate model constructed in chapter 7 uses only 886 waveforms in a $7D$ parameter space, which is $886^{\frac{1}{7}} \sim 2.64$ points per dimension. The waveforms themselves are highly oscillatory and would require many points per dimension for accurate direct interpolation, so the methods in these chapters focus on transforming the waveforms into other simple functions that are much easier to interpolate. There are at least two sacrifices made in using these transformations.

First, the computational cost to evaluate the surrogate model can be larger than when interpolating the waveform directly. The additional cost comes partially from having to reconstruct the waveform from these simple functions. This can involve rotating waveform modes from a coprecessing frame back to the inertial frame, which involves evaluating a Wigner D matrix at each timestep and can take $\mathcal{O}(1\,\mathrm{s})$. Due to the very low number of points per dimension, efficient interpolation methods such as cubic tensor-spline interpolation cannot be used and least-squares fits with carefully chosen functional forms are used instead (see, for example, Sec. 6.10 of chapter 6). These fits may use $\mathcal{O}(100)$ coefficients, each of which multiplies the product of d basis functions for a d-dimensional parameter space, and evaluating them can be more costly than cubic tensor-spline interpolation.

The second sacrifice is that the waveform modes are no longer expressed as the sum over time-dependent basis functions with parameter-dependent coefficients. A *linear surrogate model* of waveform modes $\{h^{\ell,m}(x;\vec{\lambda})\}_{\ell,m}$ computes $h^{\ell,m}(x;\vec{\lambda})$ using

$$h^{\ell,m}(x;\vec{\lambda}) = \sum_{i=1}^{N} c_i^{\ell,m}(\vec{\lambda}) b_i^{\ell,m}(x)\,. \tag{A.1}$$

Here, $\vec{\lambda}$ is a parameter vector containing (for example) the black hole masses and spins, and x can either time or frequency depending on whether the surrogate model is for time-domain or frequency-domain waveforms. The advantages of having a linear surrogate model include the ability to construct reduced-order quadratures [145, 155] or an analytic likelihood function [17], either of which can be used to speed up gravitational wave parameter estimation. In either case, the gravitational wave data observed in a detector is projected directly on to the basis $\{b_i^{\ell,m}(x)\}_{i=1}^N$ once, and all likelihoods are calculated from the coefficients $c_i^{\ell,m}(\vec{\lambda})$ without any further costly inner products with detector data. Using a *waveform decomposition* as in chapters 6 and 7, or even building linear bases for the waveform mode amplitudes and phases as done in chapter 5, leads to a *non-linear surrogate model* where the waveform modes are not expressed in the form of Eq. A.1.

Fortunately, it is rather straight-forward to build a linear surrogate model of a non-linear surrogate model. That is, if we have some underlying waveform model denoted by A (for example, performing numerical relativity simulations) and build a non-linear surrogate model $B(A)$ from the results of a handful of evaluations of A, we can then build a linear surrogate model $C(B(A))$ from many evaluations of B. Here, the notation $X(Y)$ denotes that model X is a surrogate model of Y, and X attempts to replicate the output of Y. In this case where A involves performing numerical relativity simulations, constructing $C(A)$ directly would be far too costly, as $C(X)$ will require many evaluations of X. On the other hand, constructing $B(X)$ can be done with relatively few evaluations of X, and B is sufficiently fast that we can evaluate it many times. This will allow us to construct $C(B(A))$.

In Sec. 6.8 we built a frequency domain linear surrogate model which we call NRSur4d2s_FDROM by taking the fast fourier transform (FFT) of the time domain waveforms generated by a time domain non-linear surrogate model called NRSur4d2s. In that case, C was NRSur4d2s_FDROM, and evaluating B consisted of taking the FFT of the output of the NRSur4d2s surrogate model. Here we present a generalization of the method used in Sec. 6.8 and include more detail. Building surrogate models using cubic tensor-spline interpolation has also been done for effective-one-body (EOB) models [239, 265]. The method presented here is very similar, but uses empirical interpolation rather than interpolating basis coefficients, and uses the waveform mode real and imaginary parts rather than amplitude and phase in order to yield a linear surrogate model.

Suppose we already have a waveform model B which is computationally inexpensive enough that we can evaluate it $O(10^d)$ times, where d is the dimensionality of the parameter space vectors $\vec{\lambda}$. While in our case B has always been a numerical relativity surrogate model $B(A)$, this method applies to any waveform model B meeting the above cost requirement. We will proceed by first constructing moderately dense training and validation sets. Next we will choose an appropriate set of time or frequency samples. We will then constructing a reduced basis and empirical interpolant for each mode using the training set, and estimate the errors using the validation set. Finally, we will generate cubic tensor-spline interpolants at each empirical node from increasingly dense grids in parameter space, until our surrogate error is sufficiently small.

A.2 Constructing a training set

The method presented in this chapter assumes a parameter space with topology $\mathcal{I}^d$, where $\mathcal{I}$ is the closed unit interval $[0,1]$. Each parameter dimension λ_j lies in some closed interval $[a_j, b_j]$, which can be mapped to the unit interval with the linear transformation

$$T_j(\lambda_j) = \frac{\lambda_j - a_j}{b_j - a_j} . \tag{A.2}$$

Here, $j \in \{1, \ldots, d\}$ and d is the dimensionality of the parameter space. We will denote the mapped parameters by $\vec{\mu} = \vec{T}(\vec{\lambda})$, where $\mu_j = T_j(\lambda_j)$. Given a positive integer $N \geq 2$, we can construct a $1D$ grid of N uniformly-spaced points

$$G_N = \left\{ \frac{n}{N-1} : n = 0, \ldots, N-1 \right\} . \tag{A.3}$$

Given d such integers $N_1, \ldots, N_d$, we define the d-dimensional grid by the cartesian product of the corresponding 1-dimensional grids

$$G_{N_1,\ldots,N_d} = G_{N_1} \times \ldots \times G_{N_d} . \tag{A.4}$$

We will consider building a linear surrogate model for a single waveform mode $h^{\ell,m}(x; \vec{\lambda})$, which we will denote $h(x; \vec{\lambda})$ for notational simplicity. We note, however, that in practice it is useful to build linear surrogate models for all (ℓ, m) modes simultaneously, using the same grids.

For simplicity we will restrict to d-dimensional grids $G_{N_1,\ldots,N_d}$ where $N_1 = \ldots = N_d = N$, and denote $G_{N_1,\ldots,N_d} = G_N^d$. Removing this restriction can be useful when h varies more rapidly over some parameter dimensions than others. We construct two sets of waveforms

$$H_{1,2} = \{h(x; \vec{T}^{-1}(\vec{\mu}_k) : \vec{\mu}_k \in G^d_{N_{1,2}}\} \tag{A.5}$$

by evaluating the waveform model B and storing the results, where $N_2 > N_1$. We will use H_1 as a *training set* and H_2 as a *validation set*.

A.3 Determining the domain samples

The time to evaluate the linear surrogate model $C(B)$ should be dominated by tensor-spline interpolations and evaluating empirical interpolants. Evaluating an empirical interpolant consists of multiplying a matrix of dimension $L \times n$ with a vector of dimension n, where L is the number of time (or frequency, depending on what x represents) samples and n is the number of basis vectors in the empirical interpolation basis. It is therefore useful to have a sparse sampling of the waveforms to keep L small, up to the point where either the cost of empirical interpolation becomes negligible to the cost of tensor-spline interpolation or the error due to undersampling the waveform becomes significant. We therefore wish to choose the sparsest sampling strategy that yields negligible errors when interpolating back to a dense set of times (or frequencies).

One strategy would be to use a fixed sampling rate, choosing L uniformly spaced points between x_{min} and x_{max} with

$$\delta x = \frac{1}{L-1} \left(x_{\text{max}} - x_{\text{min}} \right) . \tag{A.6}$$

A second strategy which is likely to yield a lower L for a fixed interpolation error tolerance would be to sample uniformly in the phase of $h(x)$. Since h also depends on $\vec{\lambda}$ and we must choose the same x samples for all $\vec{\lambda}$, the sampling will be only approximately uniform in the phase and the number of samples per cycle will vary between waveforms. To make this strategy explicit, we can choose one *reference waveform* $h_0(x) = h(x; \vec{\lambda}_0)$ where $\vec{\lambda}_0$ might be, for example, mass ratio $q = 1$ and zero spins. We then choose L samples uniformly spaced in

$$\varphi_0(x) = \arg\left[h_0(x) \right] . \tag{A.7}$$

This can be problematic when $|h_0(x)|$ becomes very small and $\arg[h_0(x)]$ becomes poorly resolved, as $|h(x; \vec{\lambda})|$ could still be large for some $\vec{\lambda}$. To overcome this, we choose a constant phase sampling rate $\delta\varphi$ and set minimum and maximum values of δx. We then begin at $x = x_{\text{min}}$ and compute

$$\omega_0(x) = \frac{d}{dx} \varphi_0(x) . \tag{A.8}$$

We then choose

$$\delta x = \min\left(\delta x_{\max}, \max\left(\delta x_{\min}, \frac{\delta\varphi}{\omega_0(x)}\right)\right) \tag{A.9}$$

and advance to $\min(x_{\max}, x + \delta x)$. We continue until we reach $x = x_{\max}$.

Whatever strategy is chosen, we have a method of increasing or decreasing the number of samples L. Given a densely sampled waveform $h(x)$, we can determine the interpolation error by first downsampling h to our L samples, then interpolating the real and imaginary parts of h back to the dense sampling using (for example) cubic splines. We increase L until this interpolation error is negligible for all waveforms in our training set H_1.

A.4 Constructing an empirical interpolant

Next, we construct an empirical interpolant for $h(x)$ using our training set H_1. For details, see Ref. [76] or Sec. 6.3. We begin by constructing a linear basis

$$E = \{e_i(x)\}_{i=1}^{M} \tag{A.10}$$

for $h(x)$ such that the projection errors are all below some tolerance ϵ. Specifically, we choose M large enough such that

$$\sum_{j=1}^{L} |h(x_j; \vec{T}^{-1}(\vec{\mu}_k)) - \sum_{i=1}^{M} c_i(\vec{\mu}_k) e_i(x_j)|^2 < \epsilon^2 \tag{A.11}$$

for all $\vec{\mu}_k \in H_1$ where x_j are our chosen time (or frequency) samples, and

$$c_i(\vec{\mu}) = \sum_{j=1}^{L} h(x_j; \vec{T}^{-1}(\vec{\mu})) e_i(x_j)\,. \tag{A.12}$$

The particular basis can be constructed using a singular value decomposition (SVD) of our whole training set H_1 and keeping the first M vectors, or by constructing a reduced basis (see [248] and Sec. 6.3). For large training sets, a reduced basis is favorable due to the computational difficulty of the SVD method, which requires loading the entire training set into memory.

From the basis E, we then determine M empirical nodes $\{x_i^*\}_{i=1}^{M}$ and perform a linear change of basis to obtain the empirical interpolation basis $B = \{b_i(x)\}_{i=1}^{M}$ through

$$b_i(x) = \sum_{j=1}^{M} \left(V^{-1}\right)_{ji} e_j(t)\,, \tag{A.13}$$

where V is a $M \times M$ matrix with $V_{ji} = e_j(x_i^*)$. Again, see Ref. [76] or Sec. 6.3 for details.

We will now estimate the error of the empirical interpolant and domain sampling strategy over the whole parameter space using the validation set H_2. For each $\vec{\mu} \in G^d_{N_2}$, we take the densely sampled waveform $h(x; \vec{T}^{-1}(\vec{\mu}))$ and downsample it using the sparse sampling stragegy selected in Sec. A.3. We then evaluate the empirical interpolant, obtaining

$$g(x;\vec{\mu}) = \sum_{i=1}^{M} h\left(x_i^*; \vec{T}^{-1}(\vec{\mu})\right) b_i(x)\,, \tag{A.14}$$

and interpolate $g(x;\vec{\mu})$ back to a dense sampling of x using cubic splines. Finally, we compute the error

$$\varepsilon(\vec{\mu}) = \int dx \left| h(x; \vec{T}^{-1}(\vec{\mu})) - g(x;\vec{\mu}) \right|^2 . \tag{A.15}$$

We find the largest error over the validation set

$$\varepsilon^* = \max_{\vec{\mu} \in G^d_{N_2}} \varepsilon(\vec{\mu})\,. \tag{A.16}$$

If ε^* is an acceptibly small error, we continue. Otherwise, we can attempt to reduce the error by increasing the sampling rate with a larger L and/or by using a larger linear basis by choosing a smaller ϵ. If ε^* remains large, then the training set H_1 does not adequately span the full parameter space, and we must increase N_1 and N_2 and start over at Sec. A.2.

A.5 Tensor spline interpolation

We now build cubic tensor-spline interpolants for the real and imaginary parts of the waveform at the empirical nodes, which we denote

$$h^i_R(\vec{\mu}) = h(x_i^*; T^{-1}(\vec{\mu})) \tag{A.17}$$

$$h^i_I(\vec{\mu}) = h(x_i^*; T^{-1}(\vec{\mu}))\,. \tag{A.18}$$

We will consider a single $h^i_{I,R}(\vec{\mu})$, and simply denote it by $y(\vec{\mu})$.

