Abstract

Title of dissertation:     APPLYING NUMERICAL RELATIVITY
                           TO GRAVITATIONAL WAVE ASTRONOMY

                           Sean Thomas McWilliams
                           Doctor of Philosophy, 2008

Dissertation directed by:   Professor Peter Shawhan,
                            Department of Physics

General relativity predicts the existence of gravitational waves produced by the motion of massive objects. The inspiral, merger, and ringdown of black hole binaries is expected to be one of the brightest sources in the gravitational wave sky. Interferometric detectors, such as the current ground-based Laser Interferometer Gravitational Wave Observatory (LIGO) and the future space-based Laser Interferometer Space Antenna (LISA), measure the influx of gravitational radiation from the whole sky. Each physical process that emits gravitational radiation will have a unique waveform, and prior knowledge of these waveforms is needed to distinguish a signal from the noise inherent in the interferometer. In the strong field regime of the merger, only numerical relativity, which solves the full set of Einstein's equations numerically, has been able to provide accurate waveforms.

We present a comprehensive study of the nonspinning portion of parameter space for which we have generated accurate simulations of the late inspiral through merger and ringdown, and a comparison of those results with predictions from the adiabatic Taylor-expanded post-Newtonian (PN) and effective-one-body (EOB) PN approximations. We then focus on data analysis questions using the equal-mass nonspinning as well as the 2:1, 4:1, and 6:1 mass ratio nonspinning black hole binary (BHB) waveforms. We construct a full waveform by combining our results from numerical relativity with a highly accurate Taylor PN approximation, and use it to calculate signal-to-noise ratios (SNRs) for several detectors. We measure the mass ratio scaling of the waveform amplitude through the inspiral and merger, and compare our observations with predictions from PN. Lastly, we turn our focus to parameter estimation with LISA, and investigate the increased accuracy with which parameters can be measured by including both the merger and inspiral waveforms, compared to estimates without numerical waveforms which can only incorporate the inspiral. We use the equal mass, nonspinning waveform as a test case and assess the parameter uncertainty by means of the Fisher matrix formalism.

# APPLYING NUMERICAL RELATIVITY
# TO GRAVITATIONAL WAVE ASTRONOMY

by

Sean Thomas McWilliams

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Professor Peter Shawhan, Chair
Dr. Joan Centrella, Co-Chair
Dr. John Baker
Professor Alessandra Buonanno
Professor M. Coleman Miller
Professor Manuel Tiglio

# Preface

Over the last few years, I have benefited from having access to `Hahndol`, the numerical relativity finite difference code written and maintained at NASA Goddard Space Flight Center. The waveforms shown throughout this work were all generated using `Hahndol`. Likewise, I have had the opportunity to collaborate with the scientists at NASA Goddard, and a significant portion of Chapters 3 and 4 has already been published with those collaborators as coauthors [1, 2]. For the unequal mass work, which as yet unpublished, James van Meter performed the numerical simulations, but the analysis and text presented is my own work.

The work on parameter estimation has been done in collaboration with Ira Thorpe, John Baker, and Keith Arnaud. I wrote the original driver code in C++ for generating the waveforms and incorporating LISA's response via interfacing with Synthetic LISA, which uses both Python and C++. John and I wrote the original Python code that actually ran Synthetic LISA, and used the resulting TDI observables to calculate the Fisher and covariance matrices, thus giving the parameter accuracy estimate. Keith contributed improvements to the code, and Ira subsequently converted most of the Python content to C++ to make the code more compact, and he has subsequently tested and validated the code extensively while I have been occupied with this work, for which I am exceedingly grateful. Some of the results from the code presented here were presented at the 12th Gravitational Wave Data Analysis Workshop (GWDAW) and are being prepared for publication in [3].

# Dedication

For my Rose, who supported me with her boundless inner strength and with her

     unconditional love.

For my Mom, who protected me and shared my every pain as if it were her own.

And for my Dad, my hero, who showed me how to be an honorable man.

# Acknowledgements

I'd like to express my gratitude to my colleagues at NASA Goddard, Darian Boggs, Bernard Kelly, Jim van Meter, and Ira Thorpe for their support of me throughout the completion of this work, for their assistance with many aspects of this work, and for the intellectually stimulating environment that they've always provided. I'd like to thank my advisors, Joan Centrella and John Baker at NASA and Peter Shawhan at UMD, for their patience and for their willingness to take the time to help me grow as a scientist. I'd also like to acknowledge the Laboratory for High Energy Astrophysics Fellowship, the Leon A. Herreid Fellowship, and the LISA project for providing the funding that made my research endeavors possible. I'd like to single out Tuck Stebbins, LISA's Project Scientist at Goddard, for repeatedly scrounging money for me from the project during a very lean period for LISA. I'd also like to thank the other members of my dissertation committee, Alessandra Buonanno, Cole Miller, and Manuel Tiglio, for their cooperativeness and patience through the several schedule changes and delays that, I have subsequently discovered, are an unwritten prerequisite for receiving a Ph. D..

I could never adequately thank my family enough. My brother, Tommy, and my sister, Bethy, have each expressed how proud they are of me, their little brother, and that meant more to me than they'll ever know. My mother, Mom, and my father, Dad, have always been the most supportive people in my life. Their lives have always been their children, and they are the two most selfless people I have

ever known. They are my heroes.

And lastly, no words could express what my wonderful wife, Stephanie, our perfect little boy, Sean Jr., and our precious little girl, Ella, have meant to me during the completion of this work. Stephanie knows me like no one else, and she has supported me in exactly the way I needed every step of the way. She knew I was having trouble concentrating, and she sent me away to work on my own... while she was eight months pregnant with Ella. If she could have written the thing for me, I know she would have. She is my best friend, my partner in life. And Sean Jr. and Ella are the center of our universe, my little angels. I do it all for them.

# Table of Contents

## List of Tables

# List of Figures

xi

# List of Abbreviations

| | |
|---|---|
| AMR | Adaptive mesh refinement |
| BBN | Big Bang nucleosynthesis |
| BHB | Black hole binary |
| EOB | Effective one body |
| ISCO | Innermost stable circular orbit |
| LIGO | Laser Interferometer Gravitational-Wave Observatory |
| LISA | Laser Interferometer Space Antenna |
| NASA | National Aeronautics and Space Administration |
| NR | Numerical relativity |
| PN | Post-Newtonian |
| SMBHB | Super-massive black hole binary |
| SNR | Signal-to-noise ratio |
| SPA | Stationary phase approximation |

# Chapter 1

# Introduction

*Henceforth, space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.*

-Hermann Minkowski

*We are all made of stars.*

-Moby

## 1.1 The Birth of Black Holes

To fully retrace the steps that lead to the creation of black holes, we must first follow the progression that brought about the birth of stars. Stars are the factories that take raw materials, hydrogen and helium in this case, and produce a steady supply of heavier elements for the entire universe. In fact, without stars there would be no elements heavier than beryllium [4]. Roughly one second after the Big Bang, when the universe had cooled enough to sustain stable protons and neutrons, the entire universe behaved like a fledgling star, fusing successively more massive components for the first three minutes. This period is known as Big Bang nucleosynthesis, or BBN [5, 6]. This period ended when the universe reached a bottleneck trying to fuse either 5 or 8 nucleons into a nucleus, since both configurations are unstable. Stars circumvent the BBN bottleneck by means of the triple-alpha process, whereby

1

three helium-4 nuclei produce a carbon nucleus. However, the reaction rate of the triple-alpha process is proportional to $\rho^2$, where $\rho$ is the density of the gas. Therefore, since the typical density of the universe within minutes of the Big Bang was too low for helium-4 nuclei to efficiently find each other, the triple-alpha process would take tens of thousands of years, and so would not be sufficiently expeditious to prevent the bottleneck which ended Big Bang nucleosynthesis [4]. It took hundreds of thousands of years for the universe to reach a temperature where the nuclei could capture free electrons to form the first atoms. It was at this point that the universe became optically thin, and photons could span the cosmos unimpeded.

For another billion years, very little happened. The slight inhomogeneities in the mass distribution of the early universe that were observed by COBE and, more recently, WMAP, led to pockets with higher density than the surrounding space. Eventually, clouds of gravitationally bound matter began to form, meaning that instead of the gas spreading out as it otherwise would in a rapidly expanding universe, the cloud would contract under gravity. For such cases, the gravitational attraction is counteracted primarily by the gas pressure, although rotation of the cloud and internal electromagnetic repulsion can play significant roles. If the mass exceeds a certain limit (referred to as the Jeans mass when rotation and electromagnetism are neglected), the cloud will collapse under the force of gravity, building up ever greater pressure at the core of the cloud until fusion is ignited [4].

Hydrogen fusion will continuously populate the core with helium. Where the star evolves from there depends on a number of factors, but the dominant factor, and the one which determines a star's ultimate fate, is its mass. The star may

2

continue fusing hydrogen in the envelope but be too light to fuse helium until very late in the evolution. In this case, the hydrogen-fusing envelope actually generates more luminosity than the hydrogen-fusing core did, but some of that energy goes into slowly expanding the envelope, thereby decreasing its temperature and shifting it to the red part of the visible spectrum. This type of star is referred to as a red giant. On the other hand, the star may be sufficiently massive to fuse all the elements up to iron, which requires an endothermic reaction to fuse and so will not occur spontaneously [4]. The latter possibility will lead to a particular class of supernova, or catastrophic stellar explosion, which will in turn provide the energy to fuse all the stable elements with atomic numbers higher than iron. Therefore, all the naturally-occurring elements heavier than beryllium were born from either the evolution or the death throes of stars.

Eventually, a star will populate an electron-degenerate core via fusion in its envelope, meaning the core will be supported against gravitational collapse by electron degeneracy pressure alone. If the final core is less than $1.4\,\mathrm{M}_\odot$, the Chandrasekhar limit, the star will remain electron degenerate and be called a white dwarf. However, if the star's total mass is sufficient for the envelope to populate a more massive core, electron degeneracy will be overcome, and the star will become a neutron star if neutron degeneracy is able to stop the in-fall and support the mass of the core, or else the star will collapse to a singularity, where all the mass is concentrated at a single point in space [4]. In this last case, any information contained in the star other than its mass, spin, and charge (for instance, all the data on the computers of the extremely heat-tolerant aliens living on the star) is erased from the universe.

3

This state of complete collapse is known as a black hole, because it has the amazing property that, within a certain distance from the singularity, even a photon lacks sufficient velocity to escape the gravitational pull of the singularity. The surface dividing the region where light can escape the singularity from the region where it is inexorably drawn to it by gravity is called the event horizon. Because light cannot escape, no information can be sent from inside the event horizon[1]. The region inside the event horizon is therefore causally disconnected, or in layman's terms, completely cut off from the rest of the universe.

Black holes are the most dense objects in the universe, and so, as we will detail in the next section, they create extreme curvature in spacetime. Furthermore, accelerating black holes will induce change in the surrounding spacetime curvature. Changes in the curvature of spacetime are referred to as gravitational waves. If two black holes are in orbit with one another, the mutual acceleration that they induce will cause extremely powerful emission of gravitational waves, resulting in a loss of orbital energy that drives the black holes toward coalescence. However, for a brief time before the black holes collide, the energy carried by gravitational waves exceeds the energy carried by electromagnetic waves from all the stars in the universe.

---

[1]Throughout this work, we apply only classical theory. If one includes quantum effects, then the phenomena of Hawking radiation may or may not permit information to escape through the event horizon. There is, at present, still substantial disagreement in the community as to whether Hawking radiation can carry information.

## 1.2 Gravitational Waves from Black Hole Binaries

The energy and momentum content of spacetime, which is described by the stress-energy tensor $T_{\mu\nu}$, can be related to the curvature of spacetime, described by the Einstein tensor $G_{\mu\nu}$, through the relation known, fittingly enough, as Einstein's equation,

$$G_{\mu\nu} = 8\pi T_{\mu\nu}\,, \tag{1.1}$$

where we have set $G = c = 1$. Black holes induce curvature on spacetime but, as is well known, the singularity itself has very little, if any, spatial extent. [2] Black holes result from the complete gravitational collapse of a star, and can therefore be viewed as pure curvature, or curvature in a vacuum. If we limit our focus to the general relativistic prediction for systems containing only black holes, then we can set $T_{\mu\nu}$ to zero, yielding

$$G_{\mu\nu} = 0, \tag{1.2}$$

which is the vacuum form of Einstein's equation. Furthermore, eq. (1.1) predicts that accelerating bodies will emit waves of curvature change through spacetime known as gravitational radiation. This prediction was verified indirectly by Robert Hulse and Joseph Taylor through the discovery and precision measurement of the pulsar PSR 1913+16 [7], which is part of a binary system orbiting a neutron star at a separation only a few times the distance from the Earth to the moon. Over the

---

[2]Technically, a singularity has exactly zero spatial extent by definition, but this assumes that all elementary particles are point particles. If elementary particles have some limiting minimal spatial extent, such as the Planck length $l_P$, then the "singularity" is likewise limited in the minimal size to which it can shrink.

intervening decades since its discovery, the binary's orbital period has decreased by about 75 millionths of a second each year, in excellent agreement with the expected value due to the emission of gravitational radiation as predicted by general relativity. This seemingly small change in the orbital period of the binary is an indicator of how extreme the dynamics of a system need to be in order to emit a non-negligible amount of radiation. Due to the relative weakness of the gravitational force, only the largest, most densely concentrated objects in our universe moving at substantial speeds will be a significant source of gravitational waves. The most extreme combination of dense objects and relativistic speeds in the universe will likely be black hole binaries (BHBs).

BHBs come together by different mechanisms depending on the scale in question. For example, for supermassive BHBs (SMBHBs), the individual holes are at the cores of galaxies, and so it is believed that both holes will sink to the bottom of the gravitational well of the combined galaxies when the galaxies collide, a process known as "dynamical friction" [8, 9, 10]. Then the forming BHB will most likely go through a "hardening" due to the ejection of third bodies which will bring the individual holes close enough for their mutual gravitational attraction to take over. The evolution of the BHB that has just been described is of course a very different process from that of stellar mass binaries, which would most likely be the result of the co-evolution of a binary star system. However, when all other matter has been expelled or consumed, and the BHB has reached the point of being gravitationally bound, the subsequent description of the evolution is same regardless of the scale of the holes involved. The BHB, be it SMBHB of $\mathcal{O}(10^5\,\mathrm{M_\odot} - 10^9\,\mathrm{M_\odot})$ or stellar of

$\mathcal{O}(1\,\mathrm{M}_\odot)$, will inspiral due to the emission of gravitational radiation of ever increasing strength until the two holes merge into one. The single remnant hole will not be quiescent, but rather will have perturbations that it will radiate away by emitting exponentially decaying sinusoidal waves, which are called "quasi-normal modes", and this phase is called the "ringdown" of the BHB. The merged remnant will have an inherent "spin" which reflects a conservation of the orbital angular momentum that the system had, as well as the spin of each hole, before the merger. However, as with electron spin, the spin momentum of a black hole does not refer to the rotation of an object as we would understand it in a classical sense, but rather it is a way to characterize the spacetime of a black hole. In fact, the mass, spin, and charge (which we will always assume is neutral) are the only pieces of information needed to fully characterize a black hole spacetime [11]. Furthermore, the dynamics of the inspiral, merger, and ringdown of a BHB, and the gravitational radiation that it produces, is scale-invariant, which means that a solution for a given pair of masses and spins on the individual holes can be scaled to find a solution for any total system mass (the relative sizes of the masses and spins is fixed by the initial solution).

When the binary reaches the late stages, it has cleared out any intervening gas, so the whole process only involves the curvature content of spacetime, since the black holes are pure curvature and the gravitational radiation is strictly a change in the background curvature. Therefore, the inspiral, merger, and ringdown of a BHB must satisfy the vacuum form of Einstein's equation (eq. (1.2)). This deceptively simple equation can be exceedingly difficult to solve, and while there are some exact solutions known, such as the Schwarzschild solution for a single, nonspinning,

7

chargeless black hole, the "two body problem" of two gravitationally bound black holes has no exact solution, and must be solved using either numerical or perturbative methods.

In this work we will use results from one perturbative method, the post-Newtonian (PN) expansion of various observable quantities, which is an expansion in $\frac{v}{c}$, where $v$ is the relative speed of the black holes. Specifically, two PN methods will be applied for calculating waveform phasing: the adiabatic Taylor-expanded 3.5PN, and a non-adiabatic pseudo-4PN effective one body (p4EOB) model. Adiabatic in this case means that the holes move together via changes in orbital frequency alone, so that the binary follows a sequence of quasi-circular orbits, whereas non-adiabatic evolutions admit an explicit radial velocity degree of freedom in the evolution equations. and the p4EOB, and EOB methods in general, involves the mapping of the binary system to the Hamiltonian of a deformed Schwarzschild spacetime representing the reduced system [12] (see Appendix A). These expansions are valid when the black holes are widely separated and slowly moving, but become less accurate as the dynamics become too relativistic for the expansion to be valid (i. e. $v \ll c$ is no longer the case approaching merger). Therefore, a solution in full general relativity is needed to know the correct dynamics for the late inspiral and merger-ringdown. This solution must be found numerically. However, until 2003, attempts to do so met with nothing but failure. In the ensuing years, with the experimental gravitational wave astronomers forging ahead toward their design sensitivities, and the possibility, if not probability, of a detection, the need for these numerical solutions was never more evident.

## 1.3   Gravitational Wave Detectors

The earliest efforts to detect the minute perturbations of spacetime involved the use of resonant bar detectors, and were pioneered by Joseph Weber of the University of Maryland. Weber used massive aluminum cylinders and attempted to isolate them from environmental noise and observe any oscillations from an incident gravitational wave. In 1969, Weber announced that he had made several coincident detections [13]. Ultimately, no other investigators could confirm Weber's detections, and the consensus conclusion has been that Weber failed to have a true detection, although Weber himself was not in agreement with this consensus. The race to confirm or refute Weber's findings had the effect of galvanizing a gravitational wave experimentation community where none had previously existed [14], so the unfortunate end result did have a very positive side effect.

More sensitive bar detectors have subsequently been built, but the biggest leap in gravitational wave sensitivity has come with ground-based laser interferometers, such as LIGO, VIRGO, GEO, and TAMA, which provide a lower noise floor than bar detectors and do so over a much broader frequency range, operating from the tens of Hz through the several kHz range. LIGO, an NSF-supported project, is the most sensitive of these detectors, consisting of three distinct interferometers. Two of the interferometers are collocated and occupy the same tunnel at Hanford, WA, one being two kilometers long and the other four kilometers long. The third interferometer is four kilometers long and is located in Livingston, LA. Over a period of nearly two years, LIGO scientists conducted their "S5" run, gathering a year's

worth of triple-coincident data ending in October of 2007. As of the writing of this work, no detections have been announced, although the data is still being thoroughly analyzed. An upgrade to LIGO, Advanced LIGO, is on track to begin construction in 2008, and will go online in 2013. In the intervening time before Advanced LIGO comes online, an "S6" run, referred to as "Enhanced LIGO", will use some of the improved technologies and is planned to run from mid-2009 until late 2010. It is expected to show a factor $\sim 2$ improvement over the S5 sensitivity level. The Advanced LIGO upgrade will provide roughly an order of magnitude improvement to the gravitational wave strain sensitivity across most of its bandwidth. Whereas a detection with LIGO and the other current generation ground-based interferometers is hoped for but not expected, a detection, or rather detections, are expected for Advanced LIGO.

Another type of laser interferometer, the space-based LISA, is a joint NASA-ESA project which is currently scheduled to launch in 2018. LISA will have 5 million kilometer long arms, and so will be sensitive to much longer wavelength gravitational waves compared to the ground-based interferometers. The LISA band will span from $\mathcal{O}(10^{-5}\,\text{Hz})$ to the tenths of Hz. LISA's strain sensitivity in this band will be comparable to that of LIGO in LIGO's band, but the riches of sources with large strain spectral densities in the LISA band makes LISA a unique instrument. LIGO can only hope to find a signal at low SNR, and it is very unlikely that parameters could be extracted with any significant degree of certainty. The best that could reasonably be hoped for would be that the class of object might be identified using data filtering and template matching. Advanced LIGO has the hope of finding

sources with substantial SNR and with sufficient information content that parameters might be extracted. However, the noise power relative to the signal will still be non-negligible, so uncertainties will be substantial. Advanced LIGO is therefore an incremental step, giving us a hazy picture of some relatively nearby pieces of the gravitational wave sky. LISA will measure signals with such high SNR, and extract parameters with such high fidelity, that it can really be viewed as an observatory for the gravitational wave sky. Gravitational wave astronomers will analyze the incoming data, which will contain thousands of sources from the entire sky at any given time, including galactic white dwarf binaries, extreme mass ratio inspirals (known as EMRIs), and SMBHBs. They must use the data from LISA's various channels to separate the signals, then characterize the signal of interest, trying to specify its intrinsic parameters, such as total mass or spin, and its extrinsic parameters, such as distance and sky position. BHB coalescences are of particular interest for all the interferometers. Stellar mass binaries would coalesce in LIGO and Advanced LIGO's band, whereas SMBHBs coalesce in LISA's band. BHB coalescence is a likely candidate for the first detection by a ground-based interferometer, and will be one of the richest sources of science for LISA.

## 1.4   Numerical Relativity and Data Analysis

The field of black hole numerical relativity (hereafter, numerical relativity, or NR) is an attempt to solve eq. (1.2) numerically for the case of a BHB. The aforementioned detectors will produce a data stream that contains noise in all cases

and, in the cases of LISA, as many as thousands of other signals, so nearly exact knowledge of the waveform will be needed *a priori* in order to be able to successfully identify that signal in the data stream. Until recently, the challenge of numerical relativity was viewed as formidable, and many thought the problems facing numerical relativists were impossible to solve. These problems included the inherent difficulty of evolving a singularity in a numerically stable way, finding the correct gauge choice to accommodate the treatment of the singularities and accurately and stably follow the correct dynamics, and the problem of computational resources needed to simulate the full spectrum of length scales involved, from the horizon diameters of the individual holes up to the wavelength of the emitted radiation. However, beginning with the first successful orbit of two black holes in a simulation [15], through the first demonstration of the merger and subsequent ringdown [16], and culminating with the discovery of a robust method for stably evolving any number of orbits through the merger and ringdown [17, 18], all the obstacles along the road to success seemed to fall away. Several groups went from being unable to evolve more than a fraction of an orbit to evolving several orbits through the merger and ringdown and moving beyond the case of two inspiraling equal mass Schwarzschild black holes to cases of unequal mass and even Kerr (i.e. spinning) binaries in the space of just over a year. We will discuss the methodology, and the breakthrough known as the "moving puncture" method which is most responsible for the explosion of results in the field in the last several years, in detail in Chapter 2.

Now that stable evolution is no longer an issue in numerical relativity for those using the aforementioned technique, the focus has turned to exploring other

regions of parameter space, as was mentioned, and investigating the accuracy of the waveforms that are extracted from our simulations. In Chapter 3, we will thoroughly investigate the accuracy of the equal mass, nonspinning runs which spanned seven orbits for the highest resolution case. Some of this work has already been published by the author and coauthors in [1] and [2], but much of it is previously unpublished material.

Once we have established the accuracy of the waveforms, we can apply them to problems of detection and measurement for the detectors that were discussed earlier. Specifically, in Chapter 4, we use the waveforms to calculate quantities of interest, such as the signal-to-noise ratios (SNRs) assuming matched filtering for both LIGO and Advanced LIGO as well as LISA. Since LISA will be able to not only detect, but measure parameters, we present the results of an investigation into how much the merger phase would contribute to the parameter measurement accuracy for the equal mass, nonspinning case. We compare this to an estimate of the theoretical errors of a few parameters, to assess the degree to which we can rely on the parameter values that have been calculated. We will conclude in Chapter 5 with a brief synopsis of key points, and a look ahead at the work that needs to be done to make the field of Gravitational Wave Astronomy a reality.

## 1.5   Notation and Conventions

Unless otherwise noted, geometrized units are used throughout the text, meaning $G = c = 1$. This allows us to express any observable quantities in terms of the

total mass of the system, M, given that all observables for the black hole binary solution scale invariantly with the total system mass. Two convenient conversion factors are $1\,\mathrm{M}_\odot = 5 \times 10^{-6}\,\mathrm{s}$ for time measurements and $1\,\mathrm{M}_\odot = 1.5\,\mathrm{km}$ for distance measurements.

Greek indices are used to indicate a spacetime quantity, whereas Roman indices indicate a purely spatial quantity. This distinction is also made using a preceding parenthetically enclosed superscript of "3" for spatial quantities and "4" for space-time quantities, for example, $^{(3)}g_{ij}$ and $^{(4)}g_{\mu\nu}$, respectively. However, in cases such as the examples given where the index convention specifies the dimensionality of the object and no further identification is needed, the superscript is omitted. Also, the metric signature will be (-1, 1, 1, 1).

Partial derivatives are interchangeably denoted $\frac{\partial f}{\partial x^i}$, $\partial_i f$, or $f_{,i}$, depending on convenience and clarity. Covariant derivatives are denoted $D_i f$ or $f_{;i}$.

Chapter 2

Modeling Black Hole Binary Coalescence

2.1   Recent Advances in Numerical Relativity

Numerical relativists attempt to accurately solve Einstein's equation in a computer. The field of numerical relativity has progressed at a rapid pace in the last several years. This period began with the first simulation of a complete orbit of two black holes [15]. This was followed by the successful evolution of a binary through its final orbit, merger, and ringdown [16, 17, 18]. A major breakthrough, called the "moving puncture" method [17, 18], was developed contemporaneously and independently at NASA Goddard and at the University of Texas at Brownsville, and is one of the technologies that has made these rapid advancements possible.

In the fixed puncture prescription [19], the black hole singularity is split into singular and nonsingular pieces, with the singular piece being handled analytically and not evolved. Because the coordinates of the punctures are fixed, the coordinate system becomes distorted as the binary is evolved, eventually causing the code to crash. However, by choosing appropriate gauge conditions, it was found that the singular part of the puncture could be evolved along with the nonsingular part, thus avoiding coordinate distortion. This breakthrough, referred to as the "moving puncture" method, opened the door for several groups to begin running long-lasting, stable simulations. Continued improvements in the gauge, the numerical accuracy,

and the initial data used to start the simulations have led to more orbits before merger [20, 21, 22], and the work presented here includes several equal mass, non-spinning waveforms which span roughly seven orbits prior to merger. This length, as we will demonstrate, is both necessary and sufficient for performing, among other analyses, the first validation of the post-Newtonian (PN) approximation using a numerical waveform [1].