We first give the method for building a cubic tensor-spline interpolant for y given grid data

$$Y_N = \{y(\vec{\mu}) : \vec{\mu} \in G^d_N\}\,. \tag{A.19}$$

We label the grid points of G_N^d with d indices $i_1, \ldots, i_d$ which take on values between 0 and $N - 1$ inclusive, with

$$\vec{\mu}_{i_1,\ldots,i_d} = \left(\frac{i_1}{N-1}, \ldots, \frac{i_d}{N-1}\right) \tag{A.20}$$

and

$$y\left[i_1, \ldots, i_d\right] = y(\vec{\mu}_{i_1,\ldots,i_d} \,. \tag{A.21}$$

The cubic tensor-spline interpolant requires a grid of $(N+2)^d$ coefficients. For a general grid of dimensions $N_1 \times \ldots \times N_d$, we would obtain $(N_1+2) \times \ldots \times (N_d+2)$ spline coefficients. Denoting the coefficients $s[j_1, \ldots, j_d]$, where each j index takes on values between 0 and $N + 1$ inclusive, the cubic tensor-spline interpolant TS is evaluated as

$$TS(\vec{\mu}) = \sum_{k_1=0}^{3} \ldots \sum_{k_d=0}^{3} s\left[j_1^*(\mu_1) + k_1, \ldots, j_d^*(\mu_d)\right] \prod_{\ell=1}^{d} B_{j_\ell^*(\mu_\ell)+k_\ell}(\mu_\ell) \,. \tag{A.22}$$

Here, $B_j(\mu_\ell)$ is a cubic B-spline with knots $u_j^0, \ldots, u_j^4$ given by

$$u_j^i = \max\left(0, \min\left(1, \frac{i+j-3}{N-1}\right)\right) \,, \tag{A.23}$$

and

$$j_\ell^*(\mu_\ell) = \lfloor (N-1)\mu_\ell \rfloor \tag{A.24}$$

identifies the first of 4 (potentially) nonzero B-splines in dimension ℓ. See Ref. [293] for details on B-splines. The cubic tensor-spline evaluation consists of a sum over 4^d terms, each of which is a spline coefficient multiplied with the product of d B-splines depending on $\vec{\mu}$. Since the B-splines are independent of which $h_{R,I}^i$ we are considering, the 4^d B-spline products can be evaluated a single time and stored in memory, to be reused for other empirical nodes.

Determining the spline coefficients involves equating the spline evaluation with the grid data at the N^d grid points, and imposing two boundary conditions per dimension in order to obtain $(N+2)^d$ equations for the $(N+2)^d$ unknown spline coefficients.

We have

$$(\mathcal{B})_{ij} \equiv B_j\left(\frac{i}{N-1}\right) = \begin{cases} 1, & (j=0=i) \text{ or } (j=N+1, i=N-1)\,, \\ \frac{1}{4}, & (j=1=i) \text{ or } (j=N, i=N-2)\,, \\ \frac{7}{12}, & (j=2, i=1) \text{ or } (j=N-1, i=N-2)\,, \\ \frac{1}{6}, & (j=2, i=2) \text{ or } (j=N-1, i=N-3)\,, \\ \frac{1}{6}, & 3 \le j \le N-2, i=j-1\pm 1\,, \\ \frac{2}{3}, & 3 \le j \le N-2, i=j-1\,, \\ 0, & \text{otherwise.} \end{cases} \tag{A.25}$$

We augment $\mathcal{B}$ from a $N \times (N+2)$ matrix to a $(N+2)\times(N+2)$ matrix by prepending one row and appending another, corresponding to the spline boundary conditions at the start and end of the interval. We use 'not-a-knot' boundary conditions, which give a constant derivative across the first and last non-boundary breakpoints. This leads to the extended matrix

$$\mathcal{A} = \begin{bmatrix} 1 & -2 & \frac{3}{2} & -\frac{2}{3} & \frac{1}{6} & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{7}{12} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{7}{12} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \frac{1}{6} & -\frac{2}{3} & \frac{3}{2} & -2 & 1 \end{bmatrix}, \tag{A.26}$$

where for concreteness we fix $N = 9$. For other values of N, the tri-diagonal part of $\mathcal{A}$ with diagonal $\frac{2}{3}$ and sidebands $\frac{2}{3}$ is extended. If we ignore the first and last rows of $\mathcal{A}$, each column of the the $N \times (N+2)$ submatrix $\mathcal{B}$ gives B_j evaluated at the N grid points in one dimension.

To write down the system of $(N+2)^d$ equations for the $(N+2)^d$ unknowns, we need to pad the grid data $y[]$ with zeros in order to obtain not-a-knot boundary conditions. Defining

$$\tilde{y}\left[j_1,\ldots,j_d\right] = \begin{cases} y\left[j_1-1,\ldots,j_d-1\right], & 1 \le j_\ell \le N \text{ for all } \ell = 1,\ldots,d \\ 0 & \text{otherwise}\,, \end{cases} \tag{A.27}$$

the system of equations is

$$\sum_{j_1=0}^{N+1} \cdots \sum_{j_d=0}^{N+1} s\left[j_1,\ldots,j_d\right] \prod_{\ell=1}^{d} \mathcal{A}_{i_\ell j_\ell} = \tilde{y}\left[i_1,\ldots,i_d\right] . \tag{A.28}$$

For grids with different numbers of points in each dimension, the d copies of $\mathcal{A}$ will have different sizes. Inverting this system can be done simply by inverting the matrix (or matrices for non-uniform grids) $\mathcal{A}$, obtaining

$$s\left[j_1,\ldots,j_d\right] = \sum_{i_1=0}^{N+1} \cdots \sum_{i_d=0}^{N+1} \tilde{y}\left[i_1,\ldots,i_d\right] \prod_{\ell=1}^{d} \left(\mathcal{A}^{-1}\right)_{j_l i_l} . \tag{A.29}$$

We first determine the spline coefficients for all empirical nodes $h^i_{R,I}(\vec{\mu})$ using the grid data from the training set H_1. We can then evaluate the entire tensor-spline surrogate for each $\vec{\mu} \in G_2$, and compare that to the corresponding waveform in H_2. Typically, an accurate empirical interpolant can be obtained from a fairly sparse grid. For example, in chapter 7, using $N \sim 5$ gave empirical interpolants with small validation errors. The spline interpolants, however, may require many more points per dimension. We therefore typically need to build tensor-spline interpolants from increasingly large grid sizes N until the validation errors become small. If N and d are large enough that disk space usage becomes an issue, it can be useful to store only the waveform at the empirical nodes. The tensor-spline interpolants can then be constructed, and validation errors can be computed one at a time without ever storing the full waveforms on disk.

Bibliography

1. B. P. Abbott et al. GW150914: The Advanced LIGO Detectors in the Era of First Discoveries. *Phys. Rev. Lett.* 116 (13) (2016), p. 131103. DOI: `10.1103/PhysRevLett.116.131103` (cit. on p. 1).
2. B. P. Abbott et al. The Rate of Binary Black Hole Mergers Inferred from Advanced LIGO Observations Surrounding GW150914. *Accepted for publication in Astrophys. J. Lett.; arXiv:1602.03842* (2016). arXiv: `1602.03842 [astro-ph.HE]` (cit. on p. 2).
3. B. P. Abbott et al. Binary Black Hole Mergers in the first Advanced LIGO Observing Run. *Phys. Rev.* X6 (4) (2016), p. 041015. DOI: `10.1103/PhysRevX.6.041015` (cit. on p. 2).
4. B. P. Abbott et al. Upper limits on the rates of binary neutron star and neutron-star–black-hole mergers from Advanced LIGO's first observing run (2016). arXiv: `1607.07456 [astro-ph.HE]` (cit. on p. 2).
5. B. P. Abbott et al. Search for intermediate mass black hole binaries in the first observing run of Advanced LIGO (2017). arXiv: `1704.04628 [gr-qc]` (cit. on p. 5).
6. M. Armano, H. Audley, G. Auger, J. T. Baird, M. Bassan, P. Binetruy, M. Born, D. Bortoluzzi, N. Brandt, M. Caleno, et al. Sub-femto-*g* free fall for space-based gravitational wave observatories: lisa pathfinder results. *Phys. Rev. Lett.* 116 (23 2016), p. 231101. DOI: `10.1103/PhysRevLett.116.231101` (cit. on p. 6).
7. M. van der Sluys, V. Raymond, I. Mandel, C. Röver, N. Christensen, V. Kalogera, R. Meyer, and A. Vecchio. Parameter estimation of spinning binary inspirals using markov chain monte carlo. *Classical and Quantum Gravity* 25 (18) (2008), p. 184011. URL: `http://stacks.iop.org/0264-9381/25/i=18/a=184011` (cit. on pp. 15, 170).
8. J. Veitch and A. Vecchio. Bayesian coherent analysis of in-spiral gravitational wave signals with a detector network. *Phys. Rev. D* 81 (6 2010), p. 062003. DOI: `10.1103/PhysRevD.81.062003` (cit. on pp. 15, 170).

9. C. Pankow, P. Brady, E. Ochsner, and R. O'Shaughnessy. Novel scheme for rapid parallel parameter estimation of gravitational waves from compact binary coalescences. *Phys. Rev.* D92 (2) (2015), p. 023002. DOI: `10.1103/PhysRevD.92.023002` (cit. on pp. 15, 170).

10. B. P. Abbott et al. Calibration of the Advanced LIGO detectors for the discovery of the binary black-hole merger GW150914. *Phys. Rev.* D95 (6) (2017), p. 062003. DOI: `10.1103/PhysRevD.95.062003` (cit. on p. 16).

11. A. Bohé, L. Shao, A. Taracchini, A. Buonanno, S. Babak, I. W. Harry, I. Hinder, S. Ossokine, M. Pürrer, V. Raymond, et al. Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors. *Phys. Rev. D* 95 (4 2017), p. 044028. DOI: `10.1103/PhysRevD.95.044028` (cit. on p. 18).

12. V. Connaughton et al. Fermi GBM Observations of LIGO Gravitational Wave event GW150914. *Astrophys. J.* 826 (1) (2016), p. L6. DOI: `10.3847/2041-8205/826/1/L6` (cit. on p. 171).

13. B. P. Abbott et al. Localization and broadband follow-up of the gravitational-wave transient GW150914. *Astrophys. J.* 826 (1) (2016), p. L13. DOI: `10.3847/2041-8205/826/1/L13` (cit. on p. 171).

14. J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith. A surrogate model of gravitational waveforms from numerical relativity simulations of precessing binary black hole mergers. *Phys. Rev. D* 95 (10 2017), p. 104023. DOI: `10.1103/PhysRevD.95.104023`. (Cit. on pp. vi, 73, 150–152, 155, 156, 158, 160, 171).

15. J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, C. D. Ott, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and B. Szilágyi. A Numerical Relativity Waveform Surrogate Model for Generically Precessing Binary Black Hole Mergers (2017). Submitted to Physical Review D. arXiv: `1705.07089 [gr-qc]`. (Cit. on pp. vi, 148).

16. B. Szilágyi, J. Blackman, A. Buonanno, A. Taracchini, H. P. Pfeiffer, M. A. Scheel, T. Chu, L. E. Kidder, and Y. Pan. Approaching the post-newtonian regime with numerical relativity: a compact-object binary simulation spanning 350 gravitational-wave cycles. *Phys. Rev. Lett.* 115 (3 2015), p. 031102. DOI: `10.1103/PhysRevLett.115.031102`. (Cit. on pp. 3, 11, 18, 46).

17. R. O'Shaughnessy, J. Blackman, and S. E. Field. An architecture for efficient multimodal gravitational wave parameter estimation with linear surrogate models (2017). arXiv: `1701.01137 [gr-qc]`. (Cit. on pp. 169, 174).

18. J. M. Fedrow, C. D. Ott, U. Sperhake, J. Blackman, R. Haas, C. Reisswig, and A. De Felice. Gravitational Waves from Binary Black Hole Mergers Inside of Stars (2017). arXiv: `1704.07383 [astro-ph.HE]`. (Cit. on p. 171).

19. F. Pretorius. Evolution of binary black hole spacetimes. *Phys. Rev. Lett.* 95 (2005), p. 121101. DOI: `10.1103/PhysRevLett.95.121101` (cit. on pp. v, 2).

20. B. P. Abbott et al. Properties of the binary black hole merger GW150914. *Phys. Rev. Lett.* 116 (2016), p. 241102. DOI: `10.1103/PhysRevLett.116.241102` (cit. on pp. v, 2, 15, 75, 91, 125, 150, 169).

21. C. Misner, K. Thorne, and J. Wheeler. *Gravitation*. W.H. Freeman and Company, 1973, pp. 1–1279 (cit. on pp. 1, 4, 8, 9).

22. E. Berti, V. Cardoso, and A. O. Starinets. Quasinormal modes of black holes and black branes. *Class. Quant. Grav.* 26 (2009), p. 163001. DOI: `10.1088/0264-9381/26/16/163001` (cit. on p. 1).

23. J. M. Weisberg, J. H. Taylor, and L. A. Fowler. Gravitational waves from an orbiting pulsar. *Scientific American* 245 (1981), pp. 74–82. DOI: `10.1038/scientificamerican1081-74` (cit. on p. 1).

24. J. M. Weisberg and J. H. Taylor. Relativistic binary pulsar B1913+16: Thirty years of observations and analysis. *ASP Conf. Ser.* 328 (2005), p. 25. arXiv: `astro-ph/0407149 [astro-ph]` (cit. on pp. 1, 6).

25. J. Antoniadis et al. A Massive Pulsar in a Compact Relativistic Binary. *Science* 340 (2013), p. 6131. DOI: `10.1126/science.1233232` (cit. on pp. 1, 6).

26. F. Acernese et al. Status of VIRGO. *Class. Quant. Grav.* 22 (2005), S869–S880. DOI: `10.1088/0264-9381/22/18/S01` (cit. on pp. 1, 5).

27. A. Abramovici, W. E. Althouse, R. W. P. Drever, Y. Gürsel, S. Kawamura, F. J. Raab, D. Shoemaker, L. Sievers, R. E. Spero, K. S. Thorne, et al. Ligo: the laser interferometer gravitational-wave observatory. *Science* 256 (5055) (1992), pp. 325–333. DOI: `10.1126/science.256.5055.325` (cit. on pp. 1, 5).

28. B. Abbott et al. Detector description and performance for the first coincidence observations between LIGO and GEO. *Nucl. Instrum. Meth.* A517 (2004), pp. 154–179. DOI: `10.1016/j.nima.2003.11.124` (cit. on pp. 1, 5).

29. B. Willke et al. The geo 600 gravitational wave detector. *Classical and Quantum Gravity* 19 (7) (2002), p. 1377. URL: `http://stacks.iop.org/0264-9381/19/i=7/a=321` (cit. on pp. 1, 5).

30. K. L. Dooley. Status of GEO 600. *J. Phys. Conf. Ser.* 610 (1) (2015), p. 012015. DOI: `10.1088/1742-6596/610/1/012015` (cit. on pp. 1, 5).

31. G. M. Harry. Advanced LIGO: The next generation of gravitational wave detectors. *Class.Quant.Grav.* 27 (2010), p. 084006. DOI: `10.1088/0264-9381/27/8/084006` (cit. on pp. 1, 5, 49).

32. J. Aasi et al. Advanced LIGO. *Class. Quant. Grav.* 32 (2015), p. 074001. DOI: `10.1088/0264-9381/32/7/074001` (cit. on pp. 1, 5, 49, 75).

33. The Virgo Collaboration. *Advanced Virgo Baseline Design*. [VIR-0027A-09]. 2009. URL: `https://tds.ego-gw.it/ql/?c=6589` (cit. on pp. 1, 5).

34. The Virgo Collaboration. *Advanced Virgo Technical Design Report*. [VIR-0128A-12]. 2012. URL: `https://tds.ego-gw.it/ql/?c=6940` (cit. on pp. 1, 5).

35. K. Somiya. Detector configuration of KAGRA: The Japanese cryogenic gravitational-wave detector. *Class. Quant. Grav.* 29 (2012), p. 124007. DOI: `10.1088/0264-9381/29/12/124007` (cit. on pp. 1, 5, 35, 49).

36. G. M. Harry. Advanced LIGO: The next generation of gravitational wave detectors. 27 (2010), p. 084006. DOI: `10.1088/0264-9381/27/8/084006` (cit. on pp. 1, 35).

37. C. Brans and R. H. Dicke. Mach's principle and a relativistic theory of gravitation. *Phys. Rev.* 124 (3 1961), pp. 925–935. DOI: `10.1103/PhysRev.124.925` (cit. on p. 2).

38. T. Damour and G. Esposito-Farèse. Nonperturbative strong-field effects in tensor-scalar theories of gravitation. *Phys. Rev. Lett.* 70 (15 1993), pp. 2220–2223. DOI: `10.1103/PhysRevLett.70.2220` (cit. on p. 2).

39. T. Damour and G. Esposito-Farèse. Gravitational-wave versus binary-pulsar tests of strong-field gravity. *Phys. Rev. D* 58 (4 1998), p. 042001. DOI: `10.1103/PhysRevD.58.042001` (cit. on p. 2).

40. K. G. ARUN and A. PAI. Tests of general relativity and alternative theories of gravity using gravitational wave observations. *International Journal of Modern Physics D* 22 (01) (2013), p. 1341012. DOI: `10.1142/S0218271813410125` (cit. on p. 2).

41. J. Aasi et al. Parameter estimation for compact binary coalescence signals with the first generation gravitational-wave detector network. *Phys. Rev. D* 88 (6 2013), p. 062001. DOI: `10.1103/PhysRevD.88.062001` (cit. on p. 2).

42. V. Raymond, M. V. van der Sluys, I. Mandel, V. Kalogera, C. Röver, and N. Christensen. Degeneracies in sky localization determination from a spinning coalescing binary through gravitational wave observations: a markov-chain monte carlo analysis for two detectors. *Classical and Quantum Gravity* 26 (11) (2009), p. 114007. URL: `http://stacks.iop.org/0264-9381/26/i=11/a=114007` (cit. on p. 2).

43. M. van der Sluys, I. Mandel, V. Raymond, V. Kalogera, C. Röver, and N. Christensen. Parameter estimation for signals from compact binary inspirals injected into ligo data. *Classical and Quantum Gravity* 26 (20) (2009), p. 204010. URL: `http://stacks.iop.org/0264-9381/26/i=20/a=204010` (cit. on p. 2).

44. R. J. E. Smith, K. Cannon, C. Hanna, D. Keppel, and I. Mandel. Towards rapid parameter estimation on gravitational waves from compact binaries using interpolated waveforms. *Phys. Rev. D* 87 (12 2013), p. 122002. DOI: `10.1103/PhysRevD.87.122002` (cit. on p. 2).

45. R. Fender and T. Belloni. Stellar-Mass Black Holes and Ultraluminous X-ray Sources. *Science* 337 (2012), pp. 540–. DOI: `10.1126/science.1221790` (cit. on pp. 2, 4).

46. J. Ziolkowski. Population of Galactic black holes. *MmSAI* 81 (2010), p. 294 (cit. on p. 2).

47. W. M. Farr, N. Sravan, A. Cantrell, L. Kreidberg, C. D. Bailyn, I. Mandel, and V. Kalogera. The Mass Distribution of Stellar-mass Black Holes. *Astrophysical Journal* 741, 103 (2011), p. 103. DOI: `10.1088/0004-637X/741/2/103` (cit. on pp. 2, 4).

48. F. Özel, D. Psaltis, R. Narayan, and J. E. McClintock. The Black Hole Mass Distribution in the Galaxy. *Astrophysical Journal* 725 (2010), pp. 1918–1927. DOI: `10.1088/0004-637X/725/2/1918` (cit. on p. 2).

49. L. Kreidberg, C. D. Bailyn, W. M. Farr, and V. Kalogera. Mass measurements of black holes in x-ray transients: is there a mass gap? *The Astrophysical Journal* 757 (1) (2012), p. 36. URL: `http://stacks.iop.org/0004-637X/757/i=1/a=36` (cit. on p. 2).

50. L. Lehner. Numerical relativity: a review. *Classical and Quantum Gravity* 18 (17) (2001), R25. URL: `http://stacks.iop.org/0264-9381/18/i=17/a=202` (cit. on p. 2).

51. F. Pretorius. Binary Black Hole Coalescence (2007). arXiv: `0710.1338 [gr-qc]` (cit. on p. 2).

52. J. Centrella, J. G. Baker, B. J. Kelly, and J. R. van Meter. Black-hole binaries, gravitational waves, and numerical relativity. *Rev. Mod. Phys.* 82 (4 2010), pp. 3069–3119. DOI: `10.1103/RevModPhys.82.3069` (cit. on p. 2).

53. U. Sperhake. The numerical relativity breakthrough for binary black holes. *Classical and Quantum Gravity* 32 (12) (2015), p. 124011. URL: `http://stacks.iop.org/0264-9381/32/i=12/a=124011` (cit. on p. 2).

54. M. A. Scheel, M. Boyle, T. Chu, L. E. Kidder, K. D. Matthews, and H. P. Pfeiffer. High-accuracy waveforms for binary black hole inspiral, merger, and ringdown. *Phys. Rev.* D79 (2009), p. 024003. DOI: `10.1103/PhysRevD.79.024003` (cit. on p. 2).

55. G. Lovelace, M. Boyle, M. A. Scheel, and B. Szilágyi. Accurate gravitational waveforms for binary-black-hole mergers with nearly extremal spins. *Class. Quant. Grav.* 29 (2012), p. 045003. DOI: `10.1088/0264-9381/29/4/045003` (cit. on pp. 2, 11).

56. L. T. Buchman, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi. Simulations of non-equal mass black hole binaries with spectral methods. *Phys. Rev.* D86 (2012), p. 084033. DOI: `10.1103/PhysRevD.86.084033` (cit. on pp. 2, 49, 51).

57. S. Ossokine, M. Boyle, L. E. Kidder, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi. Comparing Post-Newtonian and Numerical-Relativity Precession Dynamics (2015). arXiv: `1502.01747 [gr-qc]` (cit. on pp. 2, 96, 108, 151, 154).

58. *Simulating eXtreme Spacetimes Public Waveform Catalog*. URL: `http://www.black-holes.org/waveforms/` (cit. on pp. 2, 62, 64, 66).

59. C. W. Helstrom. *Statistical Theory of Signal Detection: International Series of Monographs in Electronics and Instrumentation*. Vol. 9. Elsevier, 2013 (cit. on pp. 2, 14).

60. B. J. Owen and B. S. Sathyaprakash. Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement. *Phys. Rev.* D60 (1999), p. 022002. DOI: `10.1103/PhysRevD.60.022002` (cit. on pp. 2, 14, 44).

61. L. Blanchet. Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries. *Living Reviews in Relativity* 9 (2006), p. 4. DOI: `10.12942/lrr-2006-4` (cit. on pp. 3, 17).

62. E. Poisson and C. Will. *Gravity: Newtonian, Post-Newtonian, Relativistic*. Cambridge University Press, 2014. ISBN: 9781107032866. URL: `https://books.google.com/books?id=PZ5cAwAAQBAJ` (cit. on pp. 3, 9, 17).

63. P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, J. T. Whelan, B. Brügmann, P. Diener, J. Gonzalez, M. Hannam, et al. A phenomenological template family for black-hole coalescence waveforms. *Classical and Quantum Gravity* 24 (19) (2007), S689. URL: `http://stacks.iop.org/0264-9381/24/i=19/a=S31` (cit. on pp. 3, 19).

64. P. Ajith et al. Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins. *Phys. Rev. Lett.* 106 (2011), p. 241101. DOI: `10.1103/PhysRevLett.106.241101` (cit. on pp. 3, 19).

65. S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé. Frequency-domain gravitational waves from non-precessing black-hole binaries. II. A phenomenological model for the advanced detector era (2015). arXiv: `1508.07253 [gr-qc]` (cit. on pp. 3, 19, 75, 85, 118, 150).

66. M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer. Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms. *Phys. Rev. Lett.* 113 (15) (2014), p. 151101. DOI: `10.1103/PhysRevLett.113.151101` (cit. on pp. v, 3, 19, 37, 62, 75, 85, 118, 125, 150, 160).

67. A. Buonanno and T. Damour. Effective one-body approach to general relativistic two-body dynamics. *Phys. Rev.* D59 (1999), p. 084006. DOI: `10.1103/PhysRevD.59.084006` (cit. on pp. 3, 18, 44).

68. Y. Pan, A. Buonanno, M. Boyle, L. T. Buchman, L. E. Kidder, H. P. Pfeiffer, and M. A. Scheel. Inspiral-merger-ringdown multipolar waveforms of nonspinning black-hole binaries using the effective-one-body formalism. *Phys. Rev.* D84 (2011), p. 124052. DOI: `10.1103/PhysRevD.84.124052` (cit. on pp. 3, 18, 56, 57, 62, 64, 72).

69. A. Taracchini, Y. Pan, A. Buonanno, E. Barausse, M. Boyle, T. Chu, G. Lovelace, H. P. Pfeiffer, and M. A. Scheel. Prototype effective-one-body model for nonprecessing spinning inspiral-merger-ringdown waveforms. *Phys. Rev.* D86 (2012), p. 024011. DOI: `10.1103/PhysRevD.86.024011` (cit. on pp. 3, 18, 44).

70. A. Taracchini et al. Effective-one-body model for black-hole binaries with generic mass ratios and spins. *Phys. Rev.* D89 (6) (2014), p. 061502. DOI: `10.1103/PhysRevD.89.061502` (cit. on pp. v, 3, 18, 49, 56–58, 75, 85, 150).

71. T. B. Littenberg, J. G. Baker, A. Buonanno, and B. J. Kelly. Systematic biases in parameter estimation of binary black-hole mergers. *Phys. Rev. D* 87 (10 2013), p. 104003. DOI: `10.1103/PhysRevD.87.104003` (cit. on pp. 3, 17).