## 2.2  Methodology

### 2.2.1  ADM Equations

The most popular approach to solving Einstein's equation (1.2) for BHBs, referred to as the "3+1" or "ADM" decomposition, entails dividing the 4 dimensional spacetime into 3 dimensional slices, with the slices spanning the subspace of the spatial dimensions of the original 4 dimensional space, but not the time dimension. The slices propagate along the time direction, so that time is represented in the sequential evolution of the purely spatial slices (see Fig. 2.1). Once we have values for the relevant fields on a single slice (which is referred to hereafter as initial data), we can evolve the fields to subsequent slices. The usual 4-metric, $g_{\mu\nu}$, is therefore more conveniently represented as three separate quantities. The 3-metric, $g_{ij}$, can obviously be calculated for each slice separately. Two new quantities, called the lapse $\alpha$ and the shift $\beta_i$, determine how much proper time and distance, respectively, should be spanned between successive slices for each grid point (see Fig. 2.2). The
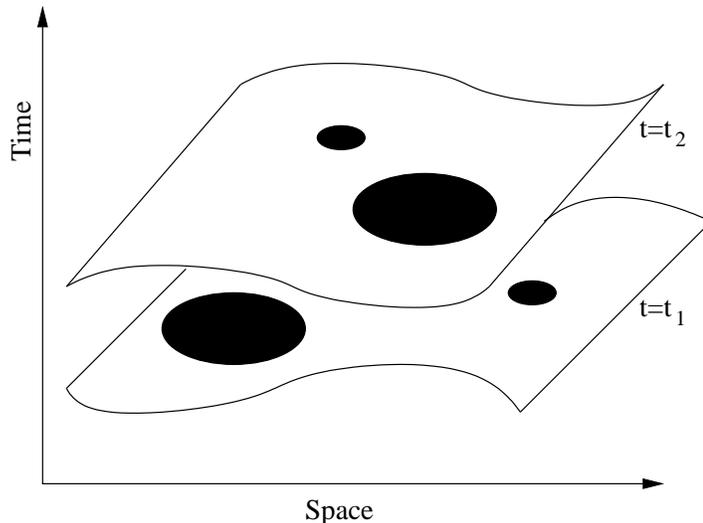
Figure 2.1: Pictorial representation of the division of 4 dimensional spacetime into 3 dimensional slices [23], with the slices spanning the subspace of the spatial dimensions, and time being represented in the sequential evolution of the purely spatial slices.

lapse and shift are related to elements of $g_{\mu\nu}$ in a straightforward way, namely

$$\alpha = (-^{(4)}g^{00})^{-1/2} \tag{2.1}$$

$$\beta_i =^{(4)} g_{0i}. \tag{2.2}$$

Therefore, we see that the lapse, shift, and spatial 3-metric can be related to the full 4-metric,

$$g_{\mu\nu} = \begin{pmatrix} \beta_k\beta^k - \alpha^2 & \beta_j \\ \beta_i & g_{ij} \end{pmatrix} \tag{2.3}$$

The slices are constrained to satisfy a set of evolution equations and a set of conservation equations, known as the "ADM" equations. These equations have

17
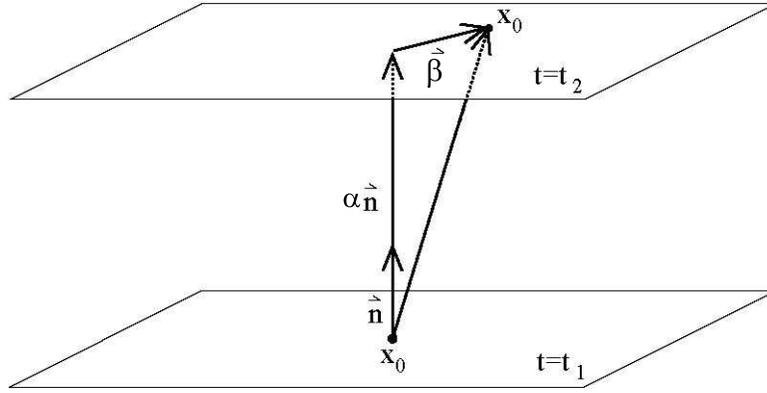
Figure 2.2: Illustration of the lapse, $\alpha$, and the shift, $\beta_i$ (or $\vec{\beta}$ in vector notation) [23], driving the evolution of a grid point from one slice to the next. $\alpha$ determines the amount of motion the grid point undergoes in the time direction, and $\beta_i$ determines the 3-dimensional spatial motion of the grid point.

been derived numerous times throughout the literature, in the original work [24], in review articles (see e. g. [25]), and in many Ph. D. theses. A full derivation is therefore not presented here. Instead, we begin with the final equations, explain their content briefly, and mention some modern improvements. The ADM equations

are [25]

$$R + K^2 - K_{ij}K^{ij} = 16\pi\rho \qquad \text{(Hamiltonian constraint)} \qquad (2.4)$$

$$D_j K^j{}_i - D_i K = 8\pi j_i \qquad \text{(momentum constraints)} \qquad (2.5)$$

$$\partial_t g_{ij} = -2\alpha K_{ij} + D_i\beta_j + D_j\beta^k \qquad \text{(evolution of spatial metric)} \qquad (2.6)$$

$$\partial_t K_{ij} = -D_i D_j \alpha + \alpha(R_{ij} - 2K_{ik}K^k{}_j + KK_{ij})$$

$$-8\pi\alpha(S_{ij} - \frac{1}{2}g_{ij}(S - \rho)) + \beta^k D_k K_{ij}$$

$$+K_{ik}D_j\beta^k + K_{kj}D_i\beta^k \qquad \text{(evolution of extrinsic curvature).} \qquad (2.7)$$

In this set of equations, $K_{ij}$ is called the extrinsic curvature, a quantity which is contained on the slice and preserves the curvature information that is lost from $G_{\mu\nu}$ by truncating to a 3 dimensional subspace, and $K$ is the trace of $K_{ij}$, i.e. $K = g^{ij}K_{ij}$. Both the Ricci curvature scalar, $R$, and the Ricci tensor, $R_{ij}$, come from contraction of the Riemann tensor, $R^d{}_{abc}$, i.e. $R = R^a{}_a$ and $R_{ij} = R^k{}_{ikj}$. The Riemann tensor, in turn, can be computed from

$$R^d{}_{abc} = \Gamma^d{}_{ac,b} - \Gamma^d{}_{bc,a} + \Gamma^e{}_{ac}\Gamma^d{}_{eb} - \Gamma^e{}_{bc}\Gamma^d{}_{ea}, \qquad (2.8)$$

where $\Gamma^a{}_{bc}$ are spatial connection coefficients that can be related to the spatial metric using

$$\Gamma^a{}_{bc} = \frac{1}{2}g^{ad}\left(g_{db,c} + g_{dc,b} - g_{bc,d}\right). \qquad (2.9)$$

$S_{ij}$ is the projection of the full stress-energy tensor on the spatial slice, $S_{ij} = g_{ik}g_{jl}T^{kl}$, with $S$ being the trace of $S_{ij}$, $S = g^{ij}S_{ij}$. $\rho$ and $j_a$ are also related to the stress-energy tensor, with $\rho \equiv n^a n^b T_{ab}$ being the total energy density measured by a normal observer, with the time-like unit normal vector $n_a = (-\alpha \quad 0 \quad 0 \quad 0)$

satisfying $^{(4)}g_{ab} = ^{(3)}g_{ab} + n_a n_b$, and $j_a \equiv -g^b{}_a n^c T_{bc}$. However, since we are interested in vacuum solutions eq. (1.2), the stress-energy projection terms and terms related to stress-energy elements can be ignored entirely. Therefore, for our purposes the constraint and evolution equations can be rewritten as

$$R + K^2 - K_{ij}K^{ij} = 0 \qquad \text{(Hamiltonian constraint)} \qquad (2.10)$$

$$D_j K^j{}_i - D_i K = 0 \qquad \text{(momentum constraint)} \qquad (2.11)$$

$$\partial_t g_{ij} = -2\alpha K_{ij} + D_i \beta_j + D_j \beta^k \qquad \text{(evolution of spatial metric)} \qquad (2.12)$$

$$\partial_t K_{ij} = -D_i D_j \alpha + \alpha(R_{ij} - 2K_{ik}K^k{}_j + KK_{ij}) + \beta^k D_k K_{ij}$$

$$+ K_{ik}D_j \beta^k + K_{kj}D_i \beta^k \qquad \text{(evolution of extrinsic curvature)}. \qquad (2.13)$$

As one might expect, not everything is determined by eq. (1.2). In fact, the gauge freedom inherent in any GR problem means that we are free to specify the lapse and shift however we wish in order to achieve the best solution. The first successful orbit of a BHB [15] was achieved using the so-called "fixed puncture" prescription, which will be discussed in section 2.3, but which, among other things, fixed the shift at the punctures to be zero, i.e. the punctures did not move through the grid. This meant that the grid had to wind itself around the fixed punctures, so to speak, in order to simulate how the real spacetime would evolve. This winding caused large errors to accumulate and eventually made the simulation unstable. This problem was later circumvented when the "moving puncture" prescription was found, which allowed a nonzero $\beta_i$ at the punctures, and has proven to be very stable.

## 2.2.2  BSSN Formulation and Further Improvements

In practice, the exact ADM formulation proved not to be ideal for long term stable evolution. A variation containing additional auxiliary evolution terms, known as the BSSN formulation, has been implemented with great success. The evolution equations, when expressed in the appropriate form as a matrix equation, will have a set of eigenvalues and eigenvectors which indicate the presence of "modes". The modes may travel at or below light speed without incident. However, superluminal modes will permit error from inside the horizon to escape via coupling to the modes, and zero speed modes will permit an accumulation of error on the grid which does not advect away. The virtue of the BSSN formulation is that all the constraint-violating modes travel at the speed of light, which is not the case for the ADM equations [26]. In addition, while the ADM system is known to only be weakly hyperbolic, the BSSN formulation has been shown to be strongly hyperbolic, meaning that no modes will grow at greater than an exponential rate, and the system is well-posed [27, 28].

The BSSN equations are obtained from the ADM equations by the following substitutions

$$g_{ij} \rightarrow e^{4\phi}\tilde{g}_{ij} \tag{2.14}$$

$$K_{ij} \rightarrow e^{4\phi}\left(\tilde{A}_{ij} + \frac{1}{3}\tilde{g}_{ij}K\right), \tag{2.15}$$

where $\tilde{g} = \det \tilde{g}_{ij} = 1$ and $\tilde{A}^i_{\ i} = 0$. Three additional variables ($\tilde{\Gamma}^i$, $\phi$, and $K$) are

also introduced. The BSSN variables obey the following evolution equations [29]

$$\partial_0 \tilde{g}_{ij} = -2\alpha \tilde{A}_{ij} \tag{2.16}$$

$$\partial_0 \phi = -\frac{1}{6}\alpha K \tag{2.17}$$

$$\partial_0 \tilde{A}_{ij} = e^{-4\phi} \left( -D_i D_j \alpha + \alpha R_{ij} \right)_{TF} +$$

$$\alpha \left( K \tilde{A}_{ij} - 2\tilde{A}_{ik} \tilde{A}^k_j \right) \tag{2.18}$$

$$\partial_0 K = -D^i D_i \alpha + \alpha \left( \tilde{A}_{ij} \tilde{A}^{ij} + \frac{1}{3}K^2 \right) \tag{2.19}$$

$$\partial_t \tilde{\Gamma}^i = \tilde{g}^{jk} \partial_j \partial_k \beta^i + \frac{1}{3}\tilde{g}^{ij} \partial_j \partial_k \beta^k + \beta^j \partial_j \tilde{\Gamma}^i -$$

$$\tilde{\Gamma}^j \partial_j \beta^i + \frac{2}{3}\tilde{\Gamma}^i \partial_j \beta^j - 2\tilde{A}^{ij} \partial_j \alpha +$$

$$2\alpha \left( \tilde{\Gamma}^i_{jk} \tilde{A}^{jk} + 6\tilde{A}^{ij} \partial_j \phi - \frac{2}{3}\tilde{g}^{ij} \partial_j K \right), \tag{2.20}$$

where $TF$ indicates that only the trace-free part is used and $\partial_0 \equiv \partial_t - \mathcal{L}_\beta$, where $\mathcal{L}_\beta$ is the Lie derivative with respect to the shift vector, $\beta^i$, which are given by eqs. (3.15)-(3.17) of [30]. Also, $R_{ij} = \tilde{R}_{ij} + R^\phi_{ij}$ is given by

$$R^\phi_{ij} = -2\tilde{D}_i \tilde{D}_j \phi - 2\tilde{g}_{ij} \tilde{D}^k \tilde{D}_k \phi + 4\tilde{D}_i \phi \tilde{D}_j \phi -$$

$$4\tilde{g}_{ij} \tilde{D}^k \phi \tilde{D}_k \phi \tag{2.21}$$

$$\tilde{R}_{ij} = -\frac{1}{2}\tilde{g}^{lm} \partial_l \partial_m \tilde{g}_{ij} + \tilde{g}_{k(i} \partial_{j)} \tilde{\Gamma}^k + \tilde{\Gamma}^k \tilde{\Gamma}_{(ij)k} +$$

$$\tilde{g}^{lm} \left( 2\tilde{\Gamma}^k_{l(i} \tilde{\Gamma}_{j)km} + \tilde{\Gamma}^k_{im} \tilde{\Gamma}_{klj} \right). \tag{2.22}$$

$\tilde{\Gamma}^i$ is replaced by $-\partial_j \tilde{g}^{ij}$ in eqs. (2.16) - (2.22) wherever it is not differentiated. The Hamiltonian and momentum constraints in the BSSN formulation take the form

$$\mathcal{H} = R - \tilde{A}_{ij} \tilde{A}^{ij} + \frac{2}{3}K^2 = 0 \tag{2.23}$$

$$e^{4\phi} \mathcal{M}^i = \partial_j \tilde{A}^{ij} + \tilde{\Gamma}^i_{jk} \tilde{A}^{jk} + 6\tilde{A}^{ij} \partial_j \phi - \frac{2}{3}\tilde{\gamma}^{ij} \partial_j K = 0. \tag{2.24}$$

Further improvement beyond the standard BSSN equations have proven necessary in order to achieve the results presented in this work. Namely, we alter the standard BSSN system through the addition of dissipation terms as in [31] and constraint-damping terms as in [32] in order to ensure robust stability. The dissipation terms are proportional to a power of the resolution in the finest grid, and so vanish in the continuum limit, but have the effect of smoothing over high frequency errors. The constraint-damping terms, as the name implies, incorporate the constraint equations into the evolution equations as identities in order to actively damp constraint violations as the system evolves from slice to slice. Using the constraints as identities can resolve issues related to the superluminal or zero speed modes that were mentioned earlier by changing the matrix form of the evolution equation.

### 2.2.3   Extraction of Gravitational Radiation

In order to analyze the gravitational radiation being emitted by the BHBs, we must first choose a null tetrad for decomposition of the gravitational radiation content. This radiation content is contained in the Weyl tensor, $C_{abcd}$, so the tetrad will ideally be able to separate the radiation part of the Weyl tensor from the non-radiation content. For this discussion, [1] we choose the following tetrad: given $\hat{\tau}$, the time-like unit vector normal to a given hypersurface (i.e. a vector normal to a slice such as the ones shown in Fig. 2.1) and $\hat{r}$, the radial unit vector, and using spherical

---

[1]This discussion necessarily follows the discussion in [33] quite closely, since we are describing the same procedure.

coordinates, we can form the tetrad

$$\vec{\ell} \equiv \frac{1}{\sqrt{2}}(\hat{\tau} + \hat{r}) \tag{2.25}$$

$$\vec{n} \equiv \frac{1}{\sqrt{2}}(\hat{\tau} - \hat{r}) \tag{2.26}$$

$$\vec{m} \equiv \frac{1}{\sqrt{2}}(\hat{\theta} + i\hat{\varphi}) \tag{2.27}$$

$$\vec{m}^* \equiv \frac{1}{\sqrt{2}}(\hat{\theta} - i\hat{\varphi}), \tag{2.28}$$

where $*$ denotes complex conjugation. In addition to having zero length, the tetrad satisfies the orthogonality relations

$$\vec{\ell} \cdot \vec{n} = -\vec{m} \cdot \vec{m}^* = -1 \tag{2.29}$$

$$\vec{\ell} \cdot \vec{m} = \vec{\ell} \cdot \vec{m}^* = \vec{n} \cdot \vec{m} = \vec{n} \cdot \vec{m}^* = 0. \tag{2.30}$$

In terms of this tetrad, the complex Weyl scalar $\psi_4$ is given by

$$\psi_4 \equiv C_{abcd}\, n^a (m^b)^* n^c (m^d)^*. \tag{2.31}$$

There are 5 Weyl scalars in total, denoted $\psi_0$ through $\psi_4$, but only $\psi_4$ is of interest for our purpose here because it contains the outbound radiation content of the system [34].

$\psi_4$ can be related to the gravitational radiation in the following way: first, we note that in the transverse-traceless (TT) gauge (see Chap. 35 in Ref. [11]),

$$\frac{1}{4}(\ddot{h}_{\hat{\theta}\hat{\theta}}^{TT} - \ddot{h}_{\hat{\varphi}\hat{\varphi}}^{TT}) = -R_{\hat{\tau}\hat{\theta}\hat{\tau}\hat{\theta}} = -R_{\hat{\tau}\hat{\varphi}\hat{r}\hat{\varphi}} = -R_{\hat{r}\hat{\theta}\hat{r}\hat{\theta}}$$

$$= R_{\hat{\tau}\hat{\varphi}\hat{\tau}\hat{\varphi}} = R_{\hat{\tau}\hat{\theta}\hat{r}\hat{\theta}} = R_{\hat{r}\hat{\varphi}\hat{r}\hat{\varphi}}, \tag{2.32a}$$

$$\frac{1}{2}\ddot{h}_{\hat{\theta}\hat{\varphi}}^{TT} = -R_{\hat{\tau}\hat{\theta}\hat{\tau}\hat{\varphi}} = -R_{\hat{r}\hat{\theta}\hat{r}\hat{\varphi}} = R_{\hat{\tau}\hat{\theta}\hat{r}\hat{\varphi}} = R_{\hat{r}\hat{\theta}\hat{\tau}\hat{\varphi}}. \tag{2.32b}$$

24

We can then set the $h_+$ and $h_\times$ polarizations of the radiation to

$$\ddot{h}_+ = \frac{1}{2}(\ddot{h}^{TT}_{\hat{\theta}\hat{\theta}} - \ddot{h}^{TT}_{\hat{\varphi}\hat{\varphi}}), \qquad (2.33a)$$

$$\ddot{h}_\times = \ddot{h}^{TT}_{\hat{\theta}\hat{\varphi}}. \qquad (2.33b)$$

Finally, we can utilize the equality of the Riemann and Weyl tensors in vacuum,

$$R_{abcd} = C_{abcd} \ (G_{\mu\nu} = 0), \qquad (2.34)$$

to yield a relationship between $\psi_4$ and the radiation,

$$\psi_4 = -(\ddot{h}_+ - i\ddot{h}_\times). \qquad (2.35)$$

One subtle point, however, is that the tetrad used for this explanation is not the one used for the simulations in this paper. The tetrad actually used to generate $\psi_4$ for all the investigations carried out in this work is referred to as the quasi-Kinnersley tetrad, which has the property of minimizing the mixing of other $\psi_n$'s with $\psi_4$ (in other words, mixing non-radiation and/or inbound radiation content into the $\psi_4$ scalar, which ideally consists of outbound radiation only), particularly in cases with spin [35, 36]. The distinction is not critical for this work other than to point out that eq. (2.35) is revised to read

$$\psi_4^{quasi-Kinnersly} = -\frac{1}{2}(\ddot{h}_+ - i\ddot{h}_\times). \qquad (2.36)$$

The final step in analyzing the gravitational radiation using $\psi_4$ is to decompose it into harmonic components. This is a useful way to gain insight into the physical processes at work, as some processes may excite specific modes, and therefore can most effectively be analyzed by eliminating the other modes. For instance,

quadrupole radiation emits at twice the orbital frequency, and so will be constrained to the $l = 2$, $m = \pm 2$ harmonic modes. Since two factors of $\vec{m}^*$ appear in the definition of $\psi_4$ in eq. (2.31), and each carries a spin-weight of $-1$, we decompose $\psi_4$ in terms of spin-weight $-2$ spherical harmonics $_{-2}Y_{\ell m}(\theta, \varphi)$ [37], given by [38]:

$$_{\pm 2}Y_{\ell m}(\theta, \varphi) = \left[ \frac{(\ell - 2)!}{(\ell + 2)!} \right]^{1/2} \left[ \alpha^{\pm}_{(\ell m)}(\theta) \, Y_{\ell m}(\theta, \varphi) + \beta^{\pm}_{(\ell m)}(\theta) \, Y_{(\ell - 1)m}(\theta, \varphi) \right], \quad (2.37)$$

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

$$\alpha^{\pm}_{(\ell m)}(\theta) = \frac{2m^2 - \ell(\ell + 1)}{\sin^2 \theta} \mp 2m(\ell - 1) \frac{\cot \theta}{\sin \theta} + \ell(\ell - 1)\cot^2 \theta, \quad (2.38\text{a})$$

$$\beta^{\pm}_{(\ell m)}(\theta) = 2 \left[ \frac{2\ell + 1}{2\ell - 1}(\ell^2 - m^2) \right]^{1/2} \left( \pm \frac{m}{\sin^2 \theta} + \frac{\cot \theta}{\sin \theta} \right). \quad (2.38\text{b})$$

We can now decompose the dimensionless Weyl scalar $Mr\psi_4$, yielding

$$Mr\psi_4(t, \vec{r}) = \sum_{\ell = 2}^{\infty} \sum_{m = -\ell}^{\ell} {_{-2}}C_{\ell m}(t) \, _{-2}Y_{\ell m}(\theta, \varphi), \quad (2.39)$$

where $M$ is the total system mass, and $r$ is the radial distance to the binary center of mass. Note that this decomposition assumes a flat space background, and so the extraction surface of the radiation must be adequately far from the coalescing BHB to not be affected significantly by its gravitational potential (e.g. $r \gg M$).

### 2.2.4 Implementation and Convergence Testing

Because we are working with a discretely sampled mesh, rather than a continuum solution, we need to monitor the behavior of our results to ensure that they converge toward the continuum limit solution at the rate we expect when we change the mesh. In other words, if our code employs all fourth order accurate methods,

meaning that the leading error term scales as the grid spacing to the fourth power, but our results are either under-convergent (i.e. consistent with lower order methods, in this case first through third), they are over-convergent (i.e. consistent with higher order methods than the ones actually applied in the code), or they do not converge, then we know there is an uncontrolled source of error that is dominating our simulations.

For the simulations in this work, time integration was performed with a fourth-order Runge-Kutta algorithm, and spatial derivatives were calculated using fourth-order-accurate finite-differencing stencils. For the outer boundary we employed a second-order-accurate Sommerfeld condition, which is an outgoing wave equation given by [39]

$$(\partial_r + \partial_t) \begin{pmatrix} g_{ij} - \delta_{ij} \\ K_{ij} \end{pmatrix} \doteq 0 . \tag{2.40}$$

The boundary is pushed to $|x_i| = 1536M$ to keep reflections far from the source. Adaptive Mesh Refinement (AMR), which adapts the resolution of the mesh on a cell-by-cell basis to meet the accuracy requirements of the simulation, was implemented via the software package PARAMESH [40, 41], and interpolation between refinement regions was fifth-order-accurate. Note that we use AMR only to resolve the sources, and the mesh will progressively become coarser far away from the sources (see Fig. 2.3). Although the radiation which reaches the outer boundary during the course of the simulation, with wavelengths of $\gtrsim 100M$, will not be well resolved in this lowest refinement region of grid-spacing $h = 32M$ (in the highest resolution run), reflections from there are causally disconnected from our extraction

radii at $R \leq 100M$. We do not use AMR to follow the radiation with a fine mesh; instead we require only that the fixed mesh resolution in the region of the extraction surfaces be sufficient to resolve the waves there. For example the extraction surface at $R = 60M$, in our highest resolution simulation, spans regions with grid spacing $h = 1M$ and $h = 2M$.

Finally, to ensure that we are achieving reasonable results from our simulation, in addition to satisfying the constraints, we test the rate of convergence of observable quantities such as the waveform phase or the time of coalescence. Since the aforementioned runs used fourth-order accurate evolution and second-order accurate initial data, one would expect the output to certainly be no higher than fourth order convergent and no lower than second order. The order of convergence will depend on the dominant source of error, and the order of the method which was responsible for generating that error. For instance, our data has been processed by an initial data calculator, boundary conditions at the outer boundary and refinement interfaces, the wave extraction algorithm, and the evolution equations. Due to technical constraints, it is not always possible to use the same order methods for all the relevant processes, but we generally expect, and have consistently observed, that the error from the evolution error dominates.

To test the convergence, we take the observable quantity $\Theta$ at three different resolutions: $h_f$, $A h_f$, and $B h_f$, where $h_f$ is the resolution in the finest grid and $A$ and $B$ are constants, with both greater than 1 and $B > A$. We then form the
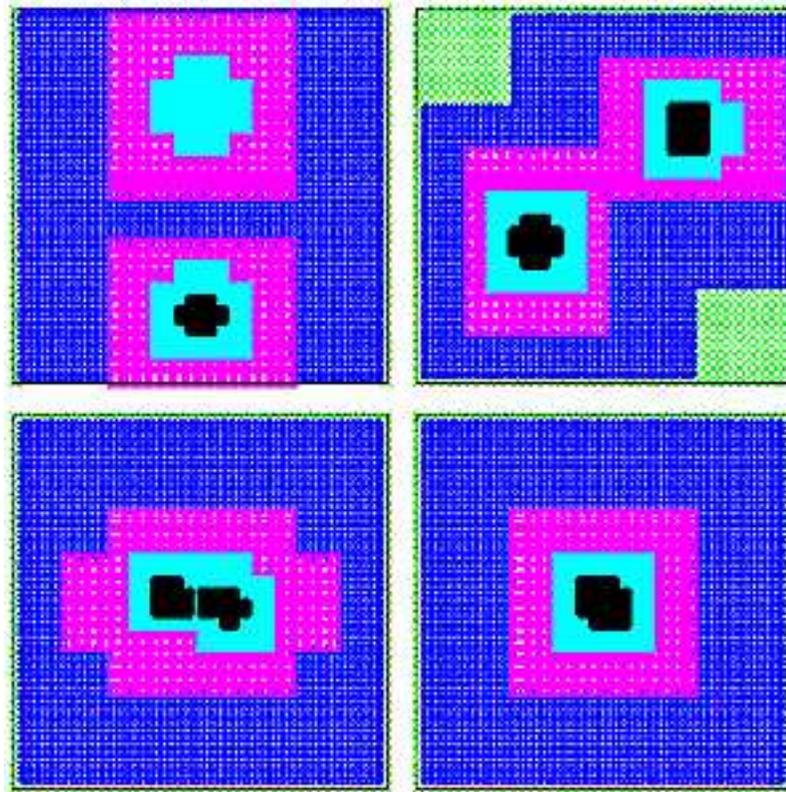
Figure 2.3: Four slices of the computational grid, demonstrating the increasing resolution as the grid moves closer to the puncture, and the adaptive nature of the grid, as the refinement levels follow the puncture, starting with the earliest slice (top right) and evolving counter-clockwise. The green, blue, pink, cyan, and black represent, in that order, ever-increasing levels of resolution. Note that the finest level for one hole disappears on the second slice, only to reappear on the third. This is not a critical issue, but is not ideal, and is an example of the added challenges that implementing AMR presents.

following differences:

$$\Delta\Theta_A = \Theta\left(A\,h_f\right) - \Theta\left(h_f\right) = \mathcal{O}\left[\left(A\,h_f\right)^n - h_f^n\right] \tag{2.41}$$

$$\Delta\Theta_B = \Theta\left(B\,h_f\right) - \Theta\left(A\,h_f\right) = \mathcal{O}\left[\left(B\,h_f\right)^n - \left(A\,h_f\right)^n\right], \tag{2.42}$$

with the result that we can now divide eq. (2.41) by eq. (2.42) to find the scaling factor for a given rate of convergence, $n$, namely,

$$\frac{\Delta\Theta_A}{\Delta\Theta_B} = \frac{\left(A\,h_f\right)^n - h_f^n}{\left(B\,h_f\right)^n - \left(A\,h_f\right)^n} \tag{2.43}$$

Therefore, in practice, we can calculate the left hand side of eq. (2.43) using eqs. (2.41), (2.42) and our data at three resolutions, then find the value for $n$ for which the right hand side most agrees. In the next Chapter, we will apply this technique to our data to assess the order of convergence and thus qualify the data for the rest of this work.