72. P. Kumar, K. Barkett, S. Bhagwat, N. Afshari, D. A. Brown, G. Lovelace, M. A. Scheel, and B. Szilágyi. Accuracy and precision of gravitational-wave models of inspiraling neutron star – black hole binaries with spin: comparison with numerical relativity in the low-frequency regime (2015). arXiv: `1507.00103 [gr-qc]` (cit. on pp. 3, 18, 75).

73. C. Prud'homme, D. V. Rovas, K. Veroy, L. Machiels, Y. Maday, A. T. Patera, and G. Turinici. Reliable real-time solution of parametrized partial differential equations: reduced-basis output bound methods. *Journal of Fluids Engineering* 124 (1) (2002), pp. 70–80 (cit. on pp. 3, 20).

74. M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera. An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique* 339 (9) (2004), pp. 667–672 (cit. on pp. 3, 20).

75. S. E. Field, C. R. Galley, F. Herrmann, J. S. Hesthaven, E. Ochsner, and M. Tiglio. Reduced basis catalogs for gravitational wave templates. *Phys. Rev. Lett.* 106 (2011), p. 221102. DOI: `10.1103/PhysRevLett.106.221102` (cit. on pp. 3, 20, 36, 38, 64, 81, 137).

76. S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio. Fast prediction and evaluation of gravitational waveforms using surrogate models. *Phys. Rev.* X4 (3) (2014), p. 031006. DOI: `10.1103/PhysRevX.4.031006`

(cit. on pp. vi, 3, 20, 25, 32, 36, 44, 45, 62–64, 68, 76, 78, 79, 81, 82, 84, 86, 87, 112, 150, 177, 178).

77. S. Hawking and W. Israel. *Three Hundred Years of Gravitation*. Philosophiae Naturalis, Principia Mathematica. Cambridge University Press, 1989. ISBN: 9780521379762. URL: `https://books.google.com/books?id=Vq787qC5PWQC` (cit. on p. 4).

78. A. Watts et al. Understanding the Neutron Star Population with the SKA. *PoS* AASKA14 (2015), p. 039. arXiv: `1501.00005 [astro-ph.HE]` (cit. on p. 4).

79. A. G. Lyne, M. Burgay, M. Kramer, A. Possenti, R. N. Manchester, F. Camilo, M. A. McLaughlin, D. R. Lorimer, N. D'Amico, B. C. Joshi, et al. A Double-Pulsar System: A Rare Laboratory for Relativistic Gravity and Plasma Physics. *Science* 303 (2004), pp. 1153–1157. DOI: `10.1126/science.1094645` (cit. on p. 4).

80. B. Kiziltan, A. Kottas, M. De Yoreo, and S. E. Thorsett. The Neutron Star Mass Distribution. *Astrophysical Journal* 778, 66 (2013), p. 66. DOI: `10.1088/0004-637X/778/1/66` (cit. on p. 4).

81. D. C. Backer, S. R. Kulkarni, C. Heiles, M. M. Davis, and W. M. Goss. A millisecond pulsar. *Nature* 300 (1982), pp. 615–618. DOI: `10.1038/300615a0` (cit. on p. 4).

82. J. W. T. Hessels, S. M. Ransom, I. H. Stairs, P. C. C. Freire, V. M. Kaspi, and F. Camilo. A radio pulsar spinning at 716 hz. *Science* 311 (5769) (2006), pp. 1901–1904. DOI: `10.1126/science.1123430` (cit. on p. 4).

83. D. P. Bennett et al. Gravitational microlensing events due to stellar-mass black holes. *The Astrophysical Journal* 579 (2) (2002), p. 639. URL: `http://stacks.iop.org/0004-637X/579/i=2/a=639` (cit. on p. 4).

84. B. C. Kelly and A. Merloni. Mass Functions of Supermassive Black Holes across Cosmic Time. *Advances in Astronomy* 2012, 970858 (2012), p. 7. DOI: `10.1155/2012/970858` (cit. on p. 4).

85. F. K. Liu, S. Li, and S. Komossa. A milliparsec supermassive black hole binary candidate in the galaxy SDSS J120136.02+300305.5. *Astrophys. J.* 786 (2014), p. 103. DOI: `10.1088/0004-637X/786/2/103` (cit. on p. 4).

86. M. C. Miller and E. J. M. Colbert. Intermediate - mass black holes. *Int. J. Mod. Phys.* D13 (2004), pp. 1–64. DOI: `10.1142/S0218271804004426` (cit. on p. 4).

87. C. D. Ott. The Gravitational Wave Signature of Core-Collapse Supernovae. *Class. Quant. Grav.* 26 (2009), p. 063001. DOI: `10.1088/0264-9381/26/6/063001` (cit. on p. 5).

88. M. Punturo et al. The einstein telescope: a third-generation gravitational wave observatory. *Classical and Quantum Gravity* 27 (19) (2010), p. 194002. URL: `http://stacks.iop.org/0264-9381/27/i=19/a=194002` (cit. on p. 5).

89. LIGO Scientific Collaboration. *Instrument Science White Paper*. `https://dcc.ligo.org/LIGO-T1600119/public`. 2016 (cit. on p. 5).

90. M. Agathos, J. Meidam, W. Del Pozzo, T. G. F. Li, M. Tompitak, J. Veitch, S. Vitale, and C. V. D. Broeck. Constraining the neutron star equation of state with gravitational wave signals from coalescing binary neutron stars. *Phys. Rev.* D92 (2) (2015), p. 023012. DOI: `10.1103/PhysRevD.92.023012` (cit. on p. 6).

91. K. Danzmann and the LISA study team. Lisa: laser interferometer space antenna for gravitational wave measurements. *Classical and Quantum Gravity* 13 (11A) (1996), A247. URL: `http://stacks.iop.org/0264-9381/13/i=11A/a=033` (cit. on p. 6).

92. D. A. Shaddock. Space-based gravitational wave detection with lisa. *Classical and Quantum Gravity* 25 (11) (2008), p. 114012. URL: `http://stacks.iop.org/0264-9381/25/i=11/a=114012` (cit. on p. 6).

93. G. Hobbs et al. The international pulsar timing array project: using pulsars as a gravitational wave detector. *Classical and Quantum Gravity* 27 (8) (2010), p. 084013. URL: `http://stacks.iop.org/0264-9381/27/i=8/a=084013` (cit. on p. 6).

94. D. R. Yardley, G. B. Hobbs, F. A. Jenet, J. P. Verbiest, Z. L. Wen, R. N. Manchester, W. A. Coles, W. van Straten, M. Bailes, N. D. Bhat, et al. The sensitivity of the Parkes Pulsar Timing Array to individual sources of gravitational waves. *Memoirs of the RAS* 407 (2010), pp. 669–680. DOI: `10.1111/j.1365-2966.2010.16949.x` (cit. on p. 6).

95. C. M. Will. The confrontation between general relativity and experiment. *Living Reviews in Relativity* 9 (1) (2006), p. 3. ISSN: 1433-8351. DOI: `10.12942/lrr-2006-3` (cit. on p. 6).

96. D. Psaltis. Probes and tests of strong-field gravity with observations in the electromagnetic spectrum. *Living Reviews in Relativity* 11 (1) (2008), p. 9. ISSN: 1433-8351. DOI: `10.12942/lrr-2008-9` (cit. on p. 6).

97. N. Yunes and X. Siemens. Gravitational-Wave Tests of General Relativity with Ground-Based Detectors and Pulsar Timing-Arrays. *Living Rev. Rel.* 16 (2013), p. 9. DOI: `10.12942/lrr-2013-9` (cit. on p. 6).

98. K. Yagi and L. C. Stein. Black Hole Based Tests of General Relativity. *Class. Quant. Grav.* 33 (5) (2016), p. 054001. DOI: `10.1088/0264-9381/33/5/054001` (cit. on p. 6).

99. C. Brans and R. H. Dicke. Mach's principle and a relativistic theory of gravitation. *Phys. Rev.* 124 (3 1961), pp. 925–935. DOI: `10.1103/PhysRev.124.925` (cit. on p. 6).

100. Y. Fujii and K. Maeda. *The Scalar-Tensor Theory of Gravitation*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2003. ISBN: 9781139436021. URL: `https://books.google.com/books?id=cRim42Z2hBsC` (cit. on p. 6).

101. T. P. Sotiriou and V. Faraoni. Black holes in scalar-tensor gravity. *Phys. Rev. Lett.* 108 (2012), p. 081103. DOI: `10.1103/PhysRevLett.108.081103` (cit. on p. 6).

102. C. M. Will. Testing scalar - tensor gravity with gravitational wave observations of inspiraling compact binaries. *Phys. Rev.* D50 (1994), pp. 6058–6067. DOI: `10.1103/PhysRevD.50.6058` (cit. on p. 7).

103. K. Hinterbichler. Theoretical Aspects of Massive Gravity. *Rev. Mod. Phys.* 84 (2012), pp. 671–710. DOI: `10.1103/RevModPhys.84.671` (cit. on p. 7).

104. C. M. Will. Bounding the mass of the graviton using gravitational wave observations of inspiralling compact binaries. *Phys. Rev.* D57 (1998), pp. 2061–2068. DOI: `10.1103/PhysRevD.57.2061` (cit. on p. 7).

105. M. Bojowald and G. M. Hossain. Loop quantum gravity corrections to gravitational wave dispersion. *Phys. Rev.* D77 (2008), p. 023508. DOI: `10.1103/PhysRevD.77.023508` (cit. on p. 7).

106. S. Alexander and N. Yunes. Chern-Simons Modified General Relativity. *Phys. Rept.* 480 (2009), pp. 1–55. DOI: `10.1016/j.physrep.2009.07.002` (cit. on p. 7).

107. P. Pani and V. Cardoso. Are black holes in alternative theories serious astrophysical candidates? The Case for Einstein-Dilaton-Gauss-Bonnet black holes. *Phys. Rev.* D79 (2009), p. 084031. DOI: `10.1103/PhysRevD.79.084031` (cit. on p. 7).

108. L. Amendola, C. Charmousis, and S. C. Davis. Solar System Constraints on Gauss-Bonnet Mediated Dark Energy. *JCAP* 0710 (2007), p. 004. DOI: `10.1088/1475-7516/2007/10/004` (cit. on p. 7).

109. P. Pani, E. Berti, V. Cardoso, and J. Read. Compact stars in alternative theories of gravity. Einstein-Dilaton-Gauss-Bonnet gravity. *Phys. Rev.* D84 (2011), p. 104035. DOI: `10.1103/PhysRevD.84.104035` (cit. on p. 7).

110. K. Yagi. A New constraint on scalar Gauss-Bonnet gravity and a possible explanation for the excess of the orbital decay rate in a low-mass X-ray binary. *Phys. Rev.* D86 (2012), p. 081504. DOI: `10.1103/PhysRevD.86.081504` (cit. on p. 7).

111. Y. Ali-Haimoud and Y. Chen. Slowly-rotating stars and black holes in dynamical Chern-Simons gravity. *Phys. Rev.* D84 (2011), p. 124033. DOI: `10.1103/PhysRevD.84.124033` (cit. on p. 7).

112. C. K. Mishra, K. G. Arun, B. R. Iyer, and B. S. Sathyaprakash. Parametrized tests of post-Newtonian theory using Advanced LIGO and Einstein Telescope. *Phys. Rev.* D82 (2010), p. 064010. DOI: `10.1103/PhysRevD.82.064010` (cit. on p. 7).

113. M. Agathos, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, J. Veitch, and S. Vitale. TIGER: A data analysis pipeline for testing the strong-field dynamics of general relativity with gravitational wave signals from coalescing compact binaries. *Phys. Rev.* D89 (8) (2014), p. 082001. DOI: `10.1103/PhysRevD.89.082001` (cit. on p. 7).

114. É. É. Flanagan and T. Hinderer. Constraining neutron-star tidal love numbers with gravitational-wave detectors. *Phys. Rev. D* 77 (2 2008), p. 021502. DOI: `10.1103/PhysRevD.77.021502` (cit. on p. 9).

115. J. A. Faber and F. A. Rasio. Binary Neutron Star Mergers. *Living Rev. Rel.* 15 (2012), p. 8. DOI: `10.12942/lrr-2012-8` (cit. on p. 9).

116. L. Gou, J. E. McClintock, R. A. Remillard, J. F. Steiner, M. J. Reid, J. A. Orosz, R. Narayan, M. Hanke, and J. García. Confirmation via the continuum-fitting method that the spin of the black hole in cygnus x-1 is extreme. *The Astrophysical Journal* 790 (1) (2014), p. 29. URL: `http://stacks.iop.org/0004-637X/790/i=1/a=29` (cit. on p. 10).

117. J. E. McClintock, R. Narayan, and J. F. Steiner. Black Hole Spin via Continuum Fitting and the Role of Spin in Powering Transient Jets. *Space Sci. Rev.* 183 (2014), pp. 295–322. DOI: `10.1007/s11214-013-0003-9` (cit. on p. 10).

118. C. S. Reynolds. Measuring Black Hole Spin using X-ray Reflection Spectroscopy. *Space Sci. Rev.* 183 (1-4) (2014), pp. 277–294. DOI: `10.1007/s11214-013-0006-6` (cit. on p. 10).