## 2.3   Black holes as punctures

The moving puncture that has proven to be, above all other advancements in numerical relativity in the last several years, the achievement that is most centrally responsible for the sudden explosion of new results which continues to come out of the field since its discovery. For decades, it was thought that a puncture couldn't possibly be permitted to move on a computational grid in a stable way, and any work on moving black holes involved excising the region inside the apparent horizon from the computational grid. Nevertheless, puncture solutions, albeit fixed punctures, have a lengthy history.

Punctures are topological holes punched in the spatial slice. The first puncture solution [42] consisted of a wormhole which connected two different regions on the same sheet, such that the topology of the spacetime was left unchanged (see Fig. 2.4a). However, this setup would be undesirable for evolutions as it admits the possibility of emitted gravitational waves from one spatial region influencing the dynamics of the puncture via interaction at the other spatial region at some later time. A more suitable setup, referred to as Brill-Lindquist initial data, consists of the wormhole connecting one spacetime sheet to a completely separate sheet, thereby changing the topology of the system overall (see Fig. 2.4c). Also, the solution is for an arbitrary number of holes. Brandt-Brügmann punctures [19] are similar to Brill-Lindquist in that they also change the overall topology, but in addition to having a solution for an arbitrary number of holes, the Brandt-Brügmann prescription allows boosts and spins to be placed on each hole separately rather than just on the whole system. The region beyond the throat of the wormhole is treated via a mapping, with further points being mapped to smaller radial components. This results in a large conformal factor as $r \to 0$, where the conformal factor,$\psi$, is given by

$$g_{ij} = \psi^4 \delta_{ij}, \tag{2.44}$$

with $\delta_{ij}$ being the 3-metric in flat space, and $g_{ij}$ is the 3-metric of the system. However, the system is asymptotically flat in both physical and mapped coordinates, since

$$K_{ij} = \psi^{-2} \hat{K}_{ij}, \tag{2.45}$$

where $\hat{K}_{ij}$ is the extrinsic curvature in flat space. For implementation, Brandt-

31
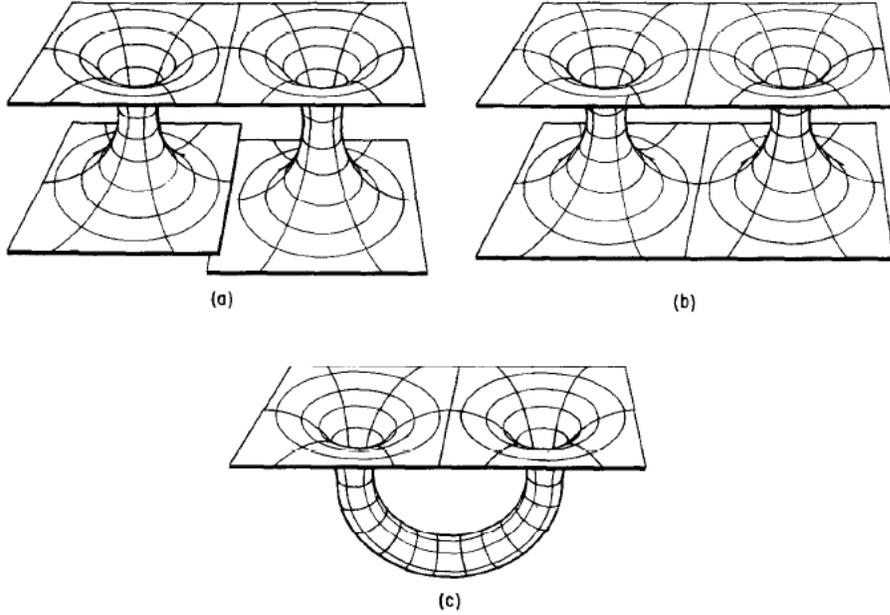
Figure 2.4: Reproduction of the three different binary punctures-on-a-sheet configurations from [43]. In the case of $c$, the binary is Misner's wormhole solution [42] connecting two asymptotically flat regions on the same sheet, whereas in $b$, the punctures lead to another shared sheet, and in $a$, the punctures each lead to their own sheet.

Brügmann provides an ansatz for the singular part of $\psi$, leaving the finite part to be solved for numerically.

For the initial data in this work, we used Brandt-Brügmann puncture data[19] generated by a second-order-accurate multigrid solver, `AMRMG` [44]. The puncture parameters were determined by the Tichy-Brügmann prescription for quasi-circular initial data [45], adjusted slightly, as informed by previous empirical experience, to reduce eccentricity. This prescription, as the name suggests, uses the PN approximation and searches for the correct orbit at a given separation which is instantaneously

circular, meaning $\dot{r} = 0$, by finding an approximate helical Killing vector. Our adjustment was simply to reduce the initial coordinate separation by roughly 2% while increasing the initial momentum of each puncture such that the product of the initial momentum with the initial coordinate separation remains constant.

The gauge condition that is used for all the results contained in this work is one approach to what has become known as the "moving puncture" method. Specifically, whereas other methods either have $\vec{\beta}(\vec{x} = \vec{x}_{punc}) = 0$, meaning the punctures are fixed in the grid, or else an excised region rather than a puncture is permitted to move across the grid, the "moving puncture" method takes advantage of the natural regularizing capacity of finite differencing in order to move punctures across a computational grid. [2] If the puncture is not placed on a grid point to start the evolution, the first time-stepping procedure effectively eliminates the singularity from the computational grid. With a properly chosen lapse and shift, the simulation can continue indefinitely without going unstable and without the computational grid being swallowed by a puncture. For the results contained in this work, the gauge choice is that suggested in [46], namely

$$\partial_t \alpha = -2\alpha K + \beta^j \partial_j \alpha \tag{2.46}$$

$$\partial_t \beta^i = \frac{3}{4} \tilde{\Gamma}^i + \beta^j \partial_j \beta^i - \eta \beta^i . \tag{2.47}$$

We use this gauge because it dispenses with the auxiliary variable $B^i$ that is present in all other gauge choices in the literature, and, more importantly, it eliminates the

[2] In addition to letting finite differencing regularize the punctures in the "moving puncture" method, another successful implementation of the method [17] involves analytically regularizing the punctures by performing a change of variables.

zero speed mode that would otherwise be contributed by $B^i$.

This method for evolving BHBs was truly a great discovery, and like many discoveries, the effectiveness of the method and a plethora of achievements made by applying the method preceded an understanding of why the moving puncture method worked. It was shown in [47] that for a single moving puncture, the region inside the event horizon evolves such that it ceases to be asymptotically flat at the puncture, but rather is cut off at a throat with a finite Schwarzschild radius. The region is still an infinite proper distance from the horizon, so the change to the puncture as it evolves is causally disconnected from the rest of the space, and the simulation should represent the same physics as it would if the puncture remained asymptotically flat. Then in [48], it was demonstrated that all current numerical codes under-resolved the punctures, but that this under-resolution, which is fully contained within the horizon, permits the system to evolve to a stationary solution, and in fact the puncture method would always be nonstationary in the limit of infinite resolution. The under-resolution is highly analogous to excision, allowing the system to evolve stably without an excessively large buildup of errors from a singular puncture. While errors from the under-resolved puncture may not be substantial enough to crash codes, they may, along with other sources of error like eccentric initial data and errors in the post-Newtonian waveform, cause our final template waveform to disagree with the true signal, thus limiting its utility. In the next Chapter, we investigate the quantity of this internal error, and so determine the level to which we can trust our waveforms.

# Chapter 3

## Determining the accuracy of waveform templates

In this Chapter, we present an analysis of the accuracy of our numerical waveforms. This constitutes an assessment of both the initial data accuracy as well as evolution accuracy. Since BHBs are expected to circularize their orbits through gravitational wave emission long before merger [49], minimal eccentricity is a requirement for an accurate simulation, and initial data is the primary factor in determining the eccentricity of a simulation. To assess errors in the evolution, simulations at multiple resolutions, and data extracted at multiple distances from the BHB, will need to be compared and analyzed. It is worth noting that the waveforms used in the analysis are the same ones presented in [1, 2]. As such, since the field is progressing at such a rapid pace, the waveforms no longer represent the most accurate available. Higher-order methods have yielded much improved results for the few cases available at the time of this writing. However, runs at multiple resolutions and with different mass ratios using the latest methods were not yet available to use in this analysis. Regardless, the intention of this Chapter is primarily to present the methodology of assessing waveform accuracy, and as a secondary point to give a snapshot of the accuracy level of state-of-the-art simulations as of a few months before the writing of this work.

## 3.1  Simulation Analysis

In order to apply our waveforms to data analysis, we must first establish their accuracy. Late-inspiral evolutions, covering more than a few orbits prior to ringdown, are more challenging than shorter merger-ringdown simulations. In this regime, frequency change, which will be important for our analysis, occurs much more slowly, so there is a greater demand for computational resources. In addition, better accuracy is required to control the accumulated phase error, which in turn constrains the numerical error that can be tolerated in the rate of energy loss. content of the system (radiation + individual horizon masses), and therefore should be conserved.

### 3.1.1  Comparing Waveforms

Using the moving puncture technique [17, 18, 46], with the gauge evolution given by eq. (2.46) and eq. (2.47), which originated from [46], implemented with fourth-order Runge-Kutta time integration, fourth-order accurate finite spatial differencing, and second-order-accurate initial data, we have simulated the evolution of a nonspinning equal-mass BHB starting at relatively wide separation, $\sim 1200M$ or $\sim 7$ orbits before the formation of a common event horizon. Here $M$ is the total mass the system would have had when the black holes were very far apart and before radiative losses became significant. We used fourth-order-accurate finite differencing techniques with AMR to resolve the dynamics near the black holes and in the wave propagation region. We carried out three runs using similar grid refinement

structures, but at different resolutions: low ($h_f = 3M/64$), medium ($h_f = 3M/80$) and high ($h_f = M/32$). Here, $h_f$ is the grid spacing in the regions with the highest resolution in each simulation, those being the regions around each black hole.

From each simulation we have measured the radiation in the form of the complex Weyl tensor component $\psi_4$ describe in Chapter 2. When calculating strain, the two complex integration constants are chosen to keep the strain close to oscillating about zero. For some applications we also examine waveforms in the form of the *strain rate* $v = \dot{h}_+ + i\dot{h}_\times$, the quantity from which radiation energy is directly obtained. To extract the waveform information from the simulation we define a series of coordinate spheres of different radii $R_{\text{ext}}/M \in \{40, 60, 100\}$ on which we measure spin-weighted spherical harmonic components of $\psi_4$ via a second-order algorithm described in [50, 51]. Decomposing $\psi_4$ into spin-weighted spherical harmonic components allows us to more clearly observe different physical processes which will radiate in particular characteristic modes, and is a natural decomposition for quasi-normal ringdown. In this work we focus exclusively on the $l = 2, m = 2$ component of the radiation, which mirrors the $l = 2, m = -2$ component, [1] when dealing with equal mass configurations (we do consider other modes for unequal mass cases). Other components are considerably smaller, containing $\sim 1\%$ as much energy [22].

---

[1]$l = 2, m = 2$ and $l = 2, m = -2$ components mirror each other by symmetry in all cases without spin. In cases with spin, spin oriented in the orbital plane will cause an asymmetry between the two components. Since the components point in opposite directions, the asymmetry means they no longer cancel each other out, and a net linear momentum will be imparted on the merging system (see e. g. [33] for a more detailed discussion).

Fig. 3.1 compares waveforms from our high-resolution simulation extracted on each of our three extraction spheres with the $R_{\text{ext}} = 40M$ and $R_{\text{ext}} = 60M$ waveforms shifted in time by the intervening propagation time derived based on a Schwarzschild black hole background. The generally good agreement for all three waveforms indicates that potentially worrying subtleties related to the relatively close distance of the extraction spheres, such as nonlinear propagation effects, or tetrad-specification sensitivity, do not seriously affect the waveforms. On the other hand, some small differences are evident.

For the early portion of the waveforms, shown in the inset, the results from the closest extraction sphere show slight differences in amplitude and phase, suggesting that $1/R_{\text{ext}}^2$ radiation details are not yet insignificant here; this is not surprising given that the extraction radius in this case is only $\sim 1/4$ of a wavelength. On the other hand, the waveform from the farthest extraction radius shows signs of dissipation for the later higher-frequency portion of the waveform. This is because the radiation has propagated significantly farther, through a lower-resolution region on the computational grid, by the time it is measured at $R_{\text{ext}} = 100M$. The resolution in most of the intermediate region is $h = 2M$, about six points per cycle for the ringdown radiation. For the rest of this work, we primarily employ the waveforms extracted at the intermediate distance $R_{\text{ext}} = 60M$, which is only weakly affected by either of the above short- and long-wavelength effects.

Fig. 3.2 shows the resulting gravitational waveform in the context of recent developments in black hole evolutions. The dashed (blue) curve gives the gravitational wave strain from the dominant $l = 2$, $m = 2$ spin-weighted spherical harmonic
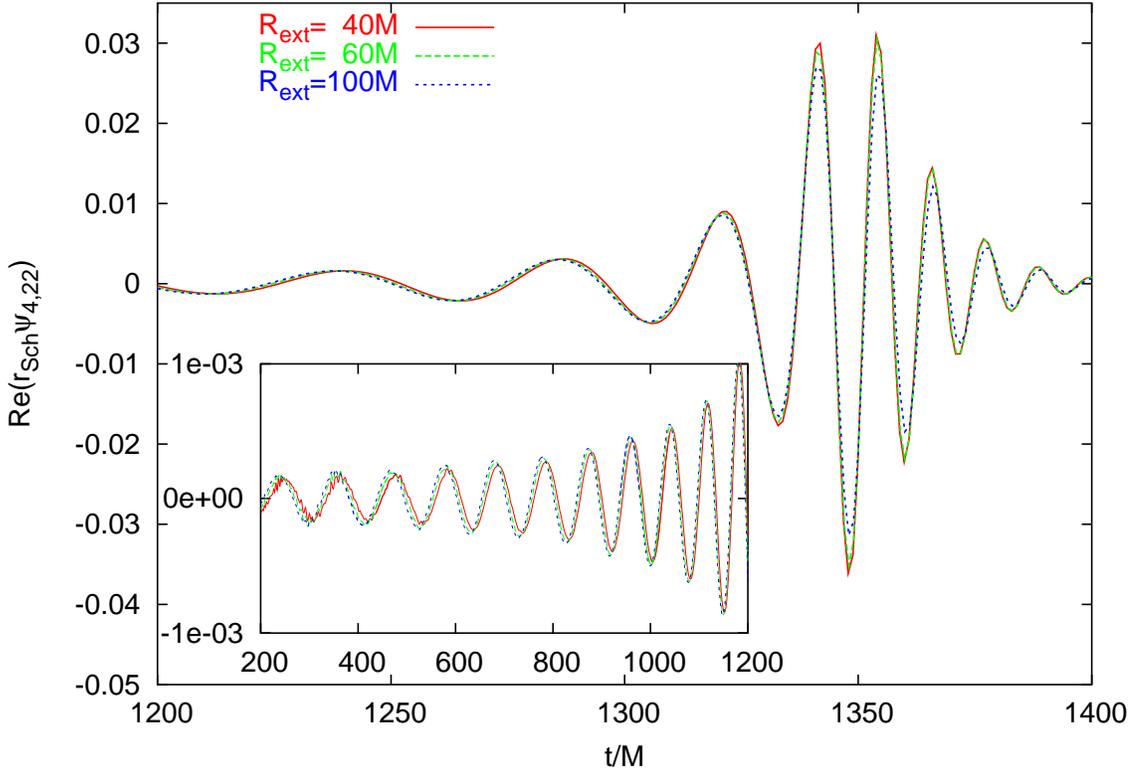
Figure 3.1: $\psi_4$ waveform calculated at three different extraction radii, and scaled by the approximate Schwarzschild areal radius. Times have been shifted to compensate for the differences in light travel time [11, 51], as appropriate to compare the $R_{ext} = 40M$ and $R_{ext} = 60M$ waveforms with the $R_{ext} = 100M$ waveform.

from our highest resolution simulation, extracted at $R_{\mathrm{ext}} = 40M$. This represents an observation made on the equatorial plane of the system, where only a single polarization component contributes to the measured strain. Time $t = 0$ is taken to be the moment of peak radiation amplitude as measured by $\psi_4$; the symbols along the time axis mark the points at which the system reaches the innermost stable

39
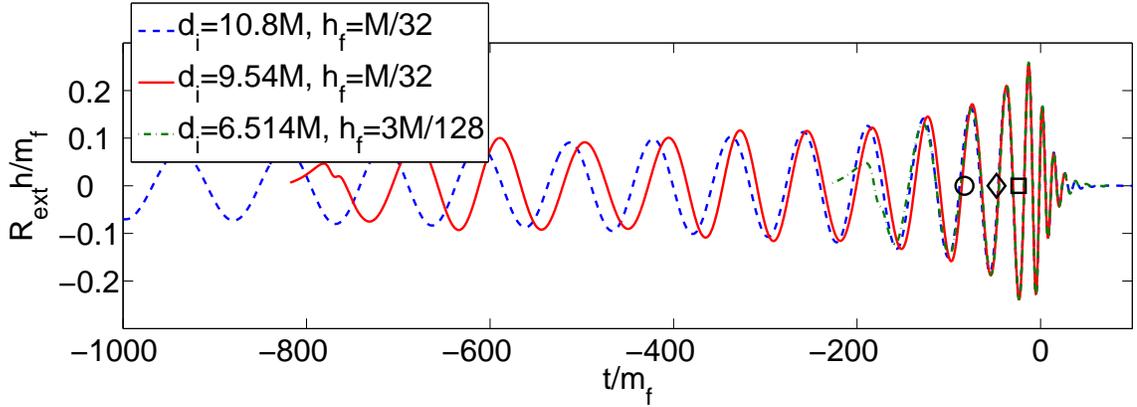
Figure 3.2:   Gravitational strain waveforms from the merger of equal-mass Schwarzschild black holes.  The late part of the merger ($t \gtrsim -50M$) is robustly determined and relatively easily calculable, while simulations of the late inspiral (early part of the waveform) are rapidly approaching the phasing accuracy required for observational applications [1]. The dashed blue line shows the more current numerical waveform that will be the focus of later analysis. The solid red line shows a comparison waveform from a run starting with the same initial data as R4 in Ref. [22] and the dash-dotted green curve shows the results from the highest resolution R1 run in Ref. [22]. All waveforms have been extracted at $R_\mathrm{ext} = 40M$ and shifted in time so that the moment of maximum $\psi_4$ radiation amplitude occurs at time $t = 0$. The initial coordinate distance between the punctures, $d_i$, is indicated in all cases.

circular orbit, or ISCO, calculated for a Schwarzschild black hole (circle), for the effective one body method (EOB) at 3PN order [52, 53] (diamond), and for the adiabatic PN method at 3PN [53, 54] (square). For comparison, we show two other

waveforms from earlier simulations carried out by this group; both were extracted at $R_{\text{ext}} = 40M$ and have been shifted in time and initial phase so that (as in [22]) the moment of peak $\psi_4$ radiation amplitude occurs at $t = 0$. As these different simulations may radiate different fractions of the initial mass, we choose the mass $m_f$ of the *post-merger* Kerr hole as the natural length scale for comparison. To within error tolerance, $m_f$ is the result of subtracting the radiation energy from the ADM mass, $M$. Because of radiative losses, $m_f \approx 0.95M$.

The solid (red) curve shows the results from a model that starts $\sim 800M$ before merger with the same initial data as run R4 in Ref. [22] but using a different gauge. Note that the gauge conditions used for the dashed (blue) curve and the solid (red) curve are equivalent to those given by case #8 in Ref. [46], while the gauge conditions used in Ref. [22] are equivalent to those given by case #3 in Ref. [46]. The dot-dash (green) curve shows a simulation that starts $\sim 200M$ before merger; this is the high resolution run R1 from Ref. [22]. All three waveforms agree to within $\sim 1\%$ for the merger-ringdown burst part of the waveform, starting near $t \sim -50M$.

Each simulation also contains a certain amount of noise which lasts for the first $\sim 100M$ of the simulation. This will be seen in the simulation presented throughout this work. The noise is primarily due to the initial gauge pulse which occurs as the gauge starts to evolve, and the initial data "junk", known as Bowen York radiation when Bowen York initial data is used, which is a result of the initial severe driving of the system by the evolution equations into the desired type of dynamics. For example, all currently-used initial data is conformally flat, meaning

the metric can be written in the form of eq. (2.44). However, a true BHB system cannot be represented with a flat metric by any transformation, so at the outset of the simulation, the description of the metric will have to change to represent meaningful physics.

### 3.1.2  Satisfying the Constraints

To verify the validity of the simulations, we look at the Hamiltonian and momentum constraints as previously discussed. For the most recent run shown in Fig. 3.2, we find the Hamiltonian constraint to be fourth order convergent in the regions leading up to the finest refinement region, but only 2.5 order convergent in the finest region, where it is likely dominated by errors from the puncture (see Fig. 3.3). For the momentum constraint, the convergence order leading up to the finest region is less clear, and may be 2.5 there as well as in the finest region (see Fig. 3.4). One possible explanation for the lower-than-expected convergence given the order of methods being implemented is that errors may be leaking from the puncture. These errors could be sufficiently small in the Hamiltonian constraint to diffuse once they reach lower refinement regions, but be larger in the momentum constraint. More investigation is needed to determine where the discrepancy lies.

### 3.1.3  Measuring and Mitigating Eccentricity

Astrophysically, BHBs in this relatively late stage of their evolution are expected to be traveling on nearly circular orbits, as any initial eccentricity would
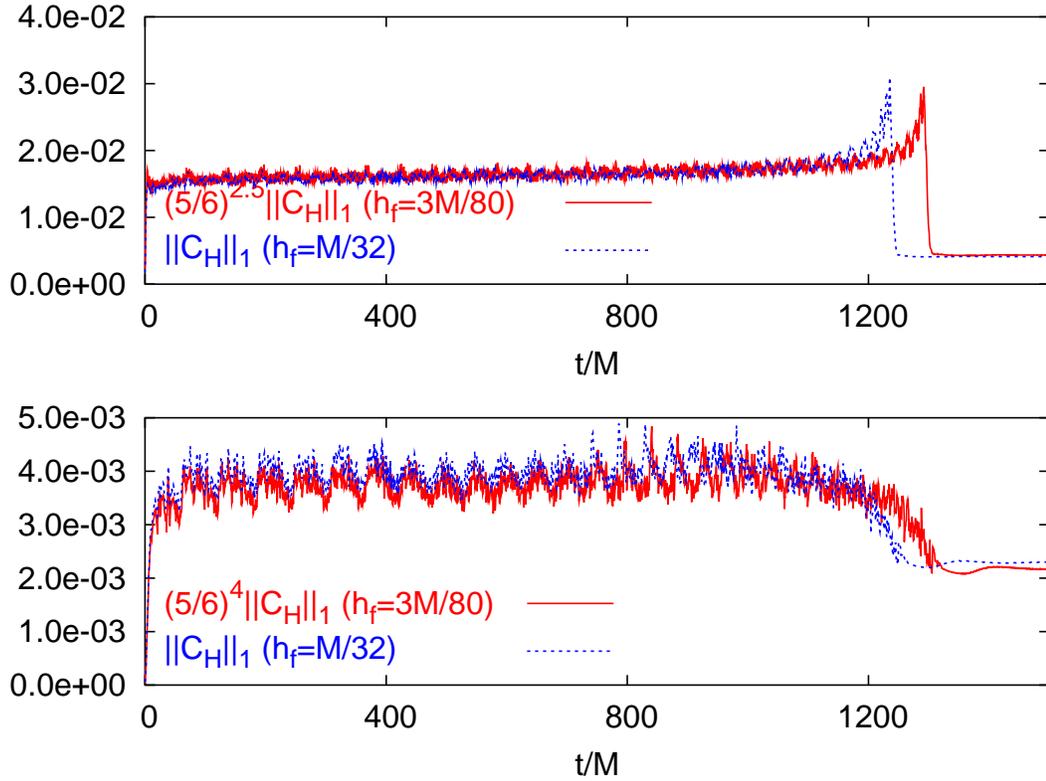
Figure 3.3: Convergence plot for the Hamiltonian constraint $C_H$. The top panel shows results from the finest grid and has been scaled so that for 2.5 order convergence the curves should superpose (see eq. (2.43)). The bottom panel shows results from the second finest grid and has been scaled so that for fourth order convergence the curves should superpose; the curves indeed appear to be fourth order convergent.

have been radiated away early in their evolution. Therefore, it is desirable to mitigate any eccentricity that ultimately is present in our simulations, in order to make the data as realistic as possible. We investigate several approaches for assessing th eccentric content and removing it, while always being careful not to introduce bias
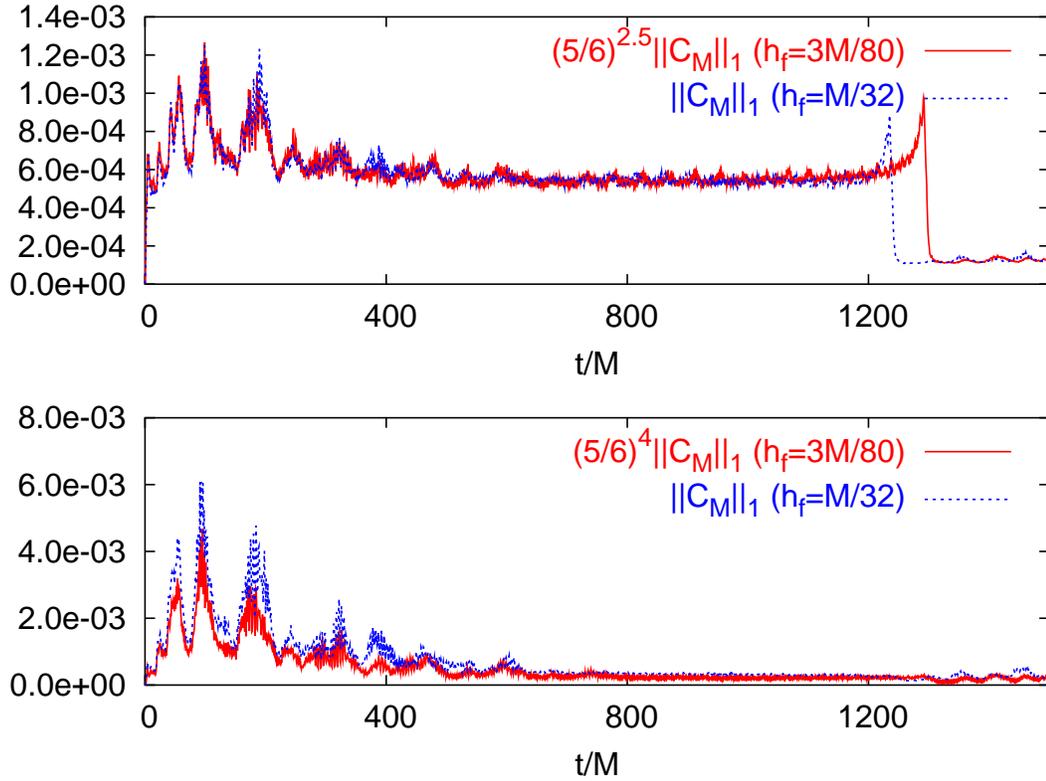
Figure 3.4: Convergence plot for the Momentum constraint $C_M$. The top panel shows results from the finest grid and has been scaled so that for 2.5 order convergence the curves should superpose (see eq. (2.43)). The bottom panel shows results from the second finest grid and has been scaled so that for fourth-order convergence the curves should superpose; the curves appear less than fourth-order convergent but better than second-order convergent.

to the underlying trend that has been predicted by the simulation.

## 3.1.3.1 Mitigation through Simple Fits

In our model, we start the black holes on nearly circular orbits with very small eccentricity $e < 0.01$, as defined below. The resulting black hole trajectories are shown in Fig. 3.5, where the tracks mark the paths of the punctures. Qualitatively, the tendency of the trajectories to not cross each other, but to form complementary spirals, is a visual indicator of the significantly reduced eccentricity compared to previous puncture tracks in the literature (e. g. in [2]).

The black hole separation as a function of time is shown in Fig. 3.6 for this model and the next shorter model from Fig. 3.2. The greater eccentricity of the previous model (solid curve) is clearly distinguished by the large oscillations around a monotonically decreasing trend. We also show all three resolutions of the more recent model. The slight deviations from the overall smooth trend give an indication of the small amount of eccentricity in these latter simulations. Note however the differences between these three resolutions, particularly in the total time between the start of the simulation and the merger. Although significant, the differences in merger time appear to converge at a rate consistent with fourth-order error.

To an excellent approximation, the $l = 2, m = 2$ harmonic of the radiation is polarized in the form expected for radiation generated by circular motion. The polar representation of the $l = 2$, $m = 2$ component of the complex waveform, $\psi_{4,22} = A_\psi(t)\exp(i\phi_\psi(t))$, is particularly natural for circularly polarized radiation, for which $A_\psi$ and $\omega = \partial\phi_\psi/\partial t$ vary only slowly. The angular polarization frequency $\omega$ then provides a meaningful instantaneous frequency obtained directly from the
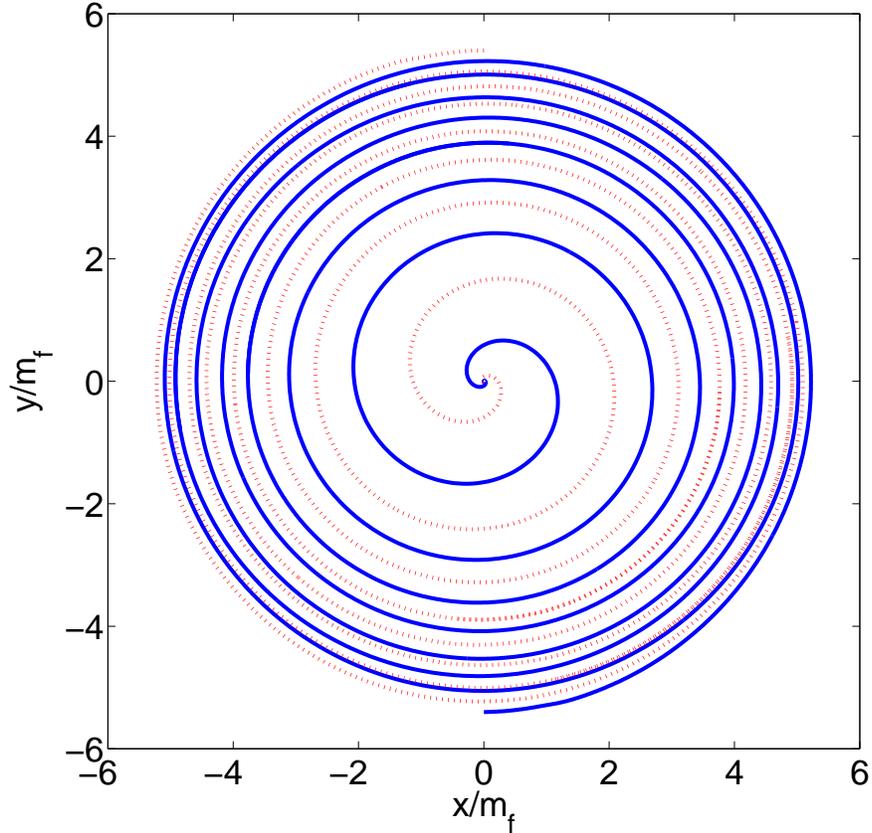
Figure 3.5: The trajectories of each of the binary system's black holes through $\sim 7$ revolutions before coalescence for our high resolution case are shown.

radiation, corresponding to twice the angular frequency of orbital motion when the black holes are still separate. Because the radiation is measured in the weak-field region of our simulations, where gauge dependence is minimal, this polarization frequency provides gauge-invariant information about the binary dynamics.

If the orbital motion is eccentric, this will leave an imprint in the radiation, causing a slight decrease in the polarization frequency of radiation generated near
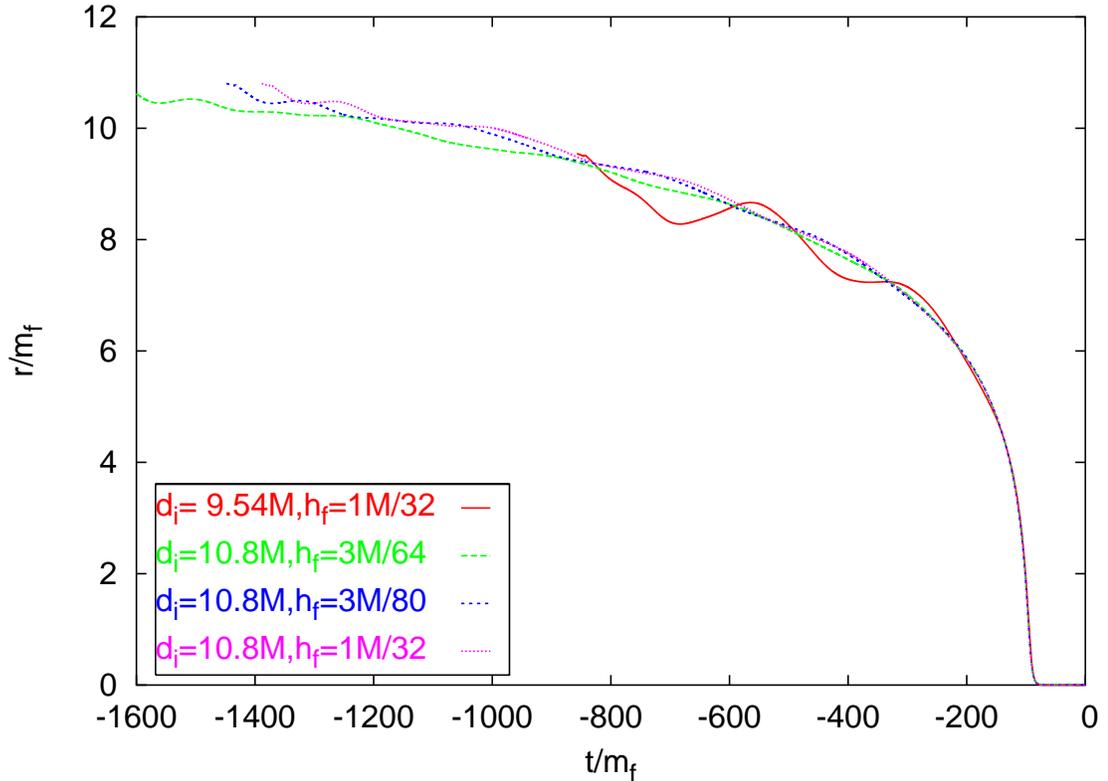
Figure 3.6: The coordinate separation between the puncture black holes is shown as a function of time. The solid line shows the results for the comparison run, which has relatively large eccentricity. The other curves show the three resolutions for our new simulations, all having noticeably less eccentricity. Note that equivalent gauge evolution equations were used in all four cases.

apocenter. We can recognize eccentric motion by identifying periodic deviations from a smooth monotonic trend in the time development of the polarization frequency $\omega(t)$.

Specifically, we looked at the polarization frequency $\omega(t) = \partial\phi/\partial t$ calculated

| Run | $AM$ | $\Omega_0 M$ | $\dot{\Omega} M^2$ | $\Phi_0$ | $e_0$ |
|---|---|---|---|---|---|
| $3M/64$ | $5 \times 10^{-4}$ | $0.017$ | $3 \times 10^{-6}$ | $1$ | $0.005$ |
| $3M/80$ | $7 \times 10^{-4}$ | $0.016$ | $4 \times 10^{-6}$ | $1$ | $0.007$ |
| $M/32$ | $8 \times 10^{-4}$ | $0.017$ | $3 \times 10^{-6}$ | $1$ | $0.008$ |

Table 3.1: Values resulting from eccentricity fitting. The magnitude of the eccentricity in these simulations (as given by $AM$ or $e_0$) appears to be at least second-order convergent.

from the strain rate $v(t) = V(t)\exp(i\phi(t))$. We see generally similar results whether we use strain, strain rate, or $\psi_4$ to define the frequency, but specifically show the strain rate because it comes out smoother than $\psi_4$ with respect to small waveform noise, but with far less noticeable detrending issues than the strain. We fit the time dependence of the frequency curves to a quartic trend, $\omega_c$, plus an eccentric modulation of the form $d\omega = A\sin(\Phi_0 + \Omega_0(t-t_0) + \dot{\Omega}(t-t_0)^2)$ where the quantities $A$, $\Phi_0$, $\Omega_0$ and $\dot{\Omega}$ are fitting constants and $t_0 = 61M$ is a time offset approximately accounting for the propagation time to $R_{\text{ext}} = 60M$. Ignoring the early part of the simulation where there are transient initial data effects, and the late part where the secular trend is very strong, we fit the curves over the time range $250M < t < 1000M$. We get similar results for eccentricity whether we apply high-frequency filtering to the waveform before fitting. The results of the fitting are summarized in Table 3.1.

Variations in the details of the fitting procedure, such as using strain instead of strain rate or smoothing high-frequency noise from $\omega(t)$, give results consistent to roughly the number of significant digits shown for $A$ and $\Omega_0$, though $\dot{\Omega}$ and $\Phi_0$ vary more significantly in some cases. The period of eccentric oscillations indicated by $\Omega_0$ is about 1.5 times the initial orbital period, decreasing slightly at a rate comparable to the rate at which the wave period grows. Allowing $A$ to evolve in time did not result in clearly improved fits.

From these fits we define eccentricity based on the effect, in Keplerian dynamics, of small eccentricity on orbital frequency, which should provide a good approximation in the adiabatic regime. Kepler's second law (conservation of angular momentum) implies that $L \propto r^2\omega$ is constant, from which it follows that the eccentric frequency deviation $d\omega/\omega$ is twice the eccentric radial deviation $dr/r$, suggesting the unitless definition of eccentricity $e \equiv A/2\omega_c$. Note that, to second order in $e$, this definition of eccentricity is also equivalent to that put forth in Ref.[55] in a post-Newtonian context. The constancy of $A$ in our fit is consistent with a linear decrease in eccentricity as frequency increases. The initial eccentricity $e_0$, obtained in this way, is also given in Table 3.1.

Even more interesting than the estimates for eccentricity provided this way are the residual curves $\omega_c(t) = \omega(t) - d\omega(t)$ of angular frequency vs. time with the eccentric part subtracted out. Note that the strictly sinusoidal nature of our fit to the eccentric modulations represented by $d\omega(t)$ implies that the underlying secular trend should be preserved. Fig. 3.7 shows the results for each of our three simulations together with the unmodified frequency $\omega(t)$ for the high-resolution case.

49
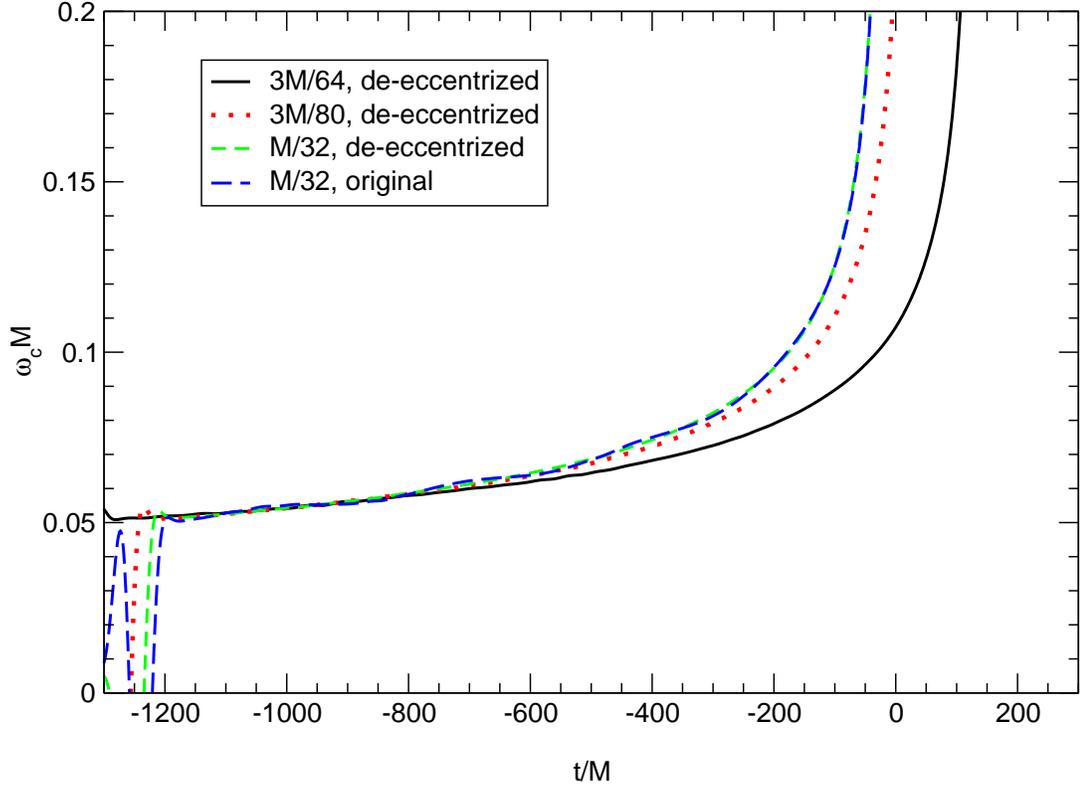
Figure 3.7: Angular frequencies with eccentricity removed using the polynomial trend, aligned such that the frequencies agree when the $h_f = M/32$ simulation is $1000M$ before the peak in $\psi_4$. Also shown is the frequency from the $h_f = M/32$ resolution before the eccentricity was removed.

The plotted curves have also been filtered to remove some high-frequency simulation noise present in the early part of the simulations. The smooth, now monotonic, trend in the curves for $\omega_c(t)$ is an indication that most of the wiggles evident in $\omega(t)$ were consistent with our model for eccentric modulations $d\omega(t)$, though the early part is affected by transient features related to the shortcomings of the initial data model.

We can take $\omega_c(t)$ as providing a record of the "hardening" process, as radiative losses bring the system through tightening orbits. The key effect of numerical simulation error evident in Fig. 3.7 is a slowing of the hardening rate at low resolutions, causing the final merger to be delayed. This delay appears to converge at fourth order in resolution. Viewing the eccentric modulation as a small perturbation on the dynamics of an optimally non-eccentric inspiral, we expect that these trends would also provide a good approximation for the frequency evolution of a merger simulation begun with optimally non-eccentric initial data. We therefore have a mechanism for effectively removing a substantial fraction of the modulation without affecting the underlying trend, and so can largely remove the eccentricity.

## 3.1.3.2 Improving Eccentricity Removal through Physically Motivated Fitting

We note, however, that one potential shortcoming of the aforementioned procedure is the possibility of mixing the trend and eccentricity components. For instance, a fourth order polynomial can have inflection points, meaning the polynomial could change its concavity due to the presence of an eccentric modulation. A more desirable trend would be one that is constrained to be monotonically increasing. To achieve this, we use the PN approximation. Specifically, we use the p4EOB model introduced in [12] (see Appendix A), which includes a pseudo-4PN term tuned to maximize the match with available numerical simulations, to approximate the circularized frequency evolution. We align this EOB frequency data with our numerical

51

frequency data by shifting the numerical data so that the peak numerical strain amplitude occurs at $t = 0$, and then shifting the EOB frequency data so that the peak in the frequency, which approximates the location of the light ring, also occurs at $t = 0$. We will hereafter refer to this type of alignment as *maquillage*, in following with [56]. We can then subtract the EOB data from the numerical data to isolate the eccentric modulation present in the extracted radiation.

Fig. 3.8 shows three different frequency predictions. Two of these have been discussed: the frequency as calculated from the extracted radiation, and the frequency as calculated using the p4EOB model. The third approach makes use of the physically intuitive evolution of the position of the puncture, which is obviously heavily dependent on the choice of gauge. The puncture position, $x^i$, is given by

$$\frac{dx^i}{dt} = -\beta^i(x^i).$$  (3.1)

Re-expressing the puncture position in polar coordinates, $r_{PT}\, e^{i\phi_{PT}}$, where $r_{PT} \equiv \sqrt{x^2 + y^2}$ and $\phi_{PT} \equiv \arctan \frac{y}{x}$, the instantaneous puncture frequency is defined as $\omega_{PT} \equiv \frac{d\phi_{PT}}{dt}$. This provides the third frequency prediction seen in Fig. 3.8. Differences among the predictions are shown in Fig. 3.9. The modulations due to eccentricity are particularly clear in the difference between the EOB and puncture track predictions in Fig. 3.9.

Somewhat surprisingly, the amplitude of the eccentric modulations will not be significantly damped in time via the emission of gravitational radiation until the strong field regime of the merger, so eccentricity in the initial data gets locked in until the final plunge. We can understand this behavior in the weak field by looking
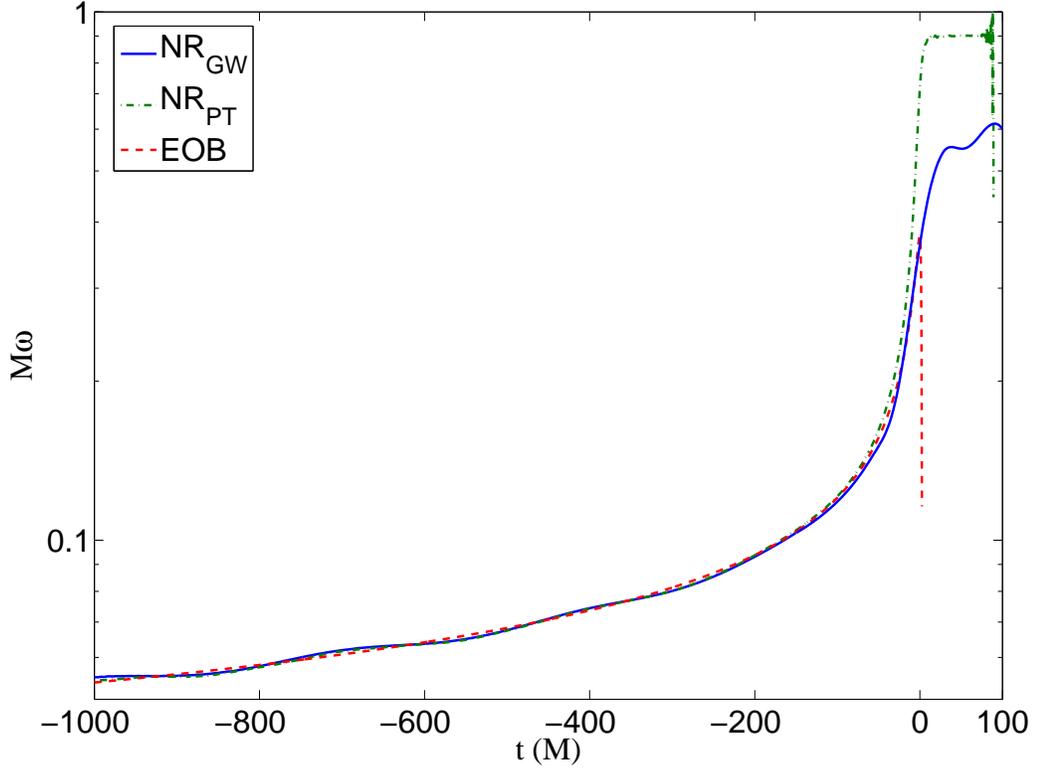
Figure 3.8: Gravitational wave frequency evolution as predicted by the p4EOB model (dashed red line), which uses a phenomenological 4PN term to fit the numerical relativity (NR) waveform, and by NR both from the puncture tracks (dash-dotted green line) and from the extracted radiation (solid blue line). The puncture orbital frequency is doubled for comparison to the extracted gravitational wave frequency prediction until the merger, when the two frequency measures decouple. The arbitrary time shift for the radiation and p4EOB frequencies through the *maquillage* procedure, and the puncture track frequency is set by first shifting it the same amount as the radiation, then shifting it forward by the coordinate distance to the extraction sphere to account for the radiation's travel time. This will be the standard puncture track *maquillage* procedure.
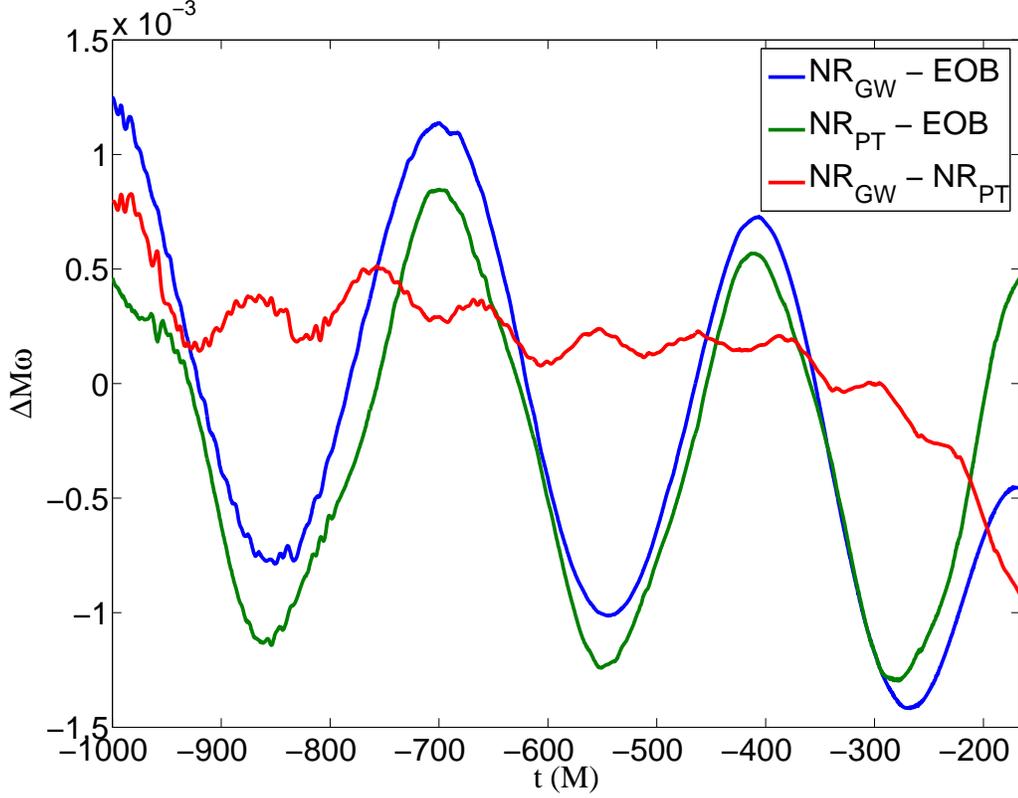
Figure 3.9: Differences in orbital frequency evolution as a function of time, demonstrating the expected behavior of common eccentricity between the puncture and radiation frequencies, and a smaller, underlying error when the eccentricity is canceled by subtracting the puncture track frequency from the radiation frequency.

at the leading order PN behavior. From Eq. (5.12) of [49], inverting $r(e)$ to lowest order in $e$, it is known that $e \propto r^{19/12}$. Using the leading order PN expressions for $r(t)$ and $\omega(t)$ (from e.g. [57] and references therein), we can string together the

relationship

$$\frac{\mathrm{d}e}{\mathrm{d}t} = \frac{\mathrm{d}e}{\mathrm{d}r}\frac{\mathrm{d}r}{\mathrm{d}t} \propto r^{-29/12} \propto \omega^{29/18} \propto (t_c - t)^{-87/144} \tag{3.2}$$

$$e(t) = \int \frac{\mathrm{d}e}{\mathrm{d}t}\mathrm{d}t \propto (t_c - t)^{57/144} \tag{3.3}$$

$$\Delta\omega_{ecc} \equiv e \cdot \omega \propto (t_c - t)^{57/144}(t_c - t)^{-3/8}$$

$$= (t_c - t)^{57/144}(t_c - t)^{-54/144} = (t_c - t)^{3/144} \approx \mathrm{const.} \tag{3.4}$$

Therefore, we expect that we will be able to accurately fit the eccentric modulation, $\Delta\omega_{ecc}$, using a sinusoid with a constant amplitude. Indeed, when we take the difference between our high resolution gravitational wave frequency and the EOB frequency (with the appropriate *maquillage* time shifts), and perform a least-squares fit assuming a power law trend for the systemic error accumulation and a constant amplitude eccentric modulation with only the leading PN order chirping behavior (where "chirp" here refers to the rate of increase in frequency of the signal), i.e. a fit of the form

$$\Delta\omega(t) = \Delta\omega_{trend}(t) + \Delta\omega_{ecc}(t) = a + bt^c + d\sin\left(f(-t)^{-3/8} + g\right), \tag{3.5}$$

we find excellent agreement with the assumption of constant amplitude (see Fig. 3.10). This result is simply an indication that the amplitude of eccentricity is radiated away on a timescale longer than that over which PN is applicable in our simulations. In other words, while it is well known that eccentricity radiates away, this has to either occur over many more cycles than we simulate in the weak field, or else in the strong field where we can't reasonably apply a PN fit.