119. G. W. Gibbons. Vacuum polarization and the spontaneous loss of charge by black holes. *Comm. Math. Phys.* 44 (3) (1975), pp. 245–264. URL: `http://projecteuclid.org/euclid.cmp/1103899346` (cit. on p. 10).

120. P. C. Peters. Gravitational Radiation and the Motion of Two Point Masses. *Physical Review* 136 (1964), pp. 1224–1232. DOI: `10.1103/PhysRev.136.B1224` (cit. on p. 10).

121. L. Wen. On the eccentricity distribution of coalescing black hole binaries driven by the Kozai mechanism in globular clusters. *Astrophys. J.* 598 (2003), pp. 419–430. DOI: `10.1086/378794` (cit. on p. 10).

122. K. S. Thorne. Multipole expansions of gravitational radiation. *Rev. Mod. Phys.* 52 (2 1980), pp. 299–339. DOI: `10.1103/RevModPhys.52.299` (cit. on p. 11).

123. *Spectral Einstein Code*. URL: `http://www.black-holes.org/SpEC.html` (cit. on p. 11).

124. M. Boyle, L. Lindblom, H. P. Pfeiffer, M. A. Scheel, and L. E. Kidder. Testing the accuracy and stability of spectral methods in numerical relativity. *Phys. Rev. D* 75 (2 2007), p. 024006. DOI: `10.1103/PhysRevD.75.024006` (cit. on p. 11).

125. B. Szilágyi, L. Lindblom, and M. A. Scheel. Simulations of binary black hole mergers using spectral methods. *Phys. Rev. D* 80 (12 2009), p. 124010. DOI: `10.1103/PhysRevD.80.124010` (cit. on p. 11).

126. M. A. Scheel, H. P. Pfeiffer, L. Lindblom, L. E. Kidder, O. Rinne, and S. A. Teukolsky. Solving einstein's equations with dual coordinate frames. *Phys. Rev. D* 74 (10 2006), p. 104006. DOI: `10.1103/PhysRevD.74.104006` (cit. on p. 11).

127. L. Lindblom, M. A. Scheel, L. E. Kidder, R. Owen, and O. Rinne. A new generalized harmonic evolution system. *Classical and Quantum Gravity* 23 (16) (2006), S447. URL: `http://stacks.iop.org/0264-9381/23/i=16/a=S09` (cit. on p. 11).

128. L. Lindblom, M. A. Scheel, L. E. Kidder, H. P. Pfeiffer, D. Shoemaker, and S. A. Teukolsky. Controlling the growth of constraints in hyperbolic evolution systems. *Phys. Rev. D* 69 (12 2004), p. 124025. DOI: `10.1103/PhysRevD.69.124025` (cit. on p. 12).

129. M. Holst, L. Lindblom, R. Owen, H. P. Pfeiffer, M. A. Scheel, and L. E. Kidder. Optimal constraint projection for hyperbolic evolution systems. *Phys. Rev. D* 70 (8 2004), p. 084017. DOI: `10.1103/PhysRevD.70.084017` (cit. on p. 12).

130. S. Dain, C. O. Lousto, and R. Takahashi. New conformally flat initial data for spinning black holes. *Phys. Rev. D* 65 (10 2002), p. 104038. DOI: `10.1103/PhysRevD.65.104038` (cit. on p. 12).

131. G. Lovelace, R. Owen, H. P. Pfeiffer, and T. Chu. Binary-black-hole initial data with nearly extremal spins. *Phys. Rev. D* 78 (8 2008), p. 084017. DOI: `10.1103/PhysRevD.78.084017` (cit. on p. 12).

132. A. Buonanno, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, and A. Taracchini. Reducing orbital eccentricity of precessing black-hole binaries. 83 (2011), p. 104034. DOI: `10.1103/PhysRevD.83.104034` (cit. on pp. 12, 64).

133. H. P. Pfeiffer, D. A. Brown, L. E. Kidder, L. Lindblom, G. Lovelace, and M. A. Scheel. Reducing orbital eccentricity in binary black hole simulations. 24 (12) (2007), S59–S81. eprint: `gr-qc/0702106` (cit. on pp. 12, 64).

134. M. Boyle and A. H. Mroué. Extrapolating gravitational-wave data from numerical simulations. *Phys. Rev. D* 80 (12 2009), p. 124045. DOI: `10.1103/PhysRevD.80.124045` (cit. on p. 12).

135. M. Boyle. Accurate gravitational waveforms from binary black-hole systems. *Ph.D. thesis, California Institute of Technology* (2008) (cit. on p. 12).

136. N. Bishop, R. Gomez, L. Lehner, and J. Winicour. Cauchy-characteristic extraction in numerical relativity. *ArXiv General Relativity and Quantum Cosmology e-prints* (1997). eprint: `gr-qc/9705033` (cit. on p. 12).

137. C. J. Handmer, B. Szilágyi, and J. Winicour. Gauge Invariant Spectral Cauchy Characteristic Extraction (2015). arXiv: `1502.06987 [gr-qc]` (cit. on p. 12).

138. C. J. Handmer, B. Szilágyi, and J. Winicour. Spectral cauchy characteristic extraction of strain, news and gravitational radiation flux. *Classical and Quantum Gravity* 33 (22) (2016), p. 225007. URL: `http://stacks.iop.org/0264-9381/33/i=22/a=225007` (cit. on pp. 12, 131).

139. T. M. Adamo, E. T. Newman, and C. Kozameh. Null geodesic congruences, asymptotically-flat spacetimes and their physical interpretation. *Living Reviews in Relativity* 15 (1) (2012). DOI: `10.12942/lrr-2012-1` (cit. on p. 13).

140. M. Boyle. Transformations of asymptotic gravitational-wave data (2015). arXiv: `1509.00862 [gr-qc]` (cit. on pp. 13, 151, 153).

141. B. Sathyaprakash and B. F. Schutz. Physics, astrophysics and cosmology with gravitational waves. *Living Reviews in Relativity* 12 (2) (2009). DOI: `10.12942/lrr-2009-2` (cit. on p. 13).

142. L. Blackburn et al. The lsc glitch group: monitoring noise transients during the fifth ligo science run. *Classical and Quantum Gravity* 25 (18) (2008), p. 184004. URL: `http://stacks.iop.org/0264-9381/25/i=18/a=184004` (cit. on p. 14).

143. N. J. Cornish and T. B. Littenberg. BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches. *Class. Quant. Grav.* 32 (13) (2015), p. 135012. DOI: `10.1088/0264-9381/32/13/135012` (cit. on p. 14).

144. P. Canizares, S. E. Field, J. R. Gair, and M. Tiglio. Gravitational wave parameter estimation with compressed likelihood evaluations. 87 (12) (2013), p. 124005 (cit. on pp. 15, 45).

145. R. Smith, S. E. Field, K. Blackburn, C.-J. Haster, M. Pürrer, V. Raymond, and P. Schmidt. Fast and accurate inference on gravitational waves from precessing compact binaries. *Phys. Rev. D* 94 (4 2016), p. 044031. DOI: `10.1103/PhysRevD.94.044031` (cit. on pp. 15, 174).

146. H.-Y. Chen and D. E. Holz. The Loudest Gravitational Wave Events (2014). arXiv: `1409.0522 [gr-qc]` (cit. on p. 16).

147. L. Lindblom, B. J. Owen, and D. A. Brown. Model Waveform Accuracy Standards for Gravitational Wave Data Analysis. 78 (2008), p. 124020. DOI: `10.1103/PhysRevD.78.124020` (cit. on pp. 16, 122, 144, 150).

148. G. Lovelace et al. Modeling the source of GW150914 with targeted numerical-relativity simulations. *Class. Quant. Grav.* 33 (24) (2016), p. 244002. DOI: `10.1088/0264-9381/33/24/244002` (cit. on p. 17).

149. T. Damour, B. R. Iyer, and B. S. Sathyaprakash. Improved filters for gravitational waves from inspiraling compact binaries. *Physical Review D* 57 (2) (1998), p. 885 (cit. on p. 17).

150. A. H. Nitz, A. Lundgren, D. A. Brown, E. Ochsner, D. Keppel, and I. W. Harry. Accuracy of gravitational waveform models for observing neutron-star-black-hole binaries in Advanced LIGO. *Phys. Rev.* D88 (12) (2013), p. 124039. DOI: `10.1103/PhysRevD.88.124039` (cit. on pp. 18, 49).

151. M. Boyle, D. A. Brown, L. E. Kidder, A. H. Mroue, H. P. Pfeiffer, M. A. Scheel, G. B. Cook, and S. A. Teukolsky. High-accuracy comparison of numerical relativity simulations with post-Newtonian expansions. *Phys. Rev.* D76 (2007), p. 124038. DOI: `10.1103/PhysRevD.76.124038` (cit. on p. 18).

152. Y. Pan, A. Buonanno, A. Taracchini, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi. Inspiral-merger-ringdown waveforms of spinning, precessing black-hole binaries in the effective-one-body formalism. 89 (2013), p. 084006. arXiv: `1307.6232 [gr-qc]` (cit. on pp. v, 3, 18, 75, 85, 118, 150, 160).

153. E. Berti, V. Cardoso, J. A. González, U. Sperhake, and B. Brügmann. Multipolar analysis of spinning binaries. *Classical and Quantum Gravity* 25 (11) (2008), p. 114035. URL: `http://stacks.iop.org/0264-9381/25/i=11/a=114035` (cit. on p. 18).

154. H. Antil, S. E. Field, F. Herrmann, R. H. Nochetto, and M. Tiglio. Two-step greedy algorithm for reduced order quadratures. *ArXiv e-prints* (2012). arXiv: `1210.0577 [cs.NA]` (cit. on p. 30).

155. P. Canizares, S. E. Field, J. Gair, V. Raymond, R. Smith, and M. Tiglio. Accelerated gravitational-wave parameter estimation with reduced order modeling. *Phys. Rev. Lett.* 114 (7) (2015), p. 071104. DOI: `10.1103/PhysRevLett.114.071104` (cit. on pp. 30, 174).

156. F. Herrmann, S. E. Field, C. R. Galley, E. Ochsner, and M. Tiglio. Towards beating the curse of dimensionality for gravitational waves using Reduced Basis. *Phys. Rev.* D86 (2012), p. 084046. DOI: `10.1103/PhysRevD.86.084046` (cit. on pp. 32, 36).

157. Sparse representations of gravitational waves from precessing compact binaries. *Physical Review Letters* 113 (2) (2014). URL: `http://dx.doi.org/10.1103/PhysRevLett.113.021101` (cit. on p. 35).

158. `http://www.gw-indigo.org` (cit. on p. 35).

159. M. Hannam. Modelling gravitational waves from precessing black-hole binaries: Progress, challenges and prospects. *Gen. Rel. Grav.* 46 (2014), p. 1767. arXiv: `1312.3641 [gr-qc]` (cit. on pp. 35, 37).

160. R. Bellman and R. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton Legacy Library. Princeton University Press, 1961. URL: `https://books.google.com/books?id=POAmAAAAMAAJ` (cit. on p. 35).

161. S. Caudill, S. E. Field, C. R. Galley, F. Herrmann, and M. Tiglio. Reduced Basis representations of multi-mode black hole ringdown gravitational waves. *Class. Quant. Grav.* 29 (2012), p. 095016. DOI: `10.1088/0264-9381/29/9/095016` (cit. on p. 36).

162. P. Binev, A. Cohen, W. Dahmen, R. A. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *SIAM J. Math. Analysis* 43 (3) (2011), p. 1457. URL: `http://dblp.uni-trier.de/db/journals/siamma/siamma43.html#BinevCDDPW11` (cit. on pp. 36, 64, 71, 137).

163. R. DeVore, G. Petrova, and P. Wojtaszczyk. Greedy algorithms for reduced bases in Banach spaces. *Constructive Approximation* 37 (3) (2013), p. 455. ISSN: 0176-4276. DOI: `10.1007/s00365-013-9186-2` (cit. on p. 36).

164. C. R. Galley, F. Herrmann, J. Silberholz, M. Tiglio, and G. Guerberoff. Statistical constraints on binary black hole inspiral dynamics. *Class. Quant. Grav.* 27 (2010), p. 245007. DOI: `10.1088/0264-9381/27/24/245007` (cit. on p. 36).

165. J. Lee and M. Verleysen. *Nonlinear Dimensionality Reduction*. Information Science and Statistics. Springer New York, 2007. ISBN: 9780387393513. URL: `https://books.google.com/books?id=o%5C_TIoyeO7AsC` (cit. on p. 36).

166. Hesthaven, Jan S., Stamm, Benjamin, and Zhang, Shun. Efficient greedy algorithms for high-dimensional parameter spaces with applications to empirical interpolation and reduced basis methods. *ESAIM: M2AN* 48 (1) (2014), pp. 259–283. DOI: `10.1051/m2an/2013100` (cit. on pp. 36, 37).

167. M. Boyle, R. Owen, and H. P. Pfeiffer. A geometric approach to the precession of compact binaries. 84 (2011), p. 124011. DOI: `10.1103/PhysRevD.84.124011` (cit. on pp. 37, 39, 77, 87, 98, 100, 101, 151).

168. M. Boyle. Angular velocity of gravitational radiation from precessing binaries and the corotating frame. 87 (10) (2013), p. 104006. DOI: `10.1103/PhysRevD.87.104006` (cit. on pp. 37, 39, 84, 87, 100).