With this additional method for approximating the eccentricity contribution

for a given frequency evolution, we can repeat the procedure demonstrated in Fig. 3.7 for the polynomial trend, and subtract the sinusoidal part of eq. (3.5) from our calculated frequency. The result, shown in Fig. 3.11, is indeed an exceedingly smooth curve. Because the p4EOB was used for the trend, rather than a polynomial, we can be more confident that the trend part of the fit will not be affected by the eccentric modulation, so that the sinusoidal part will be a more faithful representation of the eccentric content. Of course, using a trend like the p4EOB is a much more involved procedure, so the method that should be used depends on the level of fidelity required.

As described in Chapter 2, the punctures move freely through the computational grid when using the "moving puncture" method. One of the advantages that was mentioned was that, since the punctures weren't fixed, the grid did not have to evolve in a complex way, wrapping itself around the punctures in order to achieve the necessary relative motion. Another benefit of the well-behaved grid is that the coordinates are easily-interpretable. For instance, in order for the grid not to twist around the punctures during the evolution, the rays of constant $\phi$ in the evolving gauge will need to coincide with rays of constant $\phi$, modulo a constant, in another non-corotating gauge, such as the harmonic gauge used in most PN calculations, or the ADM gauge mentioned earlier in association with the $3 + 1$ decomposition. Incidentally, empirical evidence indicates that the gauge used in this work is very similar to the ADM gauge [58]. Assuming that the numerical gauge can be relied on to give phase coordinates for the puncture that are consistent with a more familiar gauge, such as ADM, we can interpret the puncture track values as being a reliable
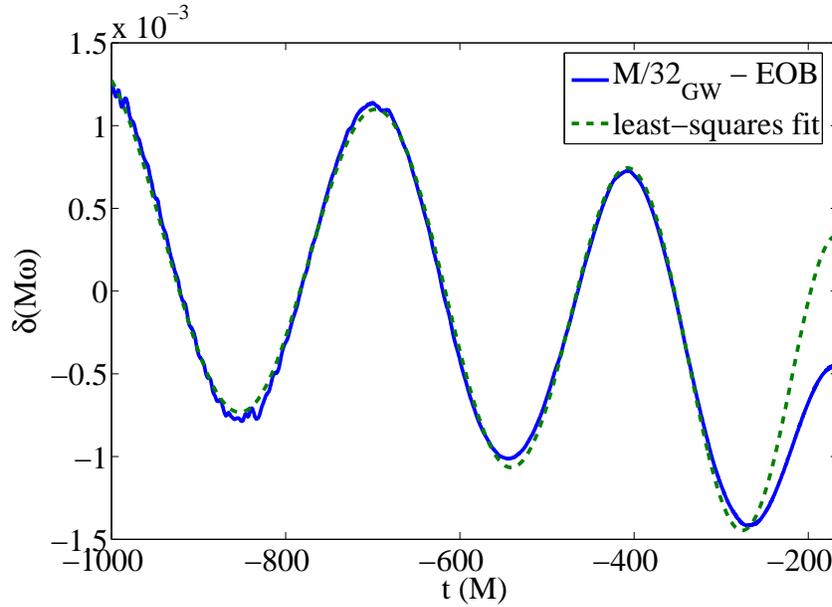
Figure 3.10: Differences in orbital frequency evolution as a function of time between the radiation and EOB prediction (solid blue), and a fit of the form given by eq. (3.5) (dashed green). The constant amplitude prediction for the eccentric modulation appears to hold up particularly well.

estimate of the orbital phase despite the fact that they are not invariant quantities. This assertion is borne out to a degree in the similarity of the eccentricity present in both predictions, as seen in Fig. 3.9.

### 3.1.4 Behavior of the Puncture Frequency through Merger

The puncture frequency has an interesting characteristic in that it levels off to a constant around merger which is independent of resolution and depends only on the
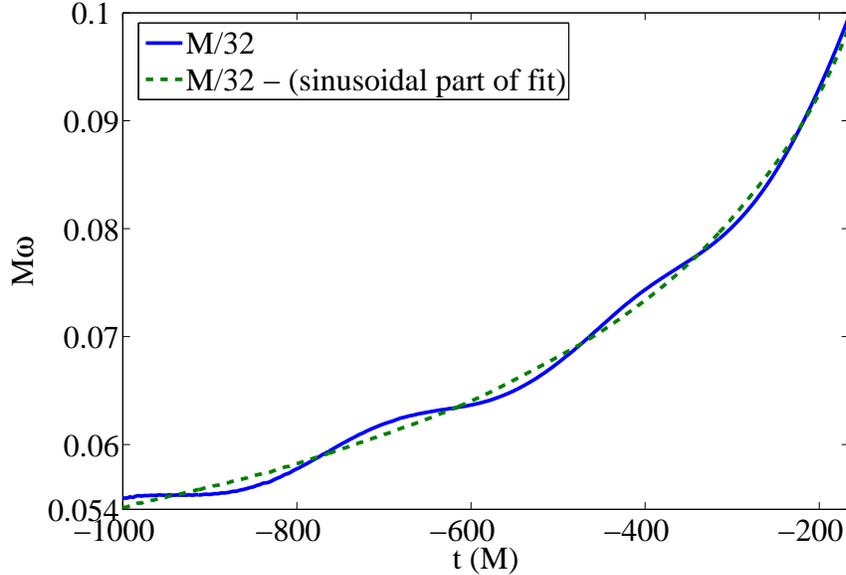
Figure 3.11: Raw frequency evolution of our highest resolution run (solid blue), as well as that same data with the eccentricity fit from Fig. 3.10 removed to leave a "circularized" frequency. This circularized frequency is likely to be of greater utility than the one based on the polynomial trend fit, since the trend in this case is a physically motivated and more accurate fit to the true trend, and therefore the fit to the modulation will ultimately be a truer representation of the correct eccentric modulation.

initial data, evolution equations, and choice of gauge. This resolution independence can be seen in Fig. 3.12. If we take $x_{PT}^i$ to be the puncture track location, and assume the center of mass is located at the origin, and $\Delta x^i$ a small deviation from the origin, then at very late times, we can assume $x_{PT}^i \approx \Delta x^i$. We can then Taylor expand the shift to ultimately find the puncture track frequency:
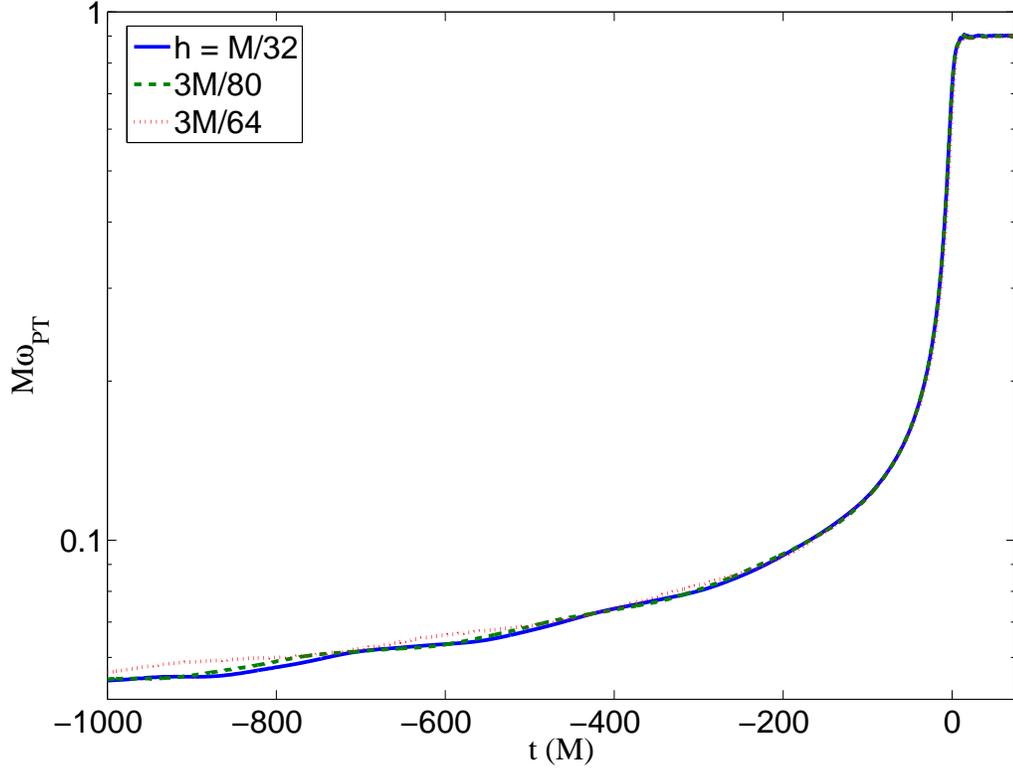
Figure 3.12: Puncture frequency evolutions at 3 different resolutions, demonstrating that they all plateau at a constant value after the merger has passed, and that that value is the same regardless of resolution.

$$\beta^i\left(x^i_{PT}\right) \approx \beta^i\left(0\right) + \beta^i\left(\Delta x^i\right) \approx \beta^i\left(0\right) + \Delta x^j \partial_j \beta^i\left(0\right) \tag{3.6}$$

$$\beta^i\left(\Delta x^i\right) = -\frac{d(\Delta x^i)}{dt} \approx \Delta x^j \partial_j \beta^i\left(0\right) \tag{3.7}$$

$$\omega_{PT} \equiv \frac{1}{\Delta x^j}\frac{d(\Delta x^i)}{dt} \approx -\partial_j \beta^i\left(0\right) \tag{3.8}$$

Therefore, the final plateau frequency from the puncture tracker depends on

59

the curl of the shift vector around the common center. The key to understanding the plateau behavior, then, is to see that eq. (3.8) becomes a steadily better approximation as the puncture track approaches the origin, and so should be valid after the merger, when a single Kerr hole has formed and is determining the evolution of the gauge, and hence the plateau value observed in the puncture frequency. Therefore, if we were to make any predictions related to the exact value of the plateaus that would be tested via simulation, we would need a pseudo-Kerr solution for punctures, which is not presently available, and it would need to be expressed in the specific numeric gauge that the simulation will have arrived at when it reaches the merger. Consequently, eq. (3.8) is not intended to be predictive (and, given the explanation that eq. (3.8) provides, any prediction will be of dubious scientific value), but simply serves as an explanation for the curious behavior of the puncture track through the merger.

## 3.2 Waveform Accuracy

Having used various waveforms at different resolutions and extraction radii in order to mitigate eccentricity, and having investigated the different kinds of error contribution in order to assess the leading factors which cause errors in our simulations, we are now ready to rigorously quantify the level of accuracy that we are able to achieve with our numerical simulations. Again, a necessary caveat is that the delay required in the preparation of this work means that the simulations presented are no longer state of the art. Nonetheless, the intent was not for the exact level of

accuracy to stand the test of time, but rather to present the methodology used to assess accuracy in our simulations.

### 3.2.1 Validating the Numerical Phase

In order to meaningfully apply our results to answer questions regarding data analysis, we must first determine our waveform accuracy, focusing primarily on the accuracy of waveform phasing information in the late-inspiral portion of our simulations. Over the course of many developmental simulations leading up to these results, we found that the early low-frequency part of the simulations is the most difficult to simulate accurately. This makes sense generally because the crucial dynamical details are in the slow loss of energy and angular momentum to the relatively fine, evolving structure of the spacetime curvature, which ultimately comprises the radiation. Timescales are also longer for this part of the dynamics, requiring high accuracy over a large number of computational iterations.

We get the most direct view of phasing information by examining the waveforms in polar decomposition. In Fig. 3.13 we show the polarization phase derived from the strain rate waveform. This time we have aligned the simulations in time from the beginning, as should be appropriate for simulations of the same initial configuration. Later in the simulations, as the frequency increases the phase increases more rapidly, and the timing differences among the simulations lead to large phase differences.

We quantitatively compare the phasing results in Fig. 3.14, showing the dif-
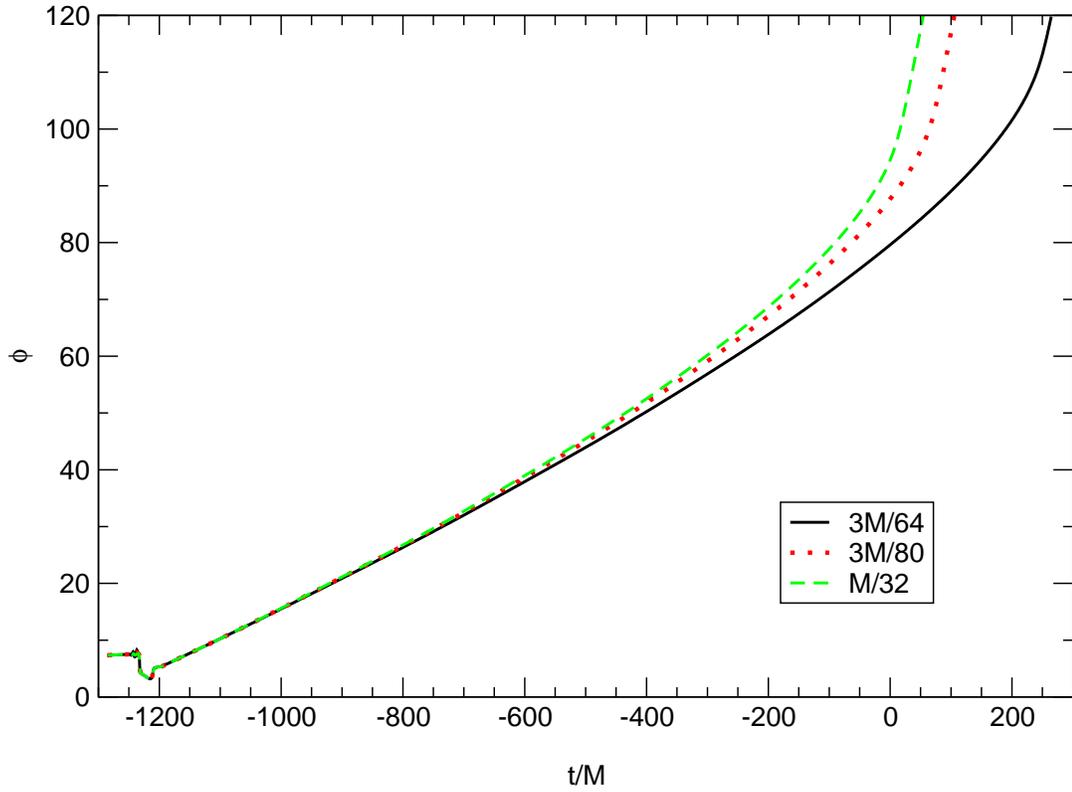
Figure 3.13: Strain rate phase. The high-resolution simulation goes through about 14 cycles before merger. The lower resolution simulations take longer to merge, and go through more cycles.

ference between the two higher-resolution runs compared with scaled versions of the difference between the two lower resolution runs. We can now clearly see that the phase errors (measured this way) grow strongly in time. The lower-resolution difference has been rescaled two ways, such that they would be expected to agree with the higher resolution difference for second- and fourth-order convergence, respectively. The comparison suggests phasing convergence somewhere between second-

and fourth-order over much of the run.

Since the phase error grows strongly in time, it is interesting to consider the effect of timing errors in these convergence comparisons. The last curve in Fig. 3.14 shows the high-medium difference shifted in time by $57M$; that is, the high resolution and medium resolution results have been differenced first, with their original time-dependence unaltered, and then the resulting difference has been shifted in time. $57M$ represents the approximate timing difference between the two higher resolution runs late in the simulation, as measured from the peaks in $\psi_4$. Viewed this way we see time-aligned phase differences taken from similar points in the physical evolution of the medium resolution run, as represented in the medium-low curve, and the high resolution run, as represented in the high-medium curve. As a result, the runs appear to converge at a faster rate, closer to fourth-order. This unconventionally time-shifted plot is not intended to serve as a rigorous assessment of convergence, but is intended to illustrate that the timing differences can have a significant impact on convergence estimates.

Above, we have compared our simulations made at different resolutions by comparing the waveforms at equal points in time, with time aligned either at the beginning of the simulation ($t = 0$ in the original run) or at the peak in $\psi_4$. Since errors cause the simulations to accelerate through their dynamical processes at somewhat different rates, such comparisons end up relating moments of quite different dynamics, and become less meaningful, depending significantly on the reference time according to which the phases are compared. This sort of comparison would not be applicable at all when there is no clear way to physically align predictions in time,
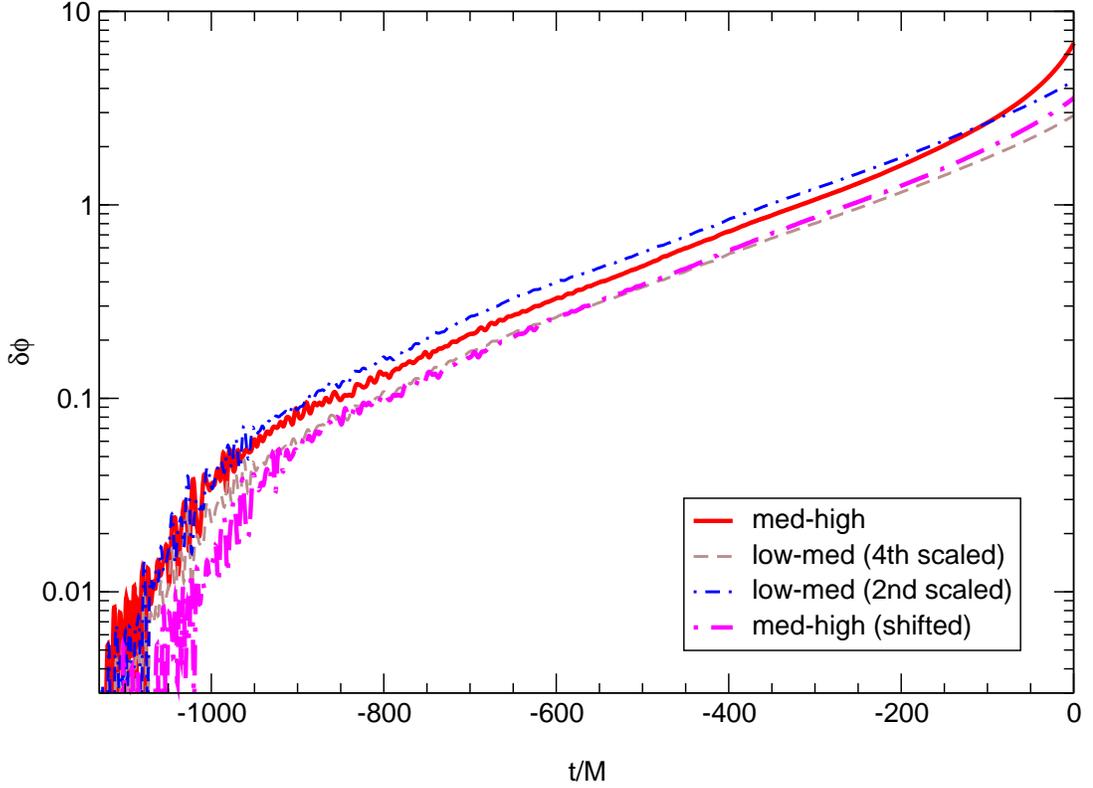
Figure 3.14: Strain rate phase three-point convergence. The higher resolution phase difference is shown with and without a phase shift to allow comparison of errors at similar dynamical points in the simulation. After shifting, the phase-error appears to be fourth-order convergent for much of the simulation.

such as when comparing numerical simulations with post-Newtonian (PN) results. However, as pointed out in a previous section, we can avoid assigning a reference time by using the gravitational-wave frequency as the reference.

For the case studied here, a quasi-circular inspiral of comparable-mass black holes, it is appropriate to consider the waveform frequency $\omega = \partial\phi/\partial t$ as a gradually

increasing monotonic function of time, $\omega \equiv \omega(t)$. Though the actual waveform frequency of our simulations is not monotonic because of small eccentricities in the inspiral trajectories, we have shown that we can fit out the eccentric deviations. The residual secular trend $\omega_c(t)$ provides a monotonic frequency, which we can apply as an independent variable against which to compare various simulations.

We show frequency-based phase differences among the simulations in Fig. 3.15, which allows us to compare the difference between the two higher-resolution simulations with that between the two lower-resolution simulations. If the simulation errors are fourth-order-convergent, then the low-medium difference should be approximately 2.784 times the medium-high difference. As is evident from the medium-high "(3M/80-M/32)" and low-medium "(3M/64-3M/80)/2.784" curves in Fig. 3.15, the errors appear slightly over-convergent with respect to fourth-order scaling. This may be caused by phase error accrued early in the lowest-resolution ($h_f = 3M/64$) simulation, due to difficulty in resolving high-frequency components in the spurious radiation content of the Bowen-York initial data pulse, as well as in an initial gauge pulse, which dominate at this time. This lowest resolution may therefore not quite be in the convergent regime during this early part of the simulation. If the phases are adjusted by a constant such that they match at some point after the main part of the Bowen-York pulse has passed, then fourth-order scaling fits more closely. This is demonstrated in Fig. 3.15 by the "(3M/64-3M/80)/2.784 (shifted)" curve, which has been vertically phase-shifted by a constant so as to agree with the "(3M/80-M/32)" curve at $\omega m_f = 0.05423$, the frequency $1000M$ before merger in our high-resolution simulation.
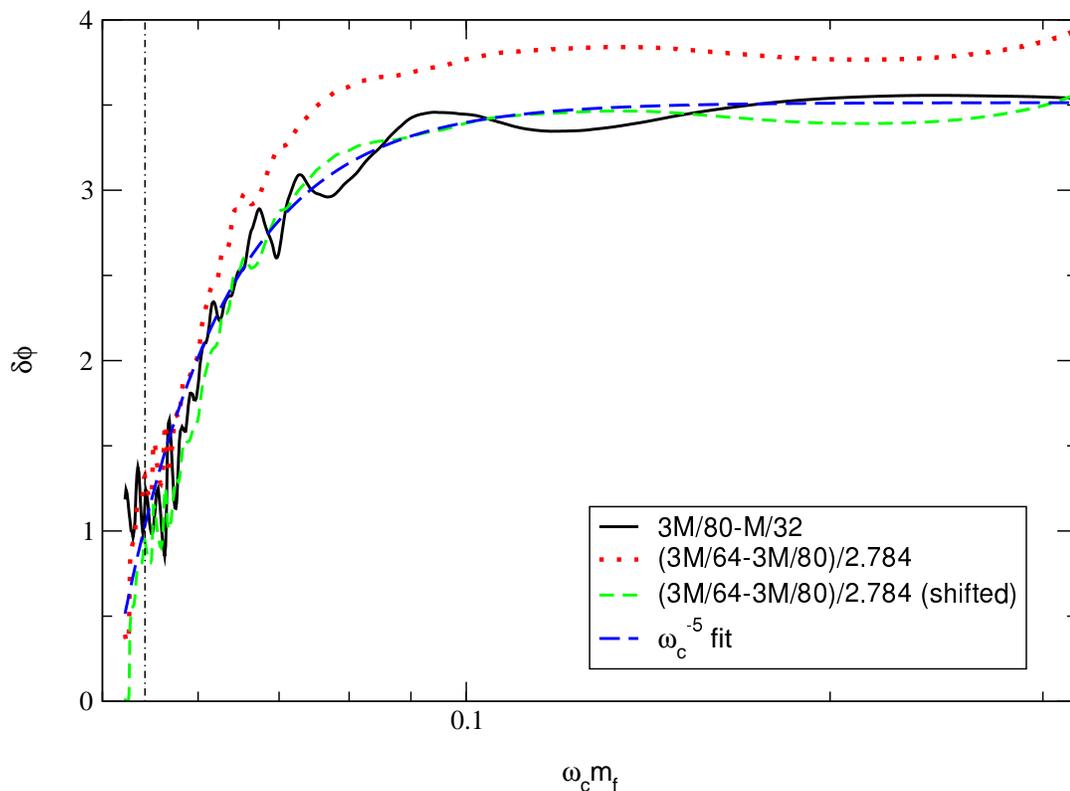
Figure 3.15: Frequency-based phase comparisons for runs at three resolutions. The solid curve represents the phase-difference between the $h_f = 3M/80$ and $h_f = M/32$ simulations; the dotted curve represents the phase-difference between the $h_f = 3M/64$ and $h_f = 3M/80$ simulations; the short-dashed curve represents the phase-difference between the $h_f = 3M/64$ and $h_f = 3M/80$ simulations, shifted vertically by a constant phase so as to agree with the higher resolution difference at $\omega_c m_f = 0.05423$ (shown as vertical dot-dashed line); and the long-dashed curve shows a fit to $\omega^{-5}$ phasing error, which might be expected if there are energy conservation violations that are constant in time. The lower-resolution curves have been scaled so that they should superpose with the higher-resolution curves in the case of fourth-order convergence.

66

As is also clear in the figure, phase error accumulates most significantly in the low-frequency portion of the simulation, $m_f\omega \lesssim 0.08$. This makes sense generally since the simulation spends much more time at lower frequencies. We can tie this observation back into our comparison between puncture track and radiation frequencies, to attempt to a fully consistent picture of the phase error. When the black holes are well separated, the dynamical development manifested in the sweeping frequency is driven by a slow loss of energy and angular momentum. A nonconservative leakage of energy $\delta\dot{E}$ will change the frequency sweep-rate $\dot{\omega}$ by $\delta\dot{\omega} \sim (\delta\dot{E})/(\partial E/\partial\omega)$, where $\partial E/\partial\omega$ indicates how the binding energy changes with frequency. Phase error $\delta\phi$ can be determined from the error in the sweep-rate by

$$\delta\phi = \delta\int\frac{\omega}{\dot{\omega}}\,\mathrm{d}\omega = -\int\frac{\omega}{\dot{\omega}^2}\delta\dot{\omega}\,\mathrm{d}\omega \tag{3.9}$$

Applying the leading-PN-order expressions for $\partial E/\partial\omega$ and $\dot{\omega}(\omega)$, we find

$$\delta\phi = -\int\frac{\omega}{(\omega^{11/3})^2}\frac{\delta\dot{E}}{(-\omega^{-1/3})}d\omega = \delta\dot{E}\int\omega^{-6}d\omega = -\frac{1}{5}\delta\dot{E}\omega^{-5} \tag{3.10}$$

Indeed this dependence fits our phase differences rather well assuming a constant leakage of energy, as shown in Fig. 3.15. Note, however, that this does not single out energy leakage as the source of error, as other errors may produce similar effects. For instance, a similar leakage of angular momentum would lead to an error proportional to $\omega^{-4}$, which would also fit reasonably well.

### 3.2.2   Comparing the Radiation Observables with the Puncture Tracks

One side effect of using the moving puncture technique is that, in addition to the typical method of extracting observable quantities such as waveform phase from the radiation, one can also potentially make use of some of the gauge versions of those quantities. For example, in the case of the fixed puncture gauge, the shift $\beta^i$ at the punctures was zero, so the punctures did not move through the grid (hence the name), and no insight could be gained from eq. (3.1), which relates the position of the puncture to the shift. However, in the case of moving punctures, eq. (3.1) has greater utility. Specifically, in order for the grid not to twist around the punctures, which can lead to a buildup of errors and ultimately make codes unstable, the puncture orbital phase will ideally coincide with the orbital phase of the binary in a non-corotational gauge modulo a constant.

The puncture phase has a potential advantage over the radiation phase as a measurement tool in that it is less vulnerable to errors from numerical dissipation. The radiation must be numerically propagated from the source to the extraction sphere, and this propagation accumulates error along the way. So, as was mentioned previously, extracting too close will lead to $1/R_{ext}$ error, but extracting too far will lead to errors biased toward the higher frequency content due to numerical dissipation, which may be compounded if the waveform has propagated through a region with coarse resolution. The puncture track, on the other hand, is a more locally defined quantity, relying predominantly on $\tilde{\Gamma}^i$, but it is a gauge quantity, and since our gauge is not analytic, defining the relationship between gauge quanti-

ties and observables calculated from the radiation in a rigorous way is challenging. However, we find empirically that our gauge is very similar to the ADM gauge [58], and regardless we do not perform any explicit gauge transformation in the upcoming analysis that requires explicit definition of the gauge.

We choose to use $\psi_4$ data for the following analysis, since we are comparing data sets representing phase evolution of the same run, rather than different resolutions or extraction radii, and we therefore need all the accuracy we can muster. Poor results are achieved when the $\psi_4$ data is smoothed using a moving average, since a rapidly growing trend tends to be damped, but, as seen in Fig. 3.16, excellent results can be achieved using a Savitsky-Golay filter. The difference between this filter and a moving average is that, whereas a moving average approximates an entire span by a single value, i.e. the average, the Savitsky-Golay filter fits the data in the moving window to a polynomial of the desired degree, in this case second order. We use this filter for all the results that come from $\psi_4$ data, as otherwise high frequency noise in the data obscures the meaningful underlying trend.

Before we perform the analysis, however, we digress briefly to illustrate why the choice of which quantity to use to calculate the phase matters. Fig. 3.17 demonstrates how different the result can be when $h_{2,2}$ is used to calculate the waveform phase instead of getting the phase directly from $\psi_4$. Note the presence of errors on the timescale of the waveform cycle, i.e. twice the orbital frequency, due to errors that were made in the zeroing procedure. These errors, if left unchecked, would obviously dominate the effect that we're attempting to measure. To further illustrate that the effect seen in Fig. 3.17 is indeed due to incomplete zeroing, we also present
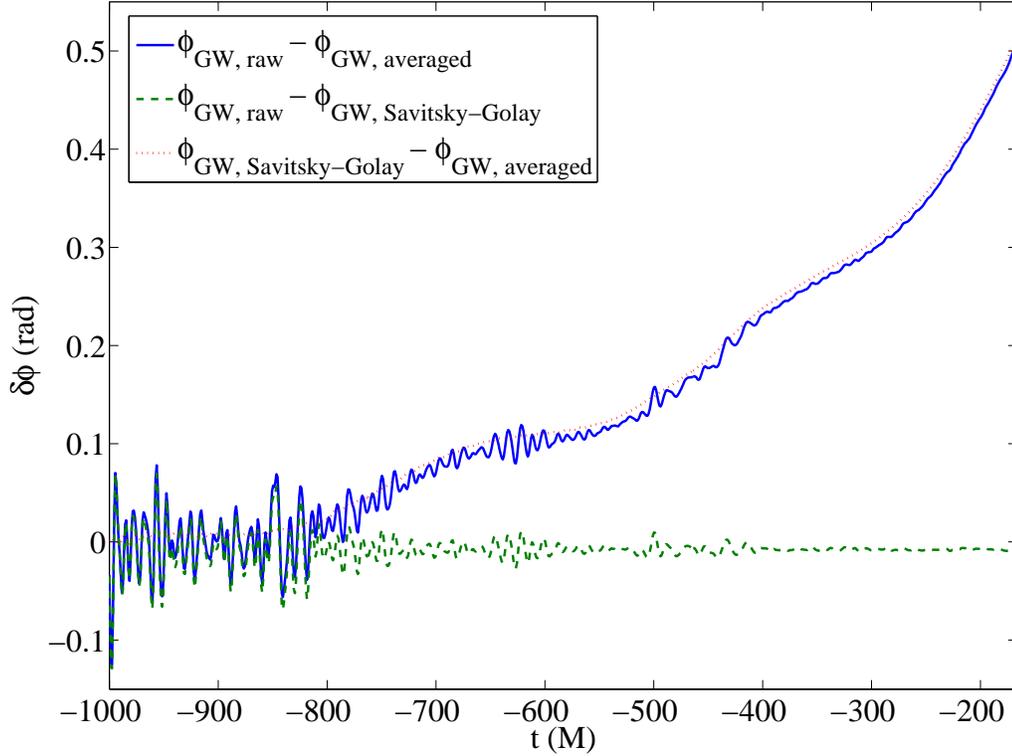
Figure 3.16: Phase difference due to smoothing using a moving average (solid) and a Savitsky-Golay filter (dotted) using the same length window and a second-order polynomial. The span was chosen to limit the moving average phase error to half a radian, but yield as smooth a first derivative (i.e. frequency) as possible. The difference between the two smoothed data sets is also shown (dashed) to demonstrate that the Savitsky-Golay filter successfully removed all the jittery content, leaving only the nonzero trend.

a toy problem in Fig. 3.18 to illustrate the point. As can be seen, even a moderate deviation from zero-mean can have a substantial effect on the frequency evolution. The explanation is analogous to Kepler's law of equal areas. If the waveform in Fig. 3.18 were viewed as a vector whose length represented the strain magnitude, and whose angle was the polarization angle, then zero-mean corresponds to the vector being centered at zero. However, if the waveform were shifted from zero-mean, so that the vector's base is no longer the origin, then the vector would have to wind at either an accelerated or decelerated rate to subtend the same arc in the same amount of time, depending on whether the vector's base was closer or farther away from the arc.

The convergence evidenced by Fig. 3.15 suggests that we can apply Richardson extrapolation to estimate the difference of our high-resolution $(M/32)$ run from the infinite-resolution limit, and use the difference between each data series and the extrapolated value as a measurement of our error. Richardson extrapolation is a simple way to numerically remove the leading error term if you have two sets of data at different resolutions and you know the convergence rate. Specifically, given two waveforms, $h_a(t)$ and $h_{a/b}(t)$, where $a$ and $a/b$ are the resolutions, we can write the exact answer $h(t)$ as

$$h(t) = h_a(t) + \alpha_o \Delta^n + \mathcal{O}(\Delta^{n+1}) \tag{3.11}$$

$$h(t) = h_{a/b}(t) + \alpha_o \left(\frac{a/b}{a}\Delta\right)^n + \mathcal{O}(\Delta^{n+1})$$

$$= h_{a/b}(t) + \alpha_o \left(\frac{\Delta}{b}\right)^n + \mathcal{O}(\Delta^{n+1}). \tag{3.12}$$
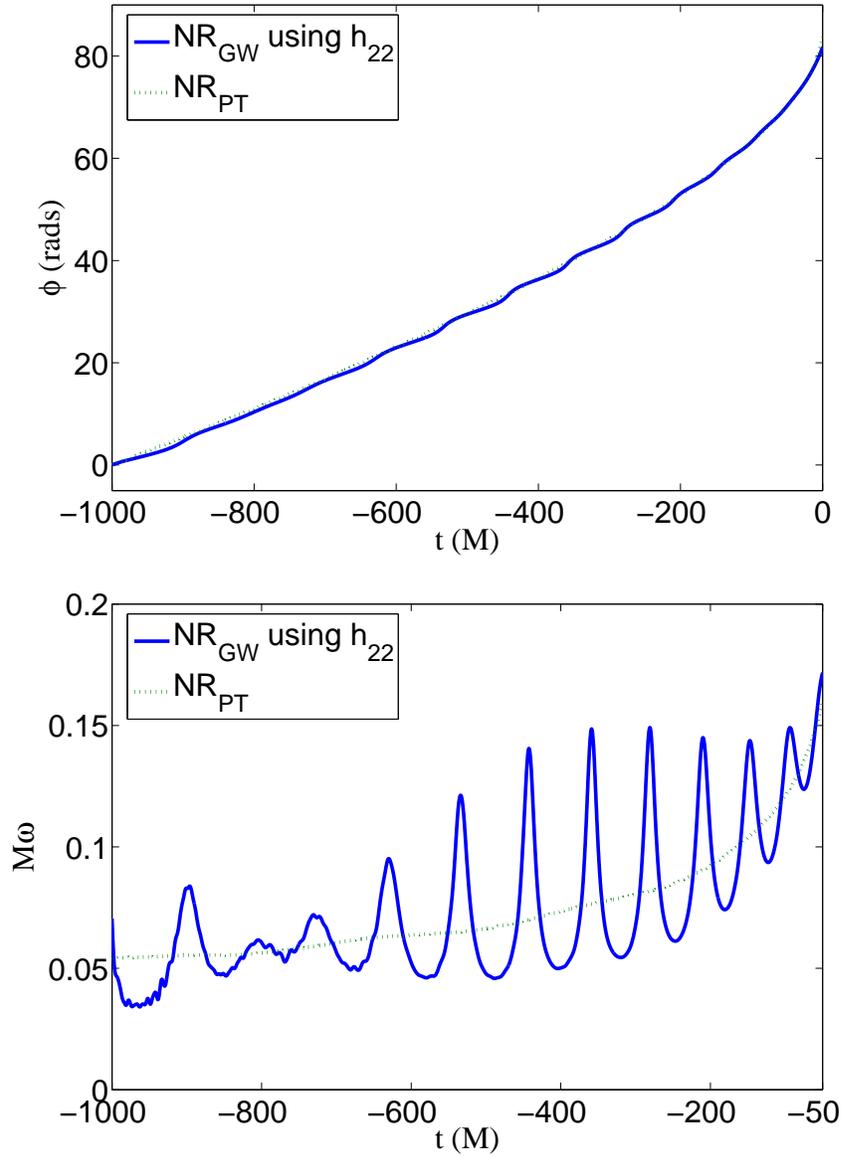
Figure 3.17: Orbital phase and frequency comparison between the puncture track value and the value extracted from the radiation when the phase is calculated from $h_{2,2}$. The underlying trend in the radiation is obscured by errors from the zeroing procedure.
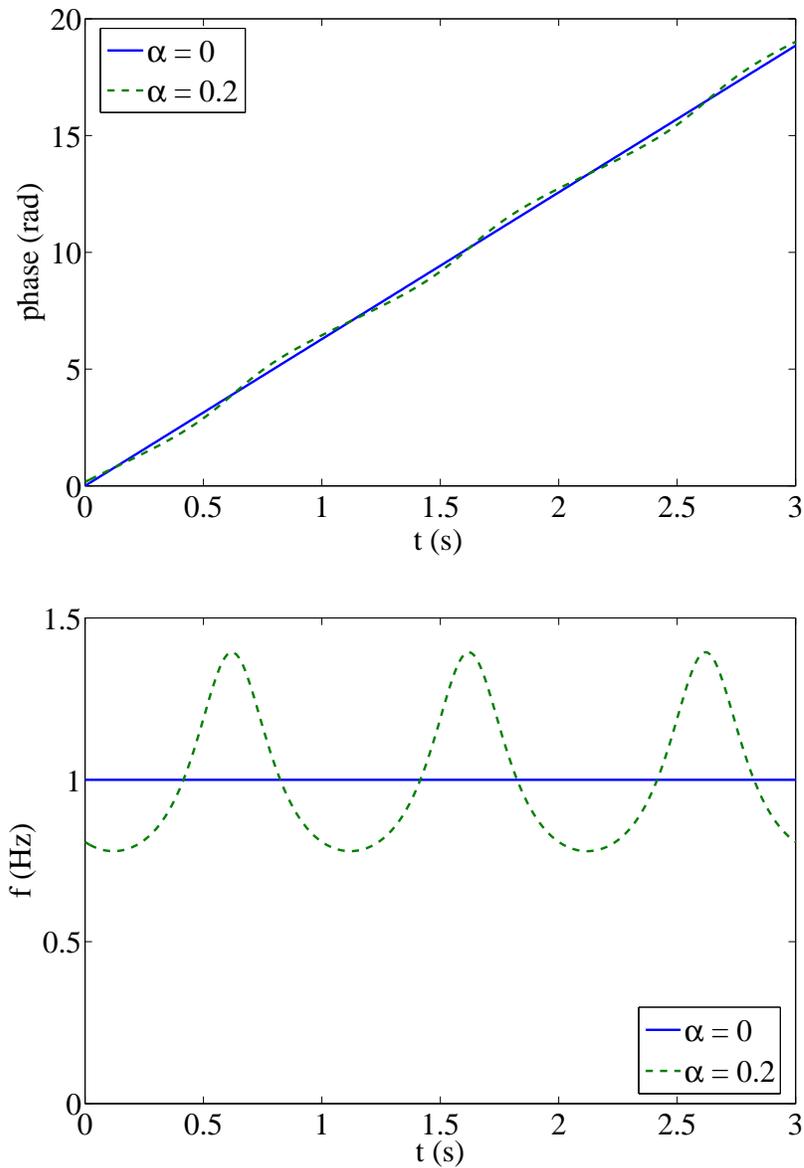
Figure 3.18: Demonstration of the error from incomplete zeroing. In this example, $h_+ = cos\phi + \alpha$ and $h_\times = sin\phi + \alpha$, so that $\alpha$ represents a deviation from zero-mean. As can be seen, a moderate deviation from zero-mean will cause a small deviation in phase, but a substantial deviation in frequency. The same characteristic scalloped pattern in the frequency of Fig. 3.17 is also clearly seen here.

We can then use these two equations to eliminate $\alpha_o$:

$$h(t) = \frac{h_a(t) - b^n h_{a/b}(t)}{1 - b^n} + \mathcal{O}(\Delta^{n+1}) \tag{3.13}$$

Using this relationship, if we assumed the fourth-order convergence suggested by Fig. 3.15, the phase error estimate for this run would be 0.93 times the difference between the phases from the $3M/80$ and $M/32$ resolution runs. However we will simply take the more conservative estimate of the actual difference between these resolutions. Note that during the last $1000M$ of our $M/32$ simulation (i.e. from $\omega m_f = 0.05423$ onwards), we estimate that roughly two and a half radians of phase error accumulate, as measured with respect to frequency, which is less than half of a gravitational wave cycle. A benchmark for accumulated waveform phasing errors is one-half of a cycle, because phase error exceeding this amount would lead to destructive interference in matched-filtering applications. For our high-resolution simulation, we estimate less than one-half cycle of gravitational wave phase error over the full simulated waveform, excluding the meaningless transients in the first $100M$. As in [1], we estimate that these phasing errors are smaller than the implicit phasing difference between the 3PN and 3.5PN expansions of $\dot{\omega}(\omega)$ after $t \approx -300M$ ($\omega_c M \approx .08$). For our data analysis considerations we will only be using the numerical waveform after this point, for which the estimated phase error is well below a half cycle.

Frequency-based phase comparisons, such as we have presented here, are better suited than time-based phase differences, which depend strongly on where the waveforms are chosen to be aligned in time. The relationship between the two can

74

be understood by considering a one-parameter family of waveform results, with a parameter $\lambda$ representing model dependence, in this case the numerical grid-spacing. The waveforms would provide phase as a function of time $\phi_\lambda(t)$, from which we can derive frequency $\omega_\lambda(t)$, which is monotonic for small eccentricity. Inverting to obtain $t_\lambda(\omega)$ one can derive the frequency-based phasing $\bar{\phi}_\lambda(\omega) \equiv \phi_\lambda(t_\lambda(\omega))$. Now considering variations $\delta \equiv d/d\lambda$ near $\lambda = 0$, one finds the relationship between frequency- and time-based phase comparisons

$$\delta\bar{\phi}(\omega) = \delta\phi(t(\omega)) + \omega\delta t(\omega). \tag{3.14}$$

This sheds some light on the often confusing issue of time alignment in the time-based comparisons shown earlier. Specifically, for a waveform that sweeps significantly through frequency, time- and frequency-based phase differences will be most similar when the time-based phase differences are aligned so that $\delta t$ vanishes where $\omega$ is largest. In our case, the net frequency-based phase differences in Fig. 3.15 are closer to net phase differences with time-aligned at the end of the waveform, as in Fig. 3.2, than when time is aligned at the lower-frequency beginning, as in Figs. 3.13 - 3.14.

Given the advantages of frequency based comparisons, and the availability of a circularized frequency to serve as our abscissa, we can compare the phase from both the extracted radiation and the puncture tracks with the T4PN waveform in order to assess the relative accuracy of each measurement technique. What we refer to as T4PN is the T4 phasing choice in [59], which is an expansion of the chirp rate $\dot{\omega}$ in terms of the frequency $\omega$, which we then numerically integrate, as well as the 2.5PN

75

amplitude from [60]. We see in Fig. 3.19 that both the radiation and the puncture track phase differences appear to be consistent with fourth-order convergence. A deviation at low frequency during the early noisy part of the waveform causes a slight phase offset in the case of the radiation, but the shapes are clearly consistent, and the use of phase means we could have freely added a phase constant to make the two radiation curves lie on top of one another better. In Fig. 3.20, the radiation phase differences both appear to outperform the puncture track ones, although by a relatively slim margin.

We should note that, due to crossings between the T4PN phase and the other data sets of interest, simple phase differences can accrue and cancel, making the comparison less meaningful. Since not only the total phase difference, but the instantaneous rate of accrual of that phase difference are indicative of the level of error, we wish to employ a metric for comparing runs which is sensitive both to net phase drifts and oscillatory deviations in phase. Therefore, for this comparison, we looked at a measure of phase difference that was constrained to increase only, namely

$$\delta\phi_*(\omega_c) = \int_{\omega_c'=\omega_o}^{\omega_c} \frac{|\omega_2(\omega_c') - \omega_1(\omega_c')|}{\dot{\omega}_c'} \, d\omega_c', \tag{3.15}$$

where $\omega_c'$ is used to distinguish the use of $\omega_c$ as the integration variable from its use as the upper limit of integration (and hence the dependent variable for $\delta\phi_*$). To distinguish from $\delta\phi$, we can refer to $\delta\phi_*$ as the "maximized phase error". We use a starting frequency of $M\omega_c = 0.054$ for the results shown in Fig. 3.20. We see that for both the high resolution and Richardson extrapolated cases, the puncture track

76
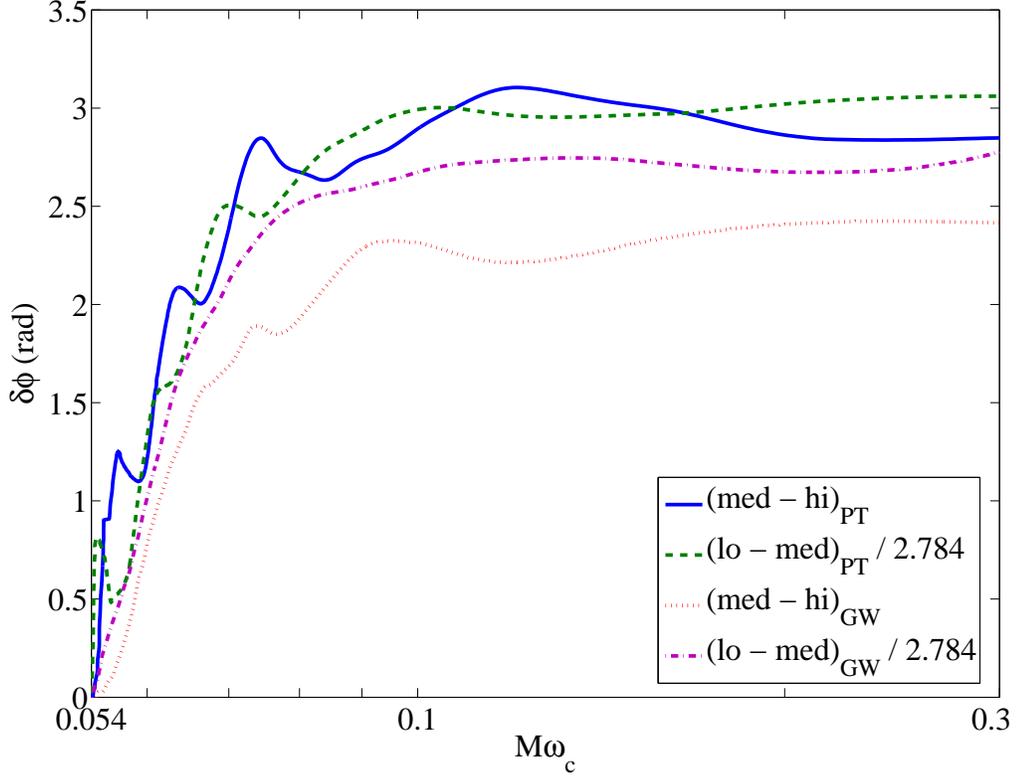
Figure 3.19: Demonstration of fourth order convergence for both the puncture tracks and the extracted radiation. All phases are set to zero at $M\omega = 0.054$, corresponding to $t = -1000M$, where $t = 0$ denotes the moment of peak strain amplitude. The frequency data for the radiation is calculated using $\psi_4$, after it has been smoothed using a Savitsky-Golay filter. The puncture track data is unsmoothed.

accumulates an additional $\sim 0.5$ rads of phase error compared to the radiation. However, it should again be noted that the radiation data required a very specific choice of which data to look at, and subsequently the data required a significant
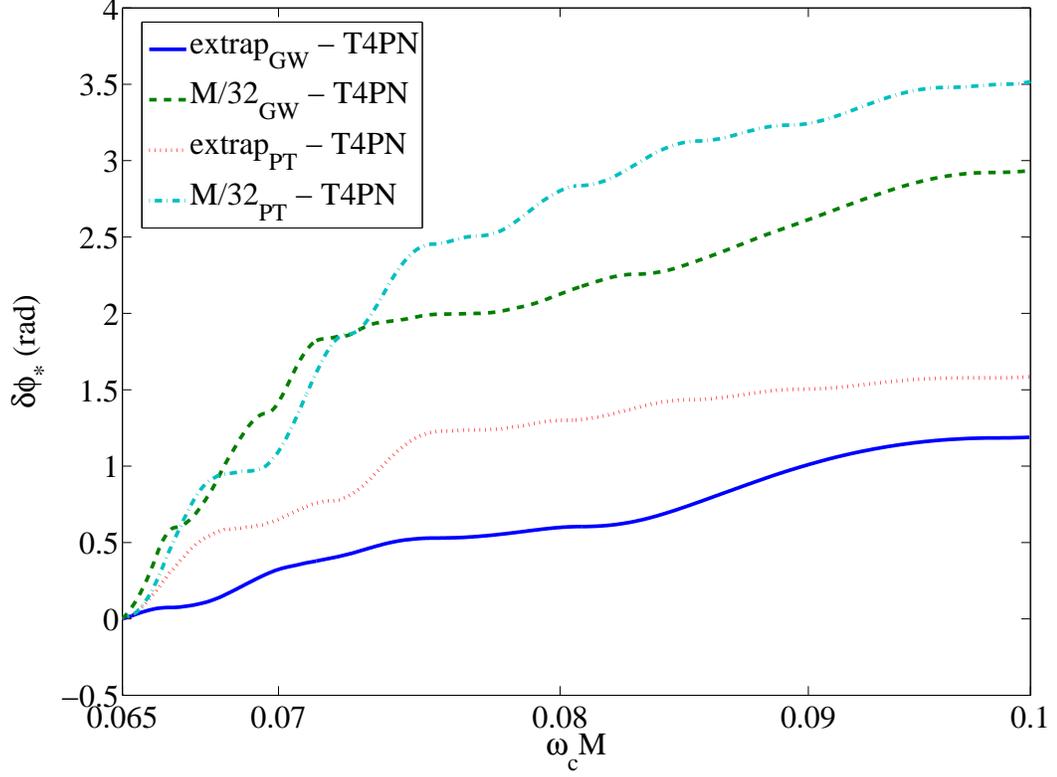
Figure 3.20: Comparison of maximized phase error, $\delta\phi_*$, measured as the difference with the T4PN prediction, for the high resolution and Richardson extrapolated cases from both the radiation (GW) and the puncture tracks (PT). The starting frequency for integration of eq. (3.15) is $M\omega_c = 0.054$, corresponding to $1000M$ before merger. The PT data shows an additional $\sim 0.5$ radians of phase error compared to the GW data for both the high resolution and extrapolated cases.

amount of smoothing, whereas the puncture track data has no such ambiguity, and the raw data is as smooth as the processed result for the radiation case.