169. P. Schmidt, M. Hannam, and S. Husa. Towards models of gravitational waveforms from generic binaries: A simple approximate mapping between precessing and non-precessing inspiral signals. 86 (2012), p. 104063. DOI: `10.1103/PhysRevD.86.104063` (cit. on pp. 37, 77).

170. A. Lundgren and R. O'Shaughnessy. Single-spin precessing gravitational waveform in closed form. 89 (4) (2014), p. 044021. DOI: `10.1103/PhysRevD.89.044021` (cit. on p. 37).

171. L. Pekowsky, R. O'Shaughnessy, J. Healy, and D. Shoemaker. Comparing gravitational waves from nonprecessing and precessing black hole binaries in the corotating frame. 88 (2013), p. 024040. arXiv: `1304.3176 [gr-qc]` (cit. on pp. 37, 44, 62, 77).

172. A. Bohé, S. Marsat, G. Faye, and L. Blanchet. Next-to-next-to-leading order spin-orbit effects in the near-zone metric and precession equations of compact binaries. 30 (2013), p. 075017. DOI: `10.1088/0264-9381/30/7/075017` (cit. on p. 39).

173. K. S. Thorne. Multipole expansions of gravitational radiation. *Rev. Mod. Phys.* 52 (2) (1980), pp. 299–339. DOI: `10.1103/RevModPhys.52.299` (cit. on p. 39).

174. B. J. Owen. Search templates for gravitational waves from inspiraling binaries: Choice of template spacing. 53 (12) (1996), pp. 6749–6761. DOI: `10.1103/PhysRevD.53.6749` (cit. on p. 44).

175. I. W. Harry, B. Allen, and B. S. Sathyaprakash. A Stochastic template placement algorithm for gravitational wave data analysis. *Phys. Rev.* D80 (2009), p. 104014. DOI: `10.1103/PhysRevD.80.104014` (cit. on p. 44).

176. P. Ajith, N. Fotopoulos, S. Privitera, A. Neunzert, and A. J. Weinstein. Effectual template bank for the detection of gravitational waves from inspiralling compact binaries with generic spins. 89 (8) (2014), p. 084041. DOI: `10.1103/PhysRevD.89.084041` (cit. on p. 44).

177. P. Ajith, M. Boyle, D. A. Brown, B. Brügmann, L. T. Buchman, et al. The NINJA-2 catalog of hybrid post-Newtonian/numerical-relativity waveforms for non-precessing black-hole binaries. 29 (12) (2012), p. 124001 (cit. on pp. 44, 50, 62).

178. I. Hinder et al. Error-analysis and comparison to analytical models of numerical waveforms produced by the NRAR Collaboration. 31 (2) (2014), p. 025012. arXiv: `1307.5307 [gr-qc]` (cit. on pp. 44, 49, 50, 62).

179. A. H. Mroue et al. Catalog of 174 Binary Black Hole Simulations for Gravitational Wave Astronomy. *Phys. Rev. Lett.* 111 (24) (2013), p. 241104. DOI: `10.1103/PhysRevLett.111.241104` (cit. on p. 44).

180. H. Antil, S. Field, F. Herrmann, R. Nochetto, and M. Tiglio. Two-step greedy algorithm for reduced order quadratures. *J. Sci. Comput.* 57 (3) (2013), pp. 604–637. ISSN: 0885-7474. DOI: `10.1007/s10915-013-9722-z` (cit. on p. 45).

181. *Triton: Code for Waveform Manipulations*. `https://github.com/moble/Triton` (cit. on p. 45).

182. L. S. Collaboration. *LSC Algorithm Library software packages* LAL, LALWRAPPER, *and* LALAPPS. URL: `http://www.lsc-group.phys.uwm.edu/lal` (cit. on pp. 45, 63, 119).

183. The Virgo Collaboration. *Advanced Virgo Baseline Design*. VIR-027A-09. 2010. URL: `https://tds.ego-gw.it/itf/tds/file.php?callFile=VIR-0027A-09.pdf` (cit. on p. 49).

184. B. P. Abbott et al. Prospects for observing and localizing gravitational-wave transients with advanced ligo and advanced virgo. 19 (2016), p. 1. DOI: `10.1007/lrr-2016-1` (cit. on pp. 49, 51).

185. J. Abadie et al. Predictions for the Rates of Compact Binary Coalescences Observable by Ground-based Gravitational-wave Detectors. 27 (2010), p. 173001. arXiv: `1003.2480 [gr-qc]` (cit. on p. 49).

186. L. S. Finn and D. F. Chernoff. Observing binary inspiral in gravitational radiation: one interferometer. 47 (1993), p. 2198 (cit. on p. 49).

187. J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. 91 (4 2015), p. 042003. DOI: `10.1103/PhysRevD.91.042003` (cit. on pp. 49, 75, 126).

188. L. Blanchet. Gravitational radiation from post-Newtonian sources and inspiralling compact binaries. 9 (2006), p. 4 (cit. on p. 49).

189. A. Buonanno and T. Damour. Effective one-body approach to general relativistic two-body dynamics. 59 (1999), p. 084006. DOI: `10.1103/PhysRevD.59.084006` (cit. on pp. 49, 62, 72).

190. A. Buonanno and T. Damour. Transition from inspiral to plunge in binary black hole coalescences. 62 (2000), p. 064015. DOI: `10.1103/PhysRevD.62.064015` (cit. on p. 49).

191. F. Pretorius. Evolution of binary black hole spacetimes. 95 (2005), p. 121101. DOI: `10.1103/PhysRevLett.95.121101` (cit. on pp. 49, 62, 75, 150).

192. M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower. Accurate evolutions of orbiting black-hole binaries without excision. 96 (2006), p. 111101. DOI: `10.1103/PhysRevLett.96.111101` (cit. on pp. 49, 62).

193. J. G. Baker, J. Centrella, D.-I. Choi, M. Koppitz, and J. van Meter. Gravitational wave extraction from an inspiraling configuration of merging black holes. 96 (2006), p. 111102. DOI: `10.1103/PhysRevLett.96.111102` (cit. on pp. 49, 62).

194. M. Campanelli, C. O. Lousto, Y. Zlochower, and D. Merritt. Maximum gravitational recoil. 98 (2007), p. 231102. eprint: `gr-qc/0702133` (cit. on p. 49).

195. J. A. González, U. Sperhake, B. Brügmann, M. Hannam, and S. Husa. Maximum kick from nonspinning black-hole binary inspiral. 98 (2007), p. 091101. eprint: `gr-qc/0610154` (cit. on p. 49).

196. A. Le Tiec, A. Buonanno, A. H. Mroué, D. A. Hemberger, G. Lovelace, H. P. Pfeiffer, L. E. Kidder, M. A. Scheel, B. Szilágyi, N. W. Taylor, et al. Periastron Advance in Spinning Black Hole Binaries: Gravitational Self-Force from Numerical Relativity. 88 (12 2013), p. 124027. arXiv: `1309.0541 [gr-qc]` (cit. on p. 49).

197. V. Cardoso, L. Gualtieri, C. Herdeiro, and U. Sperhake. Exploring new physics frontiers through numerical relativity (2014). arXiv: `1409.0014 [gr-qc]` (cit. on p. 49).

198. M. W. Choptuik, L. Lehner, and F. Pretorius. Probing Strong-Field Gravity Through Numerical Simulations. *General Relativity and Gravitation*. Ed. by A. Ashtekar, B. K. Berger, J. Isenberg, and M. MacCallum. Cambridge Books Online. Cambridge University Press, 2015, pp. 361–411. ISBN: 9781139583961. arXiv: `1502.06853 [gr-qc]` (cit. on p. 49).

199. P. Ajith, M. Hannam, S. Husa, Y. Chen, B. Brüegmann, N. Dorband, D. Mueller, F. Ohme, D. Pollney, C. Reisswig, et al. Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins. 106 (2011), p. 241101. arXiv: `0909.2867 [gr-qc]` (cit. on pp. 49, 56, 58, 75, 123).

200. L. Santamaría, F. Ohme, P. Ajith, B. Brügmann, N. Dorband, M. Hannam, S. Husa, P. Mösta, D. Pollney, C. Reisswig, et al. Matching post-Newtonian and numerical relativity waveforms: Systematic errors and a new phenomenological model for non-precessing black hole binaries. 82 (2010), p. 064016. DOI: `10.1103/PhysRevD.82.064016` (cit. on pp. 49, 56, 58, 62, 72, 75, 123).

201. T. Damour, A. Nagar, and S. Bernuzzi. Improved effective-one-body description of coalescing nonspinning black-hole binaries and its numerical-relativity completion. 87 (8) (2013), p. 084035. arXiv: `1212.4357 [gr-qc]` (cit. on pp. 49, 58, 62, 72).

202. T. Damour, A. Nagar, and M. Trias. Accuracy and effectualness of closed-form, frequency-domain waveforms for nonspinning black hole binaries. 83 (2) (2011), p. 024006. DOI: `10.1103/PhysRevD.83.024006` (cit. on p. 49).

203. I. MacDonald, A. H. Mroué, H. P. Pfeiffer, M. Boyle, L. E. Kidder, M. A. Scheel, B. Szilágyi, and N. W. Taylor. Suitability of hybrid gravitational waveforms for unequal-mass binaries. 87 (2013), p. 024009. DOI: `10.1103/PhysRevD.87.024009` (cit. on pp. 49, 50).

204. A. H. Mroué, M. A. Scheel, B. Szilágyi, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, S. Ossokine, N. W. Taylor, et al. A catalog of 174 binary black hole simulations for gravitational wave astronomy. 111 (2013), p. 241104. arXiv: `1304.6077 [gr-qc]` (cit. on pp. 49, 50, 52, 62).

205. Y. Pan, A. Buonanno, A. Taracchini, M. Boyle, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, B. Szilágyi, and A. Zenginoglu. Stability of non-spinning effective-one-body model in approximating two-body dynamics and gravitational-wave emission. 89 (2014), p. 061501. DOI: `10.1103/PhysRevD.89.061501` (cit. on pp. 50, 58).

206. B. Aylott, J. G. Baker, W. D. Boggs, M. Boyle, P. R. Brady, et al. Testing gravitational-wave searches with numerical relativity waveforms: Results from the first Numerical INJection Analysis (NINJA) project. 26 (2009), p. 165008. DOI: `10.1088/0264-9381/26/16/165008` (cit. on pp. 50, 62, 78, 96, 99).

207. M. Boyle, D. A. Brown, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, G. B. Cook, and S. A. Teukolsky. High-accuracy comparison of numerical relativity simulations with post-Newtonian expansions. 76 (12) (2007), p. 124038. arXiv: `0710.0158 [gr-qc]` (cit. on pp. 50, 57).

208. T. Damour, B. R. Iyer, and B. Sathyaprakash. A Comparison of search templates for gravitational waves from binary inspiral. 63 (2001), p. 044023. arXiv: `gr-qc/0010009 [gr-qc]` (cit. on p. 50).

209. `http://www.black-holes.org/SpEC.html` (cit. on pp. 51, 59, 62, 75, 76, 150, 151).

210. F. Foucart, L. Buchman, M. D. Duez, M. Grudich, L. E. Kidder, I. MacDonald, A. Mroue, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi. First direct comparison of nondisrupting neutron star-black hole and binary black hole merger simulations. 88 (6) (2013), p. 064017. arXiv: `1307.7685 [gr-qc]` (cit. on p. 52).

211. B. Szilágyi. Key Elements of Robustness in Binary Black Hole Evolutions using Spectral Methods. *Int. J. Mod. Phys.* D23 (7) (2014), p. 1430014. DOI: `10.1142/S0218271814300146` (cit. on pp. 52, 62, 76, 151).

212. M. A. Scheel, M. Giesler, D. A. Hemberger, G. Lovelace, K. Kuper, M. Boyle, B. Szilágyi, and L. E. Kidder. Improved methods for simulating nearly extremal binary black holes. 32 (10), 105009 (2015), p. 105009. DOI: `10.1088/0264-9381/32/10/105009` (cit. on pp. 52, 62, 64, 66, 75, 91).

213. N. W. Taylor, M. Boyle, C. Reisswig, M. A. Scheel, T. Chu, L. E. Kidder, and B. Szilágyi. Comparing gravitational waveform extrapolation to Cauchy-characteristic extraction in binary black hole simulations. 88 (12), 124010 (2013), p. 124010. DOI: `10.1103/PhysRevD.88.124010` (cit. on pp. 53, 54, 64, 67, 131).

214. L. T. Buchman and O. C. A. Sarbach. Towards absorbing outer boundaries in general relativity. 23 (2006), pp. 6709–6744 (cit. on p. 54).

215. I. MacDonald, S. Nissanke, and H. P. Pfeiffer. Suitability of post-Newtonian/numerical-relativity hybrid waveforms for gravitational wave detectors. 28 (13) (2011), p. 134002. ISSN: 0264-9381. DOI: `10.1088/0264-9381/28/13/134002` (cit. on pp. 54, 130, 163).

216. M. Boyle and A. H. Mroué. Extrapolating gravitational-wave data from numerical simulations. 80 (12) (2009), pp. 124045–14. DOI: `10.1103/PhysRevD.80.124045` (cit. on pp. 54, 95, 131, 151).

217. D. Shoemaker. *Advanced LIGO anticipated sensitivity curves*. LIGO Document T0900288-v3. 2010. URL: `https://dcc.ligo.org/cgi-bin/DocDB/ShowDocument?docid=2974` (cit. on pp. 56, 121, 160).

218. A. Buonanno, B. R. Iyer, E. Ochsner, Y. Pan, and B. S. Sathyaprakash. Comparison of post-newtonian templates for compact binary inspiral signals in gravitational-wave detectors. 80 (8) (2009), p. 084043. DOI: `10.1103/PhysRevD.80.084043` (cit. on pp. 56, 57).

219. A. Taracchini, Y. Pan, A. Buonanno, E. Barausse, M. Boyle, T. Chu, G. Lovelace, H. P. Pfeiffer, and M. A. Scheel. Prototype effective-one-body model for nonprecessing spinning inspiral-merger-ringdown waveforms. 86 (2), 024011 (2012), p. 024011. DOI: `10.1103/PhysRevD.86.024011` (cit. on pp. 56, 57, 62, 72).

220. D. McKechan, C. Robinson, and B. Sathyaprakash. A tapering window for time-domain templates and simulated signals in the detection of gravitational waves from coalescing compact binaries. 27 (2010), p. 084020. DOI: `10.1088/0264-9381/27/8/084020` (cit. on pp. 56, 106, 119, 126).

221. C. Loken, D. Gruner, L. Groer, R. Peltier, N. Bunn, M. Craig, T. Henriques, J. Dempsey, C.-H. Yu, J. Chen, et al. SciNet: Lessons Learned from Building a Power-efficient Top-20 System and Data Centre. *J. Phys.: Conf. Ser.* 256 (2010), p. 012026. DOI: `10.1088/1742-6596/256/1/012026` (cit. on p. 59).

222. J. Aasi et al. The NINJA-2 project: Detecting and characterizing gravitational waveforms modelled using numerical binary black hole simulations. 31 (2014), p. 115004. arXiv: `1401.0939 [gr-qc]` (cit. on p. 62).

223. G. M. Harry (for the LIGO Scientific Collaboration). 27 (2010), p. 084006 (cit. on p. 62).

224. F. Acernese et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. 32 (2) (2015), p. 024001. DOI: `10.1088/0264-9381/32/2/024001` (cit. on p. 62).

225. J. Abadie, B. P. Abbott, R. Abbott, M. Abernathy, T. Accadia, F. Acernese, C. Adams, R. Adhikari, P. Ajith, B. Allen, et al. TOPICAL REVIEW: Predictions for the rates of compact binary coalescences observable by ground-based gravitational-wave detectors. 27 (17) (2010), p. 173001. DOI: `10.1088/0264-9381/27/17/173001` (cit. on p. 62).

226. P. Kumar, I. MacDonald, D. Brown, H. Pfeiffer, K. Cannon, M. Boyle, L. Kidder, A. Mroue, M. Scheel, B. Szilágyi, et al. Template banks for binary black hole searches with numerical relativity waveforms. 89 (2014), p. 042002. eprint: `1310.7949` (cit. on p. 62).

227. T. Damour and A. Nagar. An Improved analytical description of inspiralling and coalescing black-hole binaries. 79 (2009), p. 081503. DOI: `10.1103/PhysRevD.79.081503` (cit. on pp. 62, 72, 75, 85, 150).

228. A. Buonanno, Y. Pan, H. P. Pfeiffer, M. A. Scheel, L. T. Buchman, and L. E. Kidder. Effective-one-body waveforms calibrated to numerical relativity simulations: Coalescence of non-spinning, equal-mass black holes. 79 (2009), p. 124028. DOI: `10.1103/PhysRevD.79.124028` (cit. on pp. 62, 72).

229. Y. Pan, A. Buonanno, L. T. Buchman, T. Chu, L. E. Kidder, H. P. Pfeiffer, and M. A. Scheel. Effective-one-body waveforms calibrated to numerical relativity simulations: coalescence of non-precessing, spinning, equal-mass black holes. 81 (2010), p. 084041. DOI: `10.1103/PhysRevD.81.084041` (cit. on pp. 62, 63, 72).

230. P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, J. T. Whelan, B. Brügmann, P. Diener, J. Gonzalez, M. H. S. Husa, et al. A phenomenological template family for black-hole coalescence waveforms. 24 (19) (2007), S689. arXiv: `0704.3764 [gr-qc]` (cit. on pp. 62, 72).

231. R. Sturani, S. Fischetti, L. Cadonati, G. Guidi, J. Healy, et al. Complete phenomenological gravitational waveforms from spinning coalescing binaries. *J.Phys.Conf.Ser.* 243 (2010), p. 012007. DOI: `10.1088/1742-6596/243/1/012007` (cit. on pp. 62, 72).

232. J. Abadie et al. Search for gravitational waves from binary black hole inspiral, merger and ringdown. 83 (2011), p. 122005. DOI: `10.1103/PhysRevD.83.122005` (cit. on pp. 62, 72).

233. J. Kaye. *The interpolation of gravitational waveforms*. PhD thesis. Brown University, 2012. URL: `%5Curl%7Bhttp://brown.edu/research/projects/scientific-computing/sites/brown.edu.research.projects.scientific-computing/files/uploads/The%20interpolation.pdf%7D` (cit. on pp. 62, 78).

234. Y. Maday, A. T. Patera, and G. Turinici. A priori convergence theory for reduced-basis approximations of single-parameter elliptic partial differential equations. *J. Sci. Comput.* 17 (1-4) (2002), p. 437. ISSN: 0885-7474. DOI: `10.1023/A:1015145924517` (cit. on p. 62).

235. K. Veroy, C. Prud'homme, and A. T. Patera. Reduced-basis approximation of the viscous burgers equation: rigorous a posteriori error bounds. *Comptes Rendus Mathematique* 337 (9) (2003), pp. 619–624. ISSN: 1631-073X. DOI: `10.1016/j.crma.2003.09.023` (cit. on p. 62).

236. C. Prud'homme, D. V. Rovas, K. Veroy, L. Machiels, Y. Maday, A. T. Patera, and G. Turinici. Reliable real-time solution of parametrized partial differential equations: reduced-basis output bound methods. *Journal of Fluids Engineering* 124 (1) (2002), p. 70. DOI: `10.1115/1.1448332` (cit. on p. 62).

237. Y. Chen, J. S. Hesthaven, Y. Maday, and J. Rodríguez. Certified reduced basis methods and output bounds for the harmonic maxwell's equations. *SIAM J. Sci. Comput.* 32 (2 2010), pp. 970–996. ISSN: 1064-8275. DOI: `http://dx.doi.org/10.1137/09075250X` (cit. on p. 62).

238. A. Quarteroni, G. Rozza, and A. Manzoni. Certified reduced basis approximation for parametrized partial differential equations and applications. *Journal of Mathematics in Industry* 1 (1) (2011), pp. 1–49. DOI: `10.1186/2190-5983-1-3` (cit. on p. 62).

239. M. Pürrer. Frequency-domain reduced order models for gravitational waves from aligned-spin compact binaries. 31 (19) (2014), p. 195010. arXiv: `1402.4146 [gr-qc]` (cit. on pp. 63, 68, 76, 78, 82, 86–88, 112, 126, 150, 174).

240. A. H. Mroué and H. P. Pfeiffer. Precessing Binary Black Holes Simulations: Quasicircular Initial Data (2012). arXiv: `1210.2958 [gr-qc]` (cit. on p. 64).

241. N. T. Bishop, R. Gomez, L. Lehner, M. Maharaj, and J. Winicour. High-powered gravitational news. *Phys. Rev. D* 56 (1997), p. 6298. DOI: `10.1103/PhysRevD.56.6298` (cit. on pp. 64, 131).

242. M. C. Babiuc, B. Szilágyi, J. Winicour, and Y. Zlochower. A characteristic extraction tool for gravitational waveforms. 84 (4 2011), p. 044057. DOI: `10.1103/PhysRevD.84.044057` (cit. on pp. 64, 131).

243. J. Winicour. Characteristic evolution and matching. 12 (3) (2009). URL: `http://www.livingreviews.org/lrr-2009-3` (cit. on p. 64).

244. C. Reisswig and D. Pollney. Notes on the integration of numerical relativity waveforms. 28 (2011), p. 195015. DOI: `10.1088/0264-9381/28/19/195015` (cit. on pp. 64, 66).

245. A. H. Mroué, H. P. Pfeiffer, L. E. Kidder, and S. A. Teukolsky. Measuring orbital eccentricity and periastron advance in quasi-circular black hole simulations. 82 (2010), p. 124016. eprint: `arXiv:1004.4697[gr-qc]` (cit. on p. 66).

246. Y. Maday, N. Nguyen, A. T. Patera, and S. H. Pau. A general multipurpose interpolation procedure: the magic points. *Communications on Pure and Applied Analysis* 8 (2009), pp. 383–404. DOI: `10.3934/cpaa.2009.8.383` (cit. on p. 68).

247. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001 (cit. on pp. 69, 88).

248. J. Blackman, B. Szilagyi, C. R. Galley, and M. Tiglio. Sparse representations of gravitational waves from precessing compact binaries. *Phys. Rev. Lett.* 113 (2 2014), p. 021101. DOI: `10.1103/PhysRevLett.113.021101`. (Cit. on pp. 32, 72, 85, 93, 138, 177).

249. *Simulating eXtreme Spacetimes*. `http://www.black-holes.org/` (cit. on p. 72).

250. B. P. Abbott et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.* 116 (6 2016), p. 061102. DOI: `10.1103/PhysRevLett.116.061102` (cit. on pp. v, 1, 75, 150, 169, 171).

251. B. P. Abbott et al. GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence. *Phys. Rev. Lett.* 116 (24) (2016), p. 241103. DOI: `10.1103/PhysRevLett.116.241103` (cit. on pp. 1, 75, 130, 150).

252. B. P. Abbott et al. Tests of general relativity with GW150914. *Phys. Rev. Lett.* 116 (2016), p. 221101. arXiv: `1602.03841 [gr-qc]` (cit. on pp. v, 2, 7, 75, 150, 171).

253. Y. Zlochower, J. Baker, M. Campanelli, and C. Lousto. Accurate black hole evolutions by fourth-order numerical relativity. 72 (2005), p. 024021. DOI: `10.1103/PhysRevD.72.024021` (cit. on pp. 75, 150).

254. Einstein Toolkit home page: `http://einsteintoolkit.org` (cit. on pp. 75, 150).

255. S. Husa, J. A. González, M. Hannam, B. Brügmann, and U. Sperhake. Reducing phase error in long numerical binary black hole evolutions with sixth order finite differencing. 25 (2008), p. 105006 (cit. on pp. 75, 150).

256. B. Brügmann, J. A. González, M. Hannam, S. Husa, U. Sperhake, and W. Tichy. Calibration of moving puncture simulations. 77 (2), 024027 (2008), p. 024027. DOI: `10.1103/PhysRevD.77.024027` (cit. on pp. 75, 150).

257. F. Herrmann, I. Hinder, D. Shoemaker, and P. Laguna. Unequal mass binary black hole plunges and gravitational recoil. 24 (2007), S33–S42. eprint: `gr-qc/0601026` (cit. on pp. 75, 150).

258. I. Ruchlin, J. Healy, C. O. Lousto, and Y. Zlochower. New Puncture Initial Data for Black-Hole Binaries: High Spins and High Boosts (2014). arXiv: `1410.8607 [gr-qc]` (cit. on p. 75).

259. C. O. Lousto and Y. Zlochower. Orbital evolution of extreme-mass-ratio black-hole binaries with numerical relativity. 106 (4 2011), p. 041101. DOI: `10.1103/PhysRevLett.106.041101` (cit. on pp. 75, 91).

260. S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé. Frequency-domain gravitational waves from non-precessing black-hole binaries. i. new numerical waveforms and anatomy of the signal. *arXiv preprint arXiv:1508.07250* (2015) (cit. on p. 75).

261. S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé. Frequency-domain gravitational waves from non-precessing black-hole binaries. I. New numerical waveforms and anatomy of the signal. *arXiv:1508.07250* (2015) (cit. on pp. 75, 85, 150).

262. A. Bohé et al. An improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors (2016). arXiv: `1611.03703 [gr-qc]` (cit. on pp. 75, 85, 150).

263. B. P. Abbott et al. An improved analysis of GW150914 using a fully spin-precessing waveform model. *Submitted to Phys. Rev. D.; arXiv:1606.01210* (2016). arXiv: `1606.01210 [gr-qc]` (cit. on pp. v, 2, 75, 76, 125, 130).

264. T. Damour, P. Jaranowski, and G. Schaefer. Effective one body approach to the dynamics of two spinning black holes with next-to-leading order spin-orbit coupling. 78 (2008), p. 024009. DOI: `10.1103/PhysRevD.78.024009` (cit. on pp. 75, 85, 150).