While at present the analysis above shows that the evolution of the puncture track phase does not provide a more accurate orbital phase than its radiation counterpart, the choice of gauge has never been driven by an interest in yielding a highly accurate gauge phase. It may, therefore, be worthwhile to monitor the accuracy of the orbital phase and other observables calculated from gauge quantities in the way presented here. It is possible, for instance, that at some point a gauge may be found which yields either such small internal error sources in the puncture tracks, or such perfectly canceling internal error sources, that the resulting phase corresponds, to a high degree of fidelity, with the radiation phase extracted at infinity, which is only approximated by our radiation measurements extracted at finite radii. We note that as our gauge choices have evolved, they have steadily led to better agreement between the puncture behavior and the dynamics predicted by the radiation.

Since we cannot disprove the possibility that such a gauge exists, we cannot dismiss the possibility that quantities from the puncture track, although gauge dependent, may ultimately yield the most accurate orbital phase for finite differencing codes. We note, however, that we cannot measure all the same observables from the puncture track that can be measured from the radiation. The mode decomposition of the radiation, as well as the amplitude, are of course impossible to calculate (aside from applying PN) using the puncture tracks. Therefore, we must emphasize that we are not suggesting an eventual replacement of the radiation observables with puncture track observables. Rather, we advocate keeping track of the subset of total radiation observables related to the orbital dynamics which can be calculated using the puncture tracks, and comparing their predictions with each other and with

PN predictions to gain insight into impact of the gauge on the dynamics, with the possibility that our choice of gauge and evolution equations may eventually lead to the aforementioned damping of error in gauge phase. Even without achieving such an ideal gauge, another measure of phase which is largely free from noise is always advantageous as a check when analyzing these waveforms which, as has been discussed, have substantial noise content and must be filtered to provide an accurate measure of phase.

### 3.2.3 Validation of PN Phase Using Numerical Relativity

Having determined that, indeed, the radiation provides a slightly more accurate measurement of phase than does the puncture track, we use the radiation phase to answer a complementary question, that being, "Can the phase evolution calculated using numerical relativity be used to validate the accuracy of the PN approximation?" It turns out that the answer is yes, as has recently been demonstrated to extremely high accuracy in [59]. However, at the time this analysis was originally carried out (the result ultimately being published in [1]), the result was somewhat unanticipated.

We again employ a frequency-based comparison of the phase predictions from the two methods. We use the T4PN as the representative of PN methods based on the qualitative agreement with numerical simulations that was demonstrated in [56] (which, given that the 3.5PN order T4PN has recently been shown to agree with the numerical simulations presented in [59] to within 0.05 radians over 30 orbits,

has proven to be a wise choice). We again show the differences for the different resolutions of simulations, along with the differences constructed from the three highest PN phase expressions, i.e. the 2.5PN, 3PN, and 3.5PN phases. Finally, we take the Richardson-extrapolated result found using our two highest resolution simulations, and difference that with the 3.5PN phase prediction. The result is shown in Fig. 3.21. Two areas of note are the regions where a portion of the "3PN-3.5PN" data set have been translated to overlap other data sets. The first such region, at a frequency of $\omega_c M \sim 0.08$, shows approximately where the rate of accumulation of error in the numerical simulation drops below that in the PN approximation, as represented by the size of the highest order PN correction term. The second point shows where the "extrap-3.5PN" data set is last equal in slope to the "3PN-3.5PN" data set. The subsequent divergence of the former, and stable behavior of the latter, indicates that after this point, the "3PN-3.5PN" data set can no longer be relied upon to give physically reasonable results. Consequently, in the region between these two areas, the small overall value and small slope (indicating a slow rate of error accumulation) for the "extrap-3.5PN" is a striking validation of the PN approximation using numerical simulations in a regime where both methods should be valid.

### 3.2.4   Waveform Amplitude

Having thoroughly assessed the phase characteristics of the waveform, we will now briefly turn to the behavior of the amplitude. Because the runs being analyzed
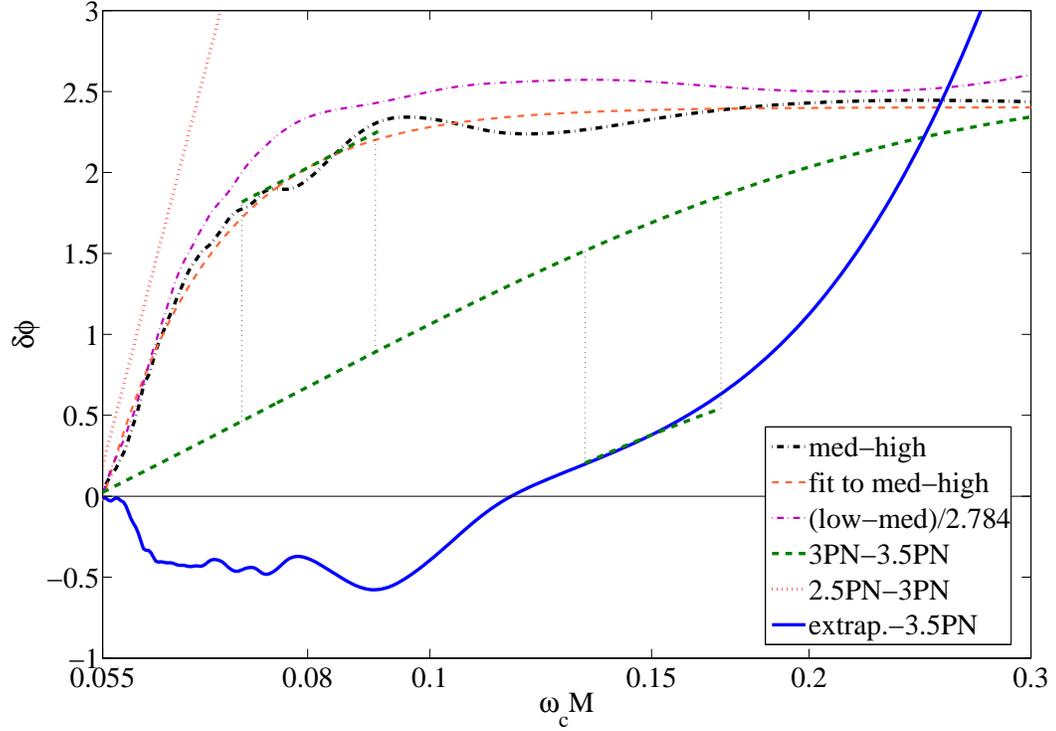
Figure 3.21: Gravitational wave phase error estimates. Differences between phasing from the 3.5PN chirp rate and Richardson extrapolation from the numerical simulations (solid blue curve) are small, and are consistent with internal error estimates for the numerical simulation results (dash-dotted) and the PN sequence (dotted).

here lack sufficient accuracy in the amplitude to perform all the analyses that were done on the phase (for example, a quantitative comparison and validation of the PN amplitude), we will provide a more qualitative description of the numerical amplitude. More recent work [61], using higher-accuracy simulations, has carried out such comparisons with PN predictions.

As can clearly be seen in Fig. 3.22, the fluctuations in the numerical amplitude

are too great for us to be able to validate one order of PN over another with any level of certainty, as a single simulation varies sufficiently to span across all orders of PN over the course of the evolution. It is noteworthy that when the different PN order amplitudes are shifted in time by making the corresponding frequencies at 3.5PN order all be equal at some matching time ($t = -328M$ in this case), i.e. $A_{3PN}(\omega_{3.5PN}(t = -328M)) = A_{2.5PN}(\omega_{3.5PN}(t = -328M)) = \ldots$, we see that the 3PN amplitude is nearly identical to the 2PN amplitude (hereafter referred to together as 2PN/3PN) until $\sim t = -100M$. However, both 2PN and 3PN are considerably larger than our numerical waveform at the time where we wish to attach a PN inspiral. The 2.5PN amplitude, on the other hand, is very close, and so allows us to construct a smooth combined PN-NR waveform. The slope of the 2PN/3PN appears to agree better at the matching point and thereafter, however, and we find it likely that the disagreement between the NR waveform and 2PN/3PN is the result of numerical dissipation in the simulation, rather than shortcomings in the PN sequence. We choose the 2.5PN for convenience since it makes a smooth hybrid waveform, and also to be conservative for our later analysis since the 2.5PN predicts less power being radiated (and hence lower SNR) than does the 2PN/3PN.

For the purpose of assigning a fractional error to the amplitude, we have a number of choices. However, we observe that the 2PN/3PN amplitude tracks the waveform steadily throughout the evolution, is consistent with 2.5PN $\sim 1000M$ before merger, and is roughly consistent with the disagreement among numerical resolutions at $t \approx -100M$, although it quickly diverges thereafter. The fractional amplitude error estimate $(3PN - NR)/3PN$ therefore ranges from $\sim 10\%$ at $t =$
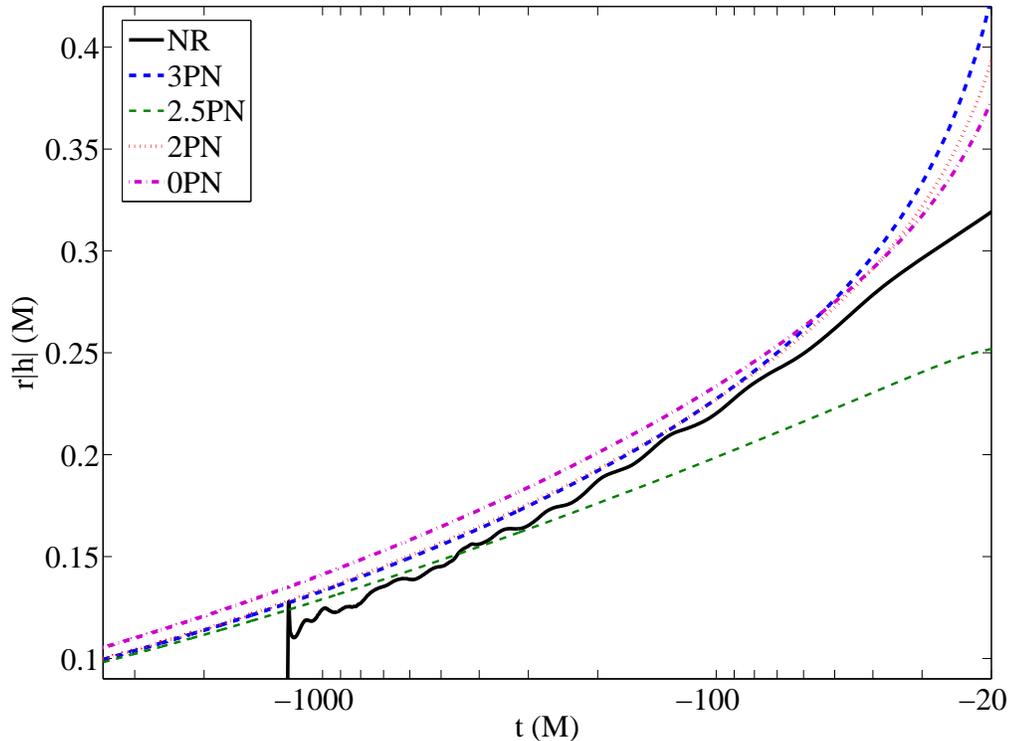
Figure 3.22: Comparison of NR amplitude with different order PN predictions.

$-1000M$ to $\sim 2\%$ at $t = -100M$.

It was observed in [62] that the amplitude predicted by numerical simulations was consistently less than the PN predictions at early times where PN is most likely to be valid. This discrepancy was seen to decrease with increasing extraction radius, indicating that the dominant error at early times was finite extraction radius. This led some groups to the practice of extrapolating waveforms with respect to extraction radius [61], as is typically done with resolution. However, care must be taken when dealing with extraction radius extrapolation, because whereas one source of error (the approximation of extracting radiation at an infinite distance) decreases

with increasing extraction radius, another error source, the error resulting from numerical dissipation as the radiation is propagated through the grid, *increases* with increasing extraction radius. Therefore, we can anticipate a crossover point, since numerical dissipation becomes stronger at higher frequencies, where extrapolating will cease to bring the result closer to the correct answer in the limit of infinite resolution at infinite extraction radius, but will cause is to systematically deviate from that answer. A study constrained to the case of the lower frequencies (as ultimately was the case for [61] due to technical issues) will be able to proceed with such radius extrapolation, but any attempt to create an accurate complete waveform must take greater care. Since the effects of finite extraction are strongest at low frequencies (when the extraction radius becomes comparable to the gravitational wavelength), an ideal complete waveform would be a hybrid of a waveform extrapolated in both resolution and radius at low frequencies, with the data used having been extracted at as far a radius as possible, combined with a high frequency late inspiral-through-merger waveform extrapolated only in resolution using closer extraction radii. As we continue to pursue the highest level of accuracy achievable, we will begin constructing more elaborate hybrids such as these.

The effects of numerical dissipation can clearly be seen in the power output as calculated from the gravitational waves. To illustrate the varying behavior of the waveform amplitude as we move from low frequencies and earlier times to high frequencies and late times near merger, we use data from a very recent run which incorporates higher order spatial differencing, and can thus provide accurate solutions at lower resolutions (only a single resolution was available at the time of this work).

The following results have 3M/64 resolution, which is equivalent to our lowest resolution in the set of equal mass, nonspinning runs which are analyzed throughout the rest of this work. However, whereas our standard runs converge at fourth order, in shorter test runs the new version of the code converged at between fifth and sixth order. As can be seen in the top panel of Fig. 3.23, power output increases with extraction radius until late in the inspiral. However, looking at the bottom panel, we see that as the system moves from late inspiral into merger, the trend flips and power output decreases with larger extraction radius, which is consistent with error from numerical dissipation coming to dominate over errors from finite extraction radius.

Lastly, we investigate the convergence of the errors from numerical dissipation with extraction radius. Specifically, we compare the strain amplitude (times the extraction radius, to compensate for the $1/r$ decay) extracted at $90M$, with the Richardson-extrapolated result using data extracted at $50M$ and $70M$. The agreement is reasonably good in the top panel of Fig. 3.24, with the dominant sources of deviation being imperfect zeroing of the waveform (which explains the oscillations at the GW frequency, which are amplified by extrapolation) and, less significantly, eccentricity. There is far less apparent convergence at late times as shown in the bottom panel of Fig. 3.24, which makes application of Richardson extrapolation in extraction radius over the range shown questionable.

The presence of numerical dissipation does not automatically make convergence impossible, since the radiation is decaying as $1/r$, so that any fractional error will also decay at that rate, and could potentially converge. However, the conver-
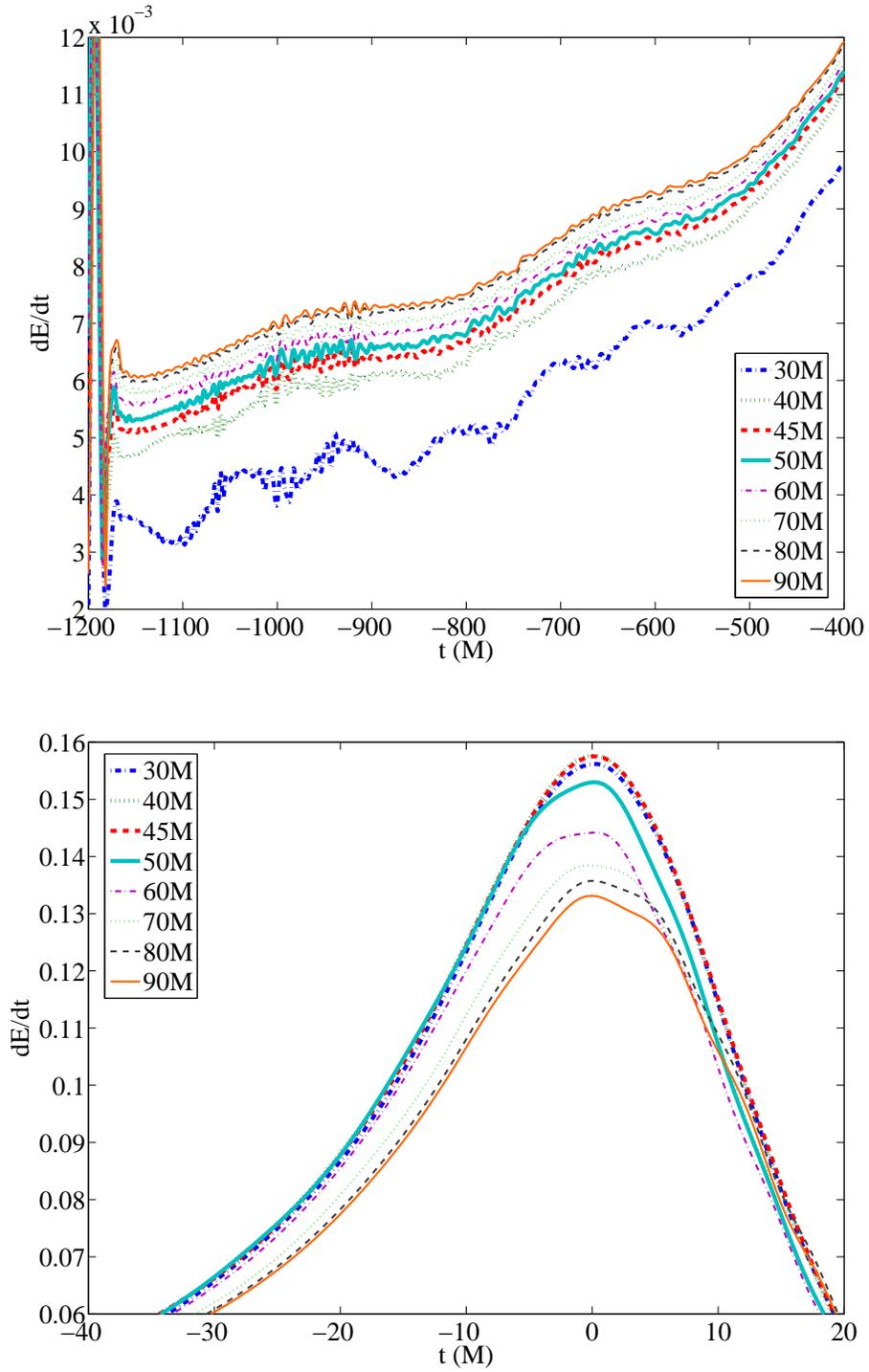
Figure 3.23: Power as calculated from the extracted radiation. For most of the inspiral, power output increases with increased extraction radius (top), but peak power decreases with increased extraction radius due to numerical dissipation (bottom).
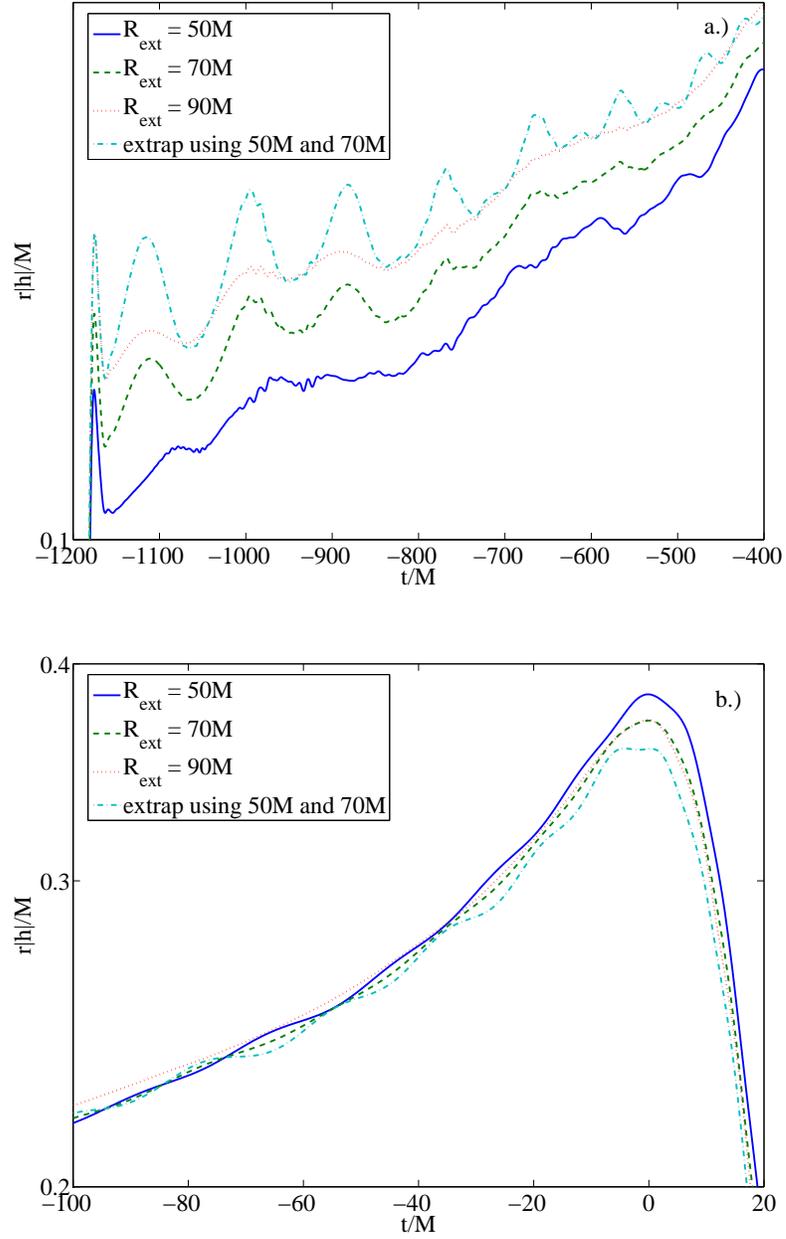
Figure 3.24: Waveform amplitude calculated from extracted radiation. The extrapolated amplitude calculated using $R_{ext} = 50M$ and $R_{ext} = 70M$ agrees with the amplitude at $R_{ext} = 90M$ during the inspiral, as seen in a.), but not in the merger, as seen in b.). The dominant error is imperfect zeroing of the waveform, as can be seen by the oscillations in a.), which are amplified by extrapolation.

gence would be to a value equivalent to the correct value plus some finite error, with the error given by a value whose integrand happened to converge when integrated from 0 to infinity. Therefore, the presence of error is no guarantee of a failure to converge and, likewise, successful convergence is not necessarily a guarantee of accuracy. These issues, though subtle, are important to be mindful of when attempting to assess and mitigate error present in our simulations.

Chapter 4

Signal Detection and Characterization

In this Chapter, we move from our detailed analysis of the waveform charac-
teristics presented in the preceding Chapter into more application-oriented issues.
Using a hybrid PN-NR waveform, we calculate various quantities related to signal
detection. In particular, we present several representations of the SNR for LIGO,
Advanced LIGO, and LISA for both the equal mass case as well as several cases of
unequal mass, nonspinning waveforms. Finally, we address the question of statistical
uncertainty that we can reasonably expect to be present in a LISA measurement of
an equal mass BHB, and discuss the subtleties related to distinguishing statistical
error from theoretical error in the waveform.

## 4.1   Making our Template

As indicated in the preceding chapter, the PN component of our hybrid PN-NR
waveforms, which we will apply in this chapter, is the T4PN as it was previously
defined. It is worth noting that this method of calculating the phase, via direct
numerical integration of the 3.5PN expansion of the chirp rate, $\dot{\omega}_{3.5PN}$, for example
in the integrand $d\omega(\omega/\dot{\omega}_{3.5PN})$, does not strictly respect the proper procedure for
PN approximation, as the latter would require additional 3.5PN expansion of the
integrand itself. However, the phase obtained in this manner will have the same

convergence properties as the original $\dot{\omega}_{3.5PN}$ expansion, which is arguably a more fundamental PN quantity because of its close relationship to the rate of energy loss, $\dot{E}$, from which it is derived. Additional expansion of the phase may compromise its accuracy. Alternately, since the T4PN procedure is done numerically, and therefore permits no way to truncate higher-order PN terms upon integrating after each step as one moves from solving $\dot{\omega}_{3.5PN}$ to get $\omega_{\sim 3.5PN}$, then to integrating $\omega_{\sim 3.5PN}$ to get $\phi_{\sim 3.5PN}$, it could be that the higher-order terms that are effectively kept instead of truncated may account for the improved accuracy. Whatever the cause, as was mentioned in Chapter 3, the accuracy of the T4PN was demonstrated in [59] to be so much better than we would have any justification to expect, that for the moment it might be best to just count our blessings and use it in blissful ignorance.

Fig. 4.1 shows our numerical waveform overlaid with the PN waveform that was just described. To generate a complete, mass-scalable waveform, we match the frequency of the numerical simulation to the PN prediction, adjusting the phases to also be equal at that point as shown in Fig. 4.1, and connecting the two halves to make a single waveform. This is done by shifting the PN waveform until the frequency equals the numerically-predicted frequency at a time in the simulation where the accuracy of the numerical data first surpasses the accuracy of the PN approximation, as estimated in [1]. Specifically, [1] predicts this point of equivalent accuracy to occur at $M\omega \sim 0.08$, which corresponds to $t = -328M$ (shown by the circle in Fig. 4.1).

It is interesting that there was no need to adjust the 2.5PN amplitude for continuity. The amplitude agreement with the numerical simulation is so good, and
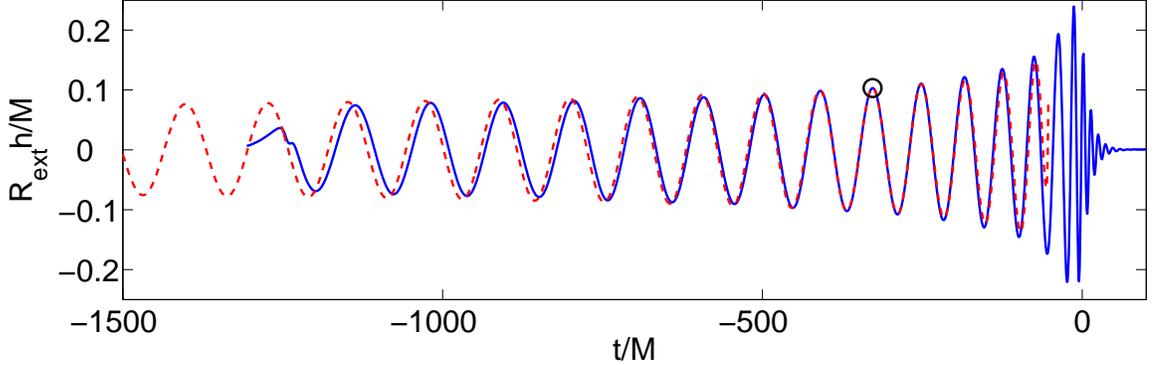
Figure 4.1: Numerical waveform overlaid with T4PN waveform as described in Appendix A. The circle indicates the point at which the waveforms are joined to form a hybrid NR-PN template waveform for data analysis applications.

hence the resulting amplitude is so nearly continuous, that the small discontinuity fails to produce any discernible artifacts in the Fourier transform $\tilde{h}(f)$ of the resulting waveform. The numerical data is shifted so that the peak amplitude occurs at $t = 0$, and the PN amplitudes are shifted so that the corresponding 3.5PN frequency matches the numerical run at $t - 328M$.

It is worth noting that the recently calculated 3PN amplitude [63] is negligibly different from the 2PN amplitude compared both to the internal errors in the numerical amplitude and to the typical scale of the individual PN terms, indicating that the 3PN correction has simply canceled out the 2.5PN correction until very late times (after $\sim -70M$) where we know from simulation that the binary has begun merging, and PN is no longer valid. Furthermore, although the 2.5PN amplitude is closest to the NR amplitude at the location where we match NR and PN, the NR amplitude trend is more similar to that of the 2PN/3PN amplitudes, so the

agreement may well be due to leakage of energy from the simulation that should be present in the radiation, causing the amplitude to be lower than it should be. Since our internal errors are too great to draw a conclusion regarding which PN order is best, we choose 2.5PN due to its aforementioned continuity with our NR data and its lower value, which will yield slightly more conservative SNR results later.

Having generated a waveform, it is informative to estimate the waveform's phasing accuracy over the course of the BHB evolution. Note in Fig. 3.15 that for the portion of our $M/32$ simulation that is used in the waveform, we estimate $\sim 0.5$ radians of phase error. If we take the difference between 3 and 3.5 PN terms to be an estimate of the phase error as in [1], we can assess the error for the PN portion of the waveform. It was shown in [64] that the analytic PN phase expression accumulates very little error, on the order of 0.1 radians, until $M\omega \sim 1 \times 10^{-4}$. Beginning our numerical phase integration at this point and evaluating up to $M\omega = 0.08$ yields a gravitational wave phase error of $\sim 3.6$ radians, such that the total accumulated phase error over the entire waveform is $\sim 4$ radians. As stated previously, an accumulated waveform phasing error of less than $\pi$ radians is the threshold below which wave-matching comparisons may be used for matched-filtering applications. We estimate that our combined waveform meets this criterion after a frequency of about $M\omega \sim 0.01$ up to the ringdown frequency, $M\omega \sim 0.5$. We therefore have a waveform with sufficient accuracy to be useful as a template for gravitational wave detection. While templates will ultimately be needed for cases of greater astrophysical interest, and still greater accuracy will be required for the template to be useful for the purpose of parameter estimation, the construction

of this waveform illustrates that the field of numerical relativity has matured to the point of being capable of producing results that are useful for gravitational wave data analysis. Furthermore, in our final analysis we calculate the expected parameter errors assuming perfect templates, so we can compare our theoretical errors for the phase and amplitude from the previous chapter to see how useful the waveforms presented here would be in actually attempting to meaningfully estimate those parameters in a real signal.

## 4.2   SNR Calculation

If one's only interest is to detect signals, then the relevant metric is the signal-to-noise ratio (SNR). The SNR is calculated assuming matched-filtering is performed on the data, and that the waveforms are perfect copies of the embedded signal. In this case, the sky- and waveform-polarization-averaged SNR is given by

$$\langle (SNR)^2 \rangle = \int d(\ln f) \left( \frac{h_{char}(f)}{h_n(f)} \right)^2 , \qquad (4.1)$$

where $h_{char}(f) \equiv 2f|\tilde{h}(f)|$ is the characteristic signal strain and $h_n(f) \equiv \sqrt{5} h_{rms}(f) = \sqrt{5 f S_n(f)}$ is the rms of the detector noise fluctuations multiplied by $\sqrt{5}$ for sky-averaging, with $\tilde{h}(f)$ and $S_n(f)$ being the Fourier transform of the signal strain and the power spectral density of the detector noise, respectively [65].

The waveform scales with luminosity distance $D_L$ and total mass $M$ as $h_{char} \propto (1 + z)M/D_L$, while the time axis for an observed wave, after redshifting, scales as $t \propto (1 + z)M$, so that the waveform shown in Fig. 4.1 is applicable over all total masses and redshifts. When needed we relate luminosity distance to redshift $z$

using cosmological parameters consistent with the most recent Wilkinson Microwave Anisotropy Probe (WMAP) results ($\Omega_\Lambda = 1 - \Omega_M = 0.72$, $h = 0.73$) [66] and the relation

$$D_L(z) = \frac{(1+z)c}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_M(1+z')^3 + \Omega_\Lambda}}. \tag{4.2}$$

For cases where an impractically long time series would be needed to cover the band with an adequate sampling rate, the waveform is extended in Fourier space to still lower frequencies (and consequently back further in time) using the quadrupole formula,

$$|\tilde{h}_{\text{quad}}(f)| = \frac{1}{2\sqrt{15}D_L} \left( \frac{[(1+z)M]^5}{\pi^4 f^7} \right)^{1/6}. \tag{4.3}$$

The PN portion of the waveform continues slightly past where its Fourier transform deviates from the quadrupole expression for $|\tilde{h}(f)|$ by $\sim 2\%$, at which point eq. (4.3) is used to extend $|\tilde{h}(f)|$ back as far as necessary. The PN segment is truncated at a higher frequency than the lowest frequency component of its Fourier transform in order to eliminate edge effects. Finally, the quasi-normal ringdown at the end of the numerical simulation is extended by fitting a damping coefficient and fundamental frequency to the data in order to mitigate edge effects at the high-frequency end of the Fourier transform.

## 4.2.1 Observing Stellar BHBs and IMBHBs with LIGO and Advanced LIGO

Ground-based interferometers are sensitive to relatively high frequency gravitational waves from coalescing stellar mass ($M \lesssim 10^2 M_\odot$) and intermediate mass

(IM) ($10^2 M_\odot \lesssim M \lesssim 10^3 M_\odot$) BHBs. In this section, we apply our combined waveform to consider the response of LIGO and Advanced LIGO to BHB coalescence, illustrating the importance of numerical simulation results for ground-based detectors.

For our analysis of LIGO, we used the design sensitivity to characterize the detector noise [67]. This sensitivity assumes that the noise is seismically limited below 40 Hz, thermally limited between 40 and 150 Hz, and shot-limited above 150 Hz. For Advanced LIGO, unlike LIGO, we had a choice of tuning configurations. We used the wide-band tuning typically associated with burst sources because of its dramatically superior sensitivity at higher frequencies, where the merger portion from many sources is predicted to occur [68]. This yielded an improved SNR for most masses compared to tunings that were optimized for only the early inspiral portion of the coalescence.

In Fig. 4.2, we show $h_{char}$ for several sources plotted relative to the $h_{rms}$ sensitivity curves for LIGO (dashed line) and Advanced LIGO (dash-dotted line). We plot these values because the height of $h_{char}$ above $h_{rms}$ is an indicator of the SNR, as can be seen by inspecting eq. (4.1). It is also informative to see that, despite the relative brevity of the merger in time, it spans a relatively broad range of frequency with significant power content.

By rescaling we can calculate the sky-averaged SNR as a function of redshifted mass, and particular luminosity distance $D_L$. In Fig. 4.3 we plot the SNR achievable by LIGO for sources at a luminosity distance $D_L = 100$ Mpc as a function of redshifted mass $(1 + z)M$. Here, the dashed line shows the SNR from the early
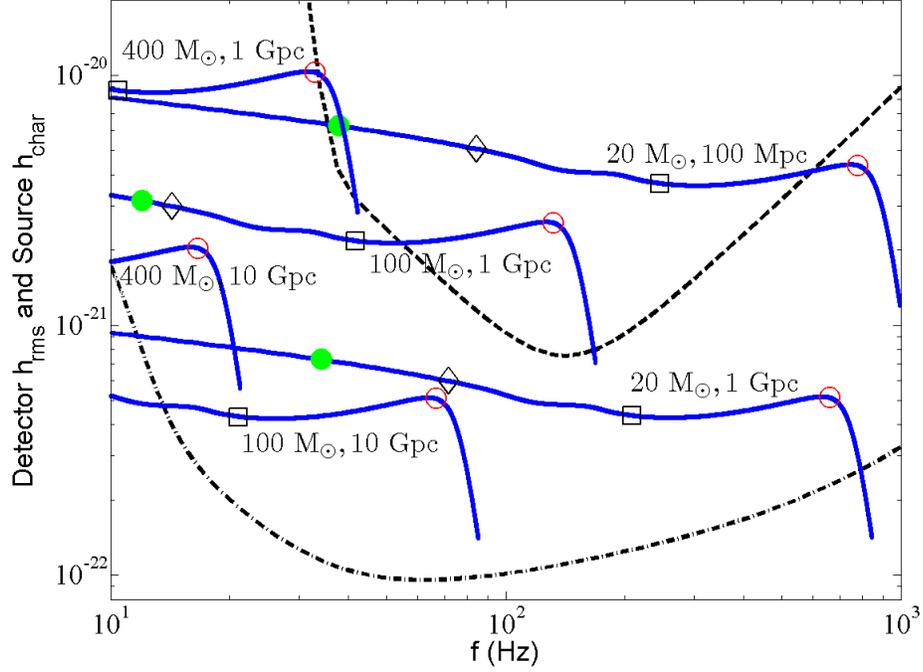
Figure 4.2: The LIGO (dashed) and Advanced LIGO (dash-dotted) rms noise amplitudes $h_n$ with the characteristic amplitudes $h_{char}$ of 6 example sources (solid). The locations on each $h_{char}$ corresponding to the peak $\psi_4$ amplitude (circle) and 1 second before the peak in the observer's frame (filled circle), as well as $t = -50M$ (square) and $t = -1000M$ (diamond) in the source's frame, are as marked. The mass given is the combined rest mass of each black hole.

inspiral in the time range $-\infty < t < -1000M$, which is roughly up to the start of our run. The dotted line shows the SNR for the late inspiral, $-1000M < t < -50M$, where $t = -50M$ is approximately the time at which the merger burst begins. The thin solid line gives the SNR for $-50M < t < \infty$, and encompasses the merger-
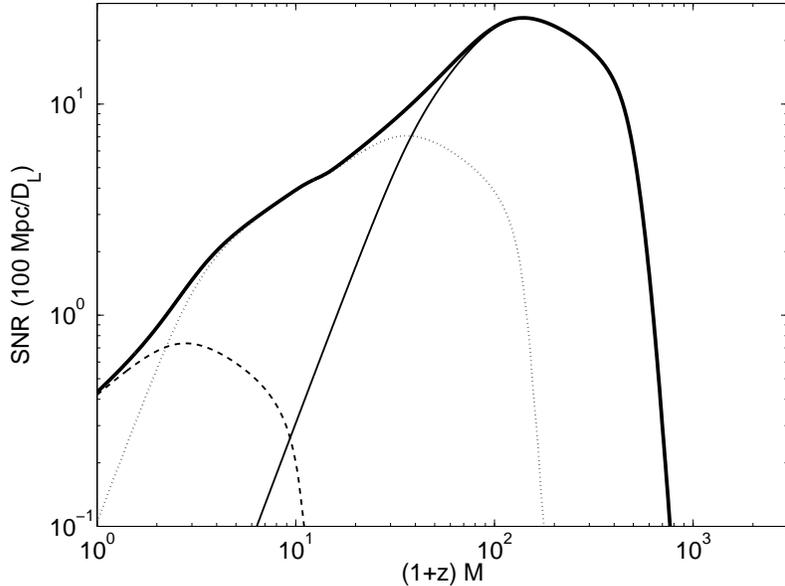
Figure 4.3: SNR for sources at luminosity distance $D_L = 100$ Mpc plotted vs. redshifted mass for LIGO. The contributions from $-\infty < t < -1000M$ (dashed), $-1000 < t < -50M$ (dotted), and $-50M < t < \infty$ (thinner solid), as well as the SNR from the entire waveform (thicker solid) are shown.

ringdown part of the signal. The thick solid line shows the SNR from the entire waveform. Note that the addition of the merger-ringdown waveforms increases the SNR and extends the detectable mass range significantly. The merger-ringdown portion $t > -50M$ dominates for all equal-mass nonspinning merger observations detectable with SNR larger than 10 at 100 Mpc.

This type of plot was first made in [65], and it is useful to compare our results with theirs. Our SNR calculations are based on a full waveform for the case of equal-

mass, nonspinning black holes. The work in [65] was done before merger waveforms were calculated and thus is based on estimates for the merger-ringdown regime. For example, they estimated a merger radiation efficiency of $\sim 10\%$, which is higher than our results but may well obtain for mergers with spin. Comparing their Fig. 4 for the SNR for LIGO with our Fig. 4.3 we note that their curve for the inspiral includes the radiation up to the merger and so should be compared to the combination (in quadrature) of our dashed and dotted curves. Our result for the merger SNR is somewhat smaller than theirs, due to the smaller amount of radiation emitted in our mergers. More recently, an analysis of SNR for LIGO using numerical relativity waveforms for the merger and PN waveforms for the inspiral was made in Ref. [56]; our results in Fig. 4.3 are similar to what they report in their Fig. 22.

Fig. 4.4 shows the SNR for sources at $D_L = 1$ Gpc for Advanced LIGO. Comparing with Fig. 4.3, we see that Advanced LIGO will have a significantly higher sensitivity to BHBs over LIGO. This point is reinforced in Fig. 4.5, which shows contours of SNR for Advanced LIGO as functions of redshift $z$ and total mass $M$. We find that for $M \sim 200 M_\odot$, Advanced LIGO should be able to achieve an SNR greater than 10 out to nearly $z = 1$ for equal-mass nonspinning binaries. From Fig. 4.4 it is evident that these high SNRs depend strongly on the merger-ringdown part of the waveform $t > -50M$.

It is important to note that astrophysical BHBs are likely to have mass ratios different from unity, and that this will reduce the SNRs computed here for the equal-mass case. For stellar BHBs, current work [69] shows that the mass ratios are rather broadly distributed. The rates for such mergers may be low, $\sim 2\,\mathrm{yr}^{-1}$
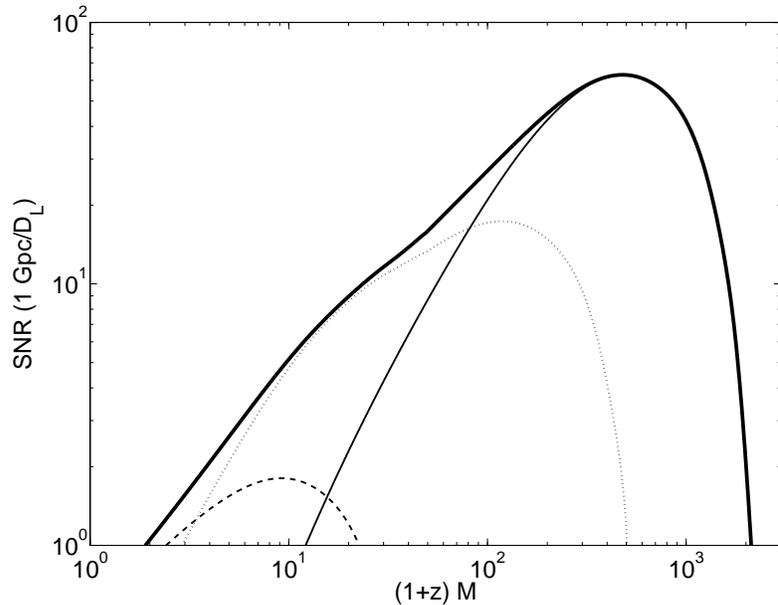
Figure 4.4: SNR for sources at luminosity distance $D_L = 1$ Gpc plotted vs. red-shifted mass for Advanced LIGO. The contributions from $-\infty < t < -1000M$ (dashed), $-1000 < t < -50M$ (dotted), and $-50M < t < \infty$ (thinner solid), as well as the SNR from the entire waveform (thicker solid) are shown.

for Advanced LIGO, depending on the evolution of the original binary through the common envelope phase. For IMBHBs, mass ratios in the range $0.1 \lesssim m_1/m_2 \lesssim 1$ are expected to be the most relevant, with potential rates of $\sim 10$ per year [70], although these rates are far more uncertain than those for stellar BHBs. We can apply the mass scalings from Ref. [65] to extrapolate these results to take into account the effect of mass ratios on the computed SNRs; specifically, SNR $\sim \eta^{1/2}$ for the inspiral based on the leading-order quadrupole behavior, and, more speculatively,
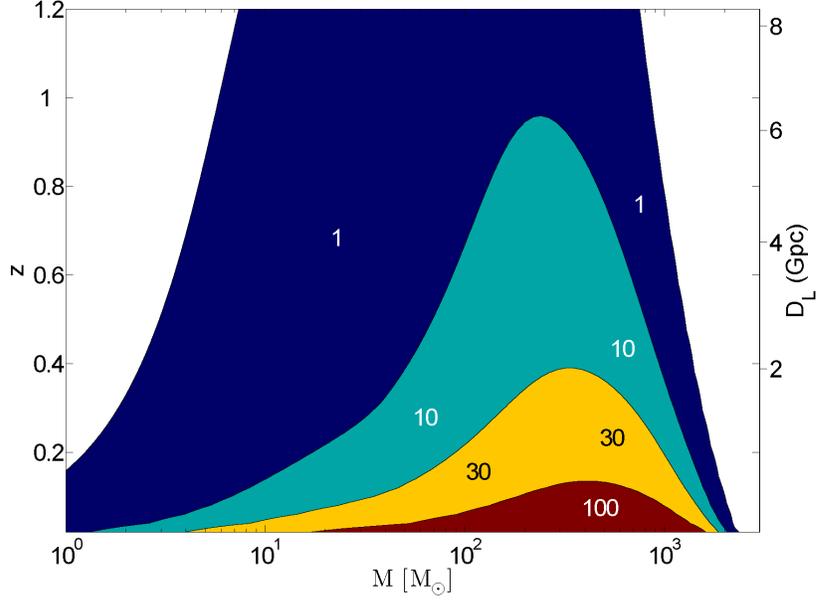
Figure 4.5: SNR contour plot with mass and redshift dependence for Advanced LIGO.

SNR $\sim \eta$ for the merger and ringdown, extrapolating from the test mass limit, where

$$\mu = m_1 m_2 / M \tag{4.4}$$

is the reduced mass and

$$\eta = \mu / M \tag{4.5}$$

is the symmetric mass ratio. The accuracy of these mass ratio scalings will be studied for the case of moderate mass ratios in an upcoming section of this Chapter. Astrophysical BHBs are also expected to be spinning and this can potentially affect the SNR, for example if the spin-orbit interaction moves the ISCO to smaller (or larger) radius and thus causes the binary to that generate more (or less) gravitational

wave cycles in the merger [71].

## 4.2.2   Observing MBHBs with LISA

The coalescing BHBs that radiate in this band will be supermassive (i. e. SMBHB), having masses $M \gtrsim 10^4 M_\odot$. Fig. 4.6 shows $h_{char}$ for several MBHBs plotted relative to the LISA sensitivity curve. We used the "standard" LISA sensitivity curve [72, 73] for frequencies above $1 \times 10^{-4}$ Hz, with shot and pointing noise contributions totaling $20\text{pm}/\sqrt{\text{Hz}}$ of laser phase noise. For $3 \times 10^{-5}\,\text{Hz} \leq f \leq 1 \times 10^{-4}\,\text{Hz}$, we employed a more conservative estimate of the acceleration noise than the one given in [72], instead assuming a steeper amplitude spectral density that falls off as $f^{-3}$ constrained to match the standard sensitivity curve at $1 \times 10^{-4}\,\text{Hz}$ [74]. Below $3 \times 10^{-5}\,\text{Hz}$, we assume the detector has no sensitivity, which is a reflection of the uncertainty of the sensitivity at such low frequencies and our desire to make conservative estimates. The sensitivity model assumes that there are no correlated noise sources, and it assumes perfect cancellation of laser phase noise which results from having an unequal arm interferometer. This cancellation is achieved through the application of time delay interferometry, or TDI (see e.g. [75] for a description of the state-of-the-art in cancellation of LISA spacecraft motion through data postprocessing). In TDI, since the laser phase noise in the different arms is correlated, different combinations of the data streams at various relative delays can be added to form new TDI observables which, if the delay sizes are known precisely, can suppress the laser phase noise to a level below the other noise sources that are independent

among the arms of the interferometer.

MBHB sources can remain in-band for LISA over a very broad frequency range. Therefore, unlike the case of LIGO and Advanced LIGO, LISA sources nearly always require the use of the quadrupole approximation procedure mentioned above to extend $h_{char}$ to sufficiently low frequencies. Also, since more massive BHBs chirp more slowly, a MBHB could potentially be in LISA's sensitive band for much longer than the mission's lifetime. To prevent unrealistic SNR values due to this excessive integration time, the quadrupole formula is only used to extend $h_{char}$ to a low enough frequency such that the total $h_{char}$ used in our calculations corresponds to 3 years of data in the detector's frame, which is a conservative estimate of the expected mission lifetime.

The SNR for LISA is shown as a function of redshifted mass, normalized for $D_L = 10\text{Gpc}$, in Fig. 4.7. The bump in the curves is caused by the binary confusion noise. Again we see the enhancement of SNR from the merger-ringdown part (thin solid line) of the waveforms, and confirm the strikingly large values of SNR obtainable by LISA for these sources seen in [65] and [56]. For systems with redshifted mass $(1 + z)M < 3 \times 10^4$, the early inspiral $t < -1000M$ portion of the waveform dominates. The highest SNRs for equal-mass nonspinning mergers are obtained for systems with $(1 + z)M > 10^6$, again dominated by the merger-ringdown portion of the waveform.

Contours of SNR for LISA are shown in Fig. 4.8 and demonstrate that LISA can observe MBHBs throughout the observable universe at large SNRs. We find it encouraging that, in addition to the large SNR values predicted for LISA overall,
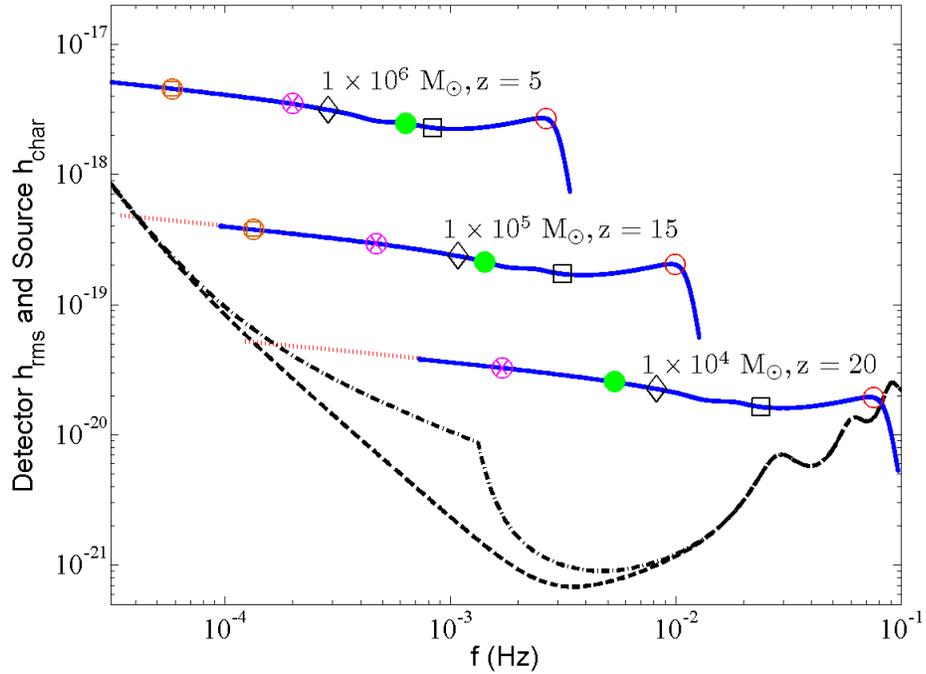
Figure 4.6: LISA rms noise amplitude $h_{rms}$ from the detector only (dashed) and from the detector combined with the anticipated white dwarf binary confusion (dash-dotted) [76] with the characteristic amplitudes $h_{char}$ of three example sources (solid). The locations on each $h_{char}$ curve corresponding to the peak $\psi_4$ amplitude (circle), 1 hour before the peak (filled circle), 1 day before the peak (circle with inscribed cross), and 1 month before the peak (circle with inscribed square) in the observer's frame, as well as $t = -50M$ (square) and $t = -1000M$ (diamond) in the source's frame, are as marked. The mass given is the combined rest mass of each black hole. When necessary, the quadrupole approximation is used to extend $h_{char}$ backward in time 3 years before the peak $\psi_4$ amplitude in the detector's frame (dotted).
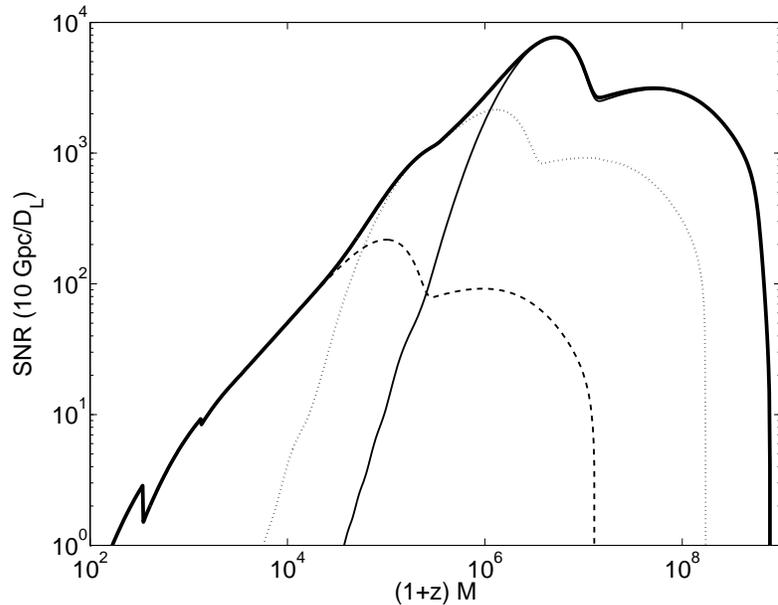
Figure 4.7: SNR for sources at luminosity distance $D_L = 10$ Gpc plotted vs. redshifted mass for LISA. The contributions from the early inspiral $-\infty < t < -1000M$ (dashed), late inspiral $-1000 < t < -50M$ (dotted), and merger-ringdown $-50M < t < \infty$ (thinner solid), as well as the SNR from the entire waveform (thicker solid) are shown.

some of the largest SNR values are obtained out to the largest redshifts in the mass range $10^5 M_\odot \leq M \leq 10^7 M_\odot$ where models of BHB populations predict that the binaries can coalesce within a Hubble time [77] and that the event rates for LISA are several per year [78]. As discussed above, the effects of unequal