265. M. Pürrer. Frequency domain reduced order model of aligned-spin effective-one-body waveforms with generic mass ratios and spins. 93 (6), 064041 (2016), p. 064041. DOI: `10.1103/PhysRevD.93.064041` (cit. on pp. v, 75, 76, 78, 82, 85–88, 112, 126, 150, 174).

266. K. Jani, J. Healy, J. A. Clark, L. London, P. Laguna, and D. Shoemaker. Georgia Tech Catalog of Gravitational Waveforms. *Class. Quant. Grav.* 33 (20) (2016), p. 204001. DOI: `10.1088/0264-9381/33/20/204001` (cit. on p. 75).

267. S. Babak, A. Taracchini, and A. Buonanno. Validating the effective-one-body model of spinning, precessing binary black holes against numerical relativity (2016). arXiv: `1607.05661 [gr-qc]` (cit. on p. 75).

268. B. P. Abbott et al. Effects of waveform model systematics on the interpretation of GW150914 (2016). arXiv: `1611.07531 [gr-qc]` (cit. on pp. vi, 3, 75, 76, 130).

269. J. Blackman, S. E. Field, C. R. Galley, B. Szilágyi, M. A. Scheel, M. Tiglio, and D. A. Hemberger. Fast and Accurate Prediction of Numerical Relativity Waveforms from Binary Black Hole Coalescences Using Surrogate Models. *Phys. Rev. Lett.* 115 (12) (2015), p. 121102. DOI: `10.1103/PhysRevLett.`

`115.121102`. (Cit. on pp. vi, 18, 60, 76, 78, 86–88, 112, 129, 132, 150, 168, 171).

270. H. P. Pfeiffer, L. E. Kidder, M. A. Scheel, and S. A. Teukolsky. A multidomain spectral method for solving elliptic equations. 152 (2003), pp. 253–273. DOI: `10.1016/S0010-4655(02)00847-0` (cit. on pp. 76, 151).

271. G. Lovelace, R. Owen, H. P. Pfeiffer, and T. Chu. Binary-black-hole initial data with nearly-extremal spins. 78 (2008), p. 084017. DOI: `10.1103/PhysRevD.78.084017` (cit. on pp. 76, 151).

272. L. Lindblom, M. A. Scheel, L. E. Kidder, R. Owen, and O. Rinne. A new generalized harmonic evolution system. 23 (2006), S447. DOI: `10.1088/0264-9381/23/16/S09` (cit. on pp. 76, 151).

273. B. Szilágyi, L. Lindblom, and M. A. Scheel. Simulations of binary black hole mergers using spectral methods. 80 (2009), p. 124010. arXiv: `0909.3557 [gr-qc]` (cit. on pp. 76, 151).

274. M. A. Scheel, M. Boyle, T. Chu, L. E. Kidder, K. D. Matthews and H. P. Pfeiffer. High-accuracy waveforms for binary black hole inspiral, merger, and ringdown. 79 (2009), p. 024003. eprint: `arXiv:gr-qc/0810.1767` (cit. on pp. 76, 151).

275. P. Schmidt, M. Hannam, S. Husa, and P. Ajith. Tracking the precession of compact binaries from their gravitational-wave signal. 84 (2011), p. 024046. eprint: `arxiv:1012.2879` (cit. on pp. 77, 87).

276. O'Shaughnessy, R. and Vaishnav, B. and Healy, J. and Meeks, Z. and Shoemaker, D. *Efficient asymptotic frame selection for binary black hole spacetimes using asymptotic radiation*. 2011. eprint: `arXiv:1109.5224` (cit. on pp. 77, 87).

277. K. Cannon, J. Emberson, C. Hanna, D. Keppel, and H. Pfeiffer. Interpolation in waveform space: enhancing the accuracy of gravitational waveform families using numerical relativity. 87 (4) (2013), p. 044008. DOI: `10.1103/PhysRevD.87.044008` (cit. on p. 78).

278. K. Cannon, C. Hanna, and D. Keppel. Interpolating compact binary waveforms using the singular value decomposition. 85 (2012), p. 081504. DOI: `10.1103/PhysRevD.85.081504` (cit. on p. 78).

279. T. A. Apostolatos, C. Cutler, G. J. Sussman, and K. S. Thorne. Spin-induced orbital precession and its modulation of the gravitational waveforms from merging binaries. 49 (1994), pp. 6274–6297 (cit. on pp. 78, 125).

280. L. E. Kidder. Coalescing binary systems of compact objects to postNewtonian 5/2 order. 5. Spin effects. *Phys. Rev.* D52 (1995), pp. 821–847. DOI: `10.1103/PhysRevD.52.821` (cit. on p. 78).

281. D. Christodoulou and R. Ruffini. Reversible Transformations of a Charged Black Hole. 4 (1971), pp. 3552–3555. DOI: `10.1103/PhysRevD.4.3552` (cit. on p. 78).

282. S. Caudill, S. E. Field, C. R. Galley, F. Herrmann, and M. Tiglio. Reduced basis representations of multi-mode black hole ringdown gravitational waves. 29 (9) (2012), p. 095016 (cit. on pp. 81, 137).

283. *GWFrames*. `https://github.com/moble/GWFrames` (cit. on pp. 84, 100).

284. F. Herrmann, S. E. Field, C. R. Galley, E. Ochsner, and M. Tiglio. Towards beating the curse of dimensionality for gravitational waves using Reduced Basis. 86 (2012), p. 084046. DOI: `10.1103/PhysRevD.86.084046` (cit. on p. 86).

285. D. A. Hemberger, M. A. Scheel, J. Blackman, S. E. Field, C. R. Galley, P. Schmidt, R. Smith, S. Ossokine, et al. (in prep) Numerical Relativity Simulations of Precessing Binary Black Hole Mergers for Surrogate Modeling (2017) (cit. on pp. 87, 95).

286. M. Boyle, L. E. Kidder, S. Ossokine, and H. P. Pfeiffer. Gravitational-wave modes from precessing black-hole binaries (2014). arXiv:1409.4431. arXiv: `1409.4431` (cit. on pp. 87, 105, 151).

287. C. Cutler and E. E. Flanagan. Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral wave form? 49 (1994), p. 2658. DOI: `10.1103/PhysRevD.49.2658` (cit. on pp. 90, 142).

288. D. Foreman-Mackey. Corner.py: scatterplot matrices in python. *The Journal of Open Source Software* 24 (2016). DOI: `10.21105/joss.00024` (cit. on p. 94).

289. M. Boyle. Angular velocity of gravitational radiation from precessing binaries and the corotating frame. 87 (10) (2013), p. 104006. DOI: `10.1103/PhysRevD.87.104006` (cit. on pp. 104, 147, 151).

290. R. R. Hocking. A biometrics invited paper. the analysis and selection of variables in linear regression. *Biometrics* 32 (1) (1976), pp. 1–49. ISSN: 0006341X, 15410420. URL: `http://www.jstor.org/stable/2529336` (cit. on pp. 113, 134).

291. P. Schmidt, F. Ohme, and M. Hannam. Towards models of gravitational waveforms from generic binaries II: Modelling precession effects with a single effective precession parameter. 91 (2) (2015), p. 024043. DOI: `10.1103/PhysRevD.91.024043` (cit. on pp. 123, 125).

292. A. Buonanno, Y. Chen, Y. Pan, and M. Vallisneri. Quasiphysical family of gravity-wave templates for precessing binaries of spinning compact objects: Application to double-spin precessing binaries. 70 (2004), p. 104003 (cit. on p. 125).

293. C. De Boor. *A Practical Guide to Splines*. Applied Mathematical Sciences v. 27. Springer-Verlag, 1978. ISBN: 9783540903567 (cit. on pp. 128, 179).

294. T. A. Davis. Algorithm 832: umfpack v4.3—an unsymmetric-pattern multifrontal method. *ACM Trans. Math. Softw.* 30 (2) (2004), pp. 196–199. ISSN: 0098-3500. DOI: `10.1145/992200.992206` (cit. on p. 128).

295. T. A. Davis. A column pre-ordering strategy for the unsymmetric-pattern multifrontal method. *ACM Trans. Math. Softw.* 30 (2) (2004), pp. 165–195. ISSN: 0098-3500. DOI: `10.1145/992200.992205` (cit. on p. 128).

296. X. S. Li and J. W. Demmel. Superlu_dist: a scalable distributed-memory sparse direct solver for unsymmetric linear systems. *ACM Trans. Mathematical Software* 29 (2) (2003), pp. 110–140 (cit. on p. 128).

297. X. Li, J. Demmel, J. Gilbert, iL. Grigori, M. Shao, and I. Yamazaki. *SuperLU Users' Guide*. Tech. rep. LBNL-44289. `http://crd.lbl.gov/~xiaoye/SuperLU/`. Last update: August 2011. Lawrence Berkeley National Laboratory, 1999 (cit. on p. 128).

298. J. C. Bustillo, A. Bohé, S. Husa, A. M. Sintes, M. Hannam, et al. Comparison of subdominant gravitational wave harmonics between post-Newtonian and numerical relativity calculations and construction of multi-mode hybrids (2015). arXiv: `1501.00918 [gr-qc]` (cit. on pp. 130, 163).

299. M. Boyle. Uncertainty in hybrid gravitational waveforms: optimizing initial orbital frequencies for binary black-hole simulations. 84 (6 2011), p. 064013. DOI: `10.1103/PhysRevD.84.064013` (cit. on pp. 130, 163).

300. F. Ohme, M. Hannam, and S. Husa. Reliability of complete gravitational waveform models for compact binary coalescences. 84 (6 2011), p. 064029. DOI: `10.1103/PhysRevD.84.064029` (cit. on pp. 130, 163).

301. J. Winicour. Characteristic evolution and matching. 1 (1998), p. 5. DOI: `10.12942/lrr-2012-2` (cit. on p. 131).

302. D. Pollney and C. Reisswig. Gravitational memory in binary black hole mergers. 732 (2011), p. L13. DOI: `10.1088/2041-8205/732/1/L13` (cit. on p. 131).

303. M. Favata. The gravitational-wave memory effect. 27 (2010), p. 084036. DOI: `10.1088/0264-9381/27/8/084036` (cit. on p. 131).

304. T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. The MIT Press, 2009 (cit. on p. 137).

305. W. Hoffmann. Iterative algorithms for gram-schmidt orthogonalization. *Computing* 41 (4) (1989), pp. 335–348. ISSN: 1436-5057. DOI: `10.1007/BF02241222` (cit. on p. 137).

306. C. R. Galley and P. Schmidt. Fast and efficient evaluation of gravitational waveforms via reduced-order spline interpolation (2016). arXiv: `1611.07529 [gr-qc]` (cit. on p. 137).

307. K. Cannon et al. Singular value decomposition applied to compact binary coalescence gravitational-wave signals. 82 (2010), p. 044025. DOI: `10.1103/PhysRevD.82.044025` (cit. on p. 137).

308. S. Field. GreedyCPP home page: `https://bitbucket.org/sfield83/greedycpp` (cit. on p. 138).

309. C. R. Galley. *Hierarchical reduced basis building strategies*. (unpublished). 2014 (cit. on p. 138).

310. Hesthaven, Jan S., Stamm, Benjamin, and Zhang, Shun. Efficient greedy algorithms for high-dimensional parameter spaces with applications to empirical interpolation and reduced basis methods. *ESAIM: M2AN* 48 (1) (2014), pp. 259–283. DOI: `10.1051/m2an/2013100` (cit. on p. 138).

311. LIGO Scientific Collaboration. *Advanced LIGO anticipated sensitivity curves*. `https://dcc.ligo.org/LIGO-T0900288/public`. 2011 (cit. on p. 142).

312. L. Lindblom. Use and abuse of the model waveform accuracy standards. 80 (6) (2009), p. 064019. arXiv: `0907.0457` (cit. on p. 144).

313. *Scri*. `https://github.com/moble/scri` (cit. on p. 151).

314. H.-J. Bungartz and M. Griebel. Sparse grids. *Acta numerica* 13 (2004), pp. 147–269 (cit. on pp. 152, 165).

315. L. Blanchet. Gravitational radiation from post-Newtonian sources and inspiralling compact binaries. 17 (2014), p. 2. URL: `http://www.livingreviews.org/lrr-2014-2` (cit. on p. 154).

316. J. C. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley, 2003 (cit. on pp. 157, 166).

317. F. Bashforth and J. C. Adams. *An attempt to test the theories of capillary action by comparing the theoretical and measured forms of drops of fluid*. Cambridge University Press, 1883. URL: `http://ci.nii.ac.jp/ncid/BA66330495` (cit. on pp. 157, 166).

318. `http://www.black-holes.org/surrogates/` (cit. on pp. 157, 163).

319. P. D. Lasky, E. Thrane, Y. Levin, J. Blackman, and Y. Chen. Detecting gravitational-wave memory with ligo: implications of gw150914. *Phys. Rev. Lett.* 117 (6 2016), p. 061102. DOI: `10.1103/PhysRevLett.117.061102`. (Cit. on pp. 131, 168, 170, 171).

320. A. Loeb. Electromagnetic Counterparts to Black Hole Mergers Detected by LIGO. *Astrophys. J.* 819 (2) (2016), p. L21. DOI: `10.3847/2041-8205/819/2/L21` (cit. on p. 171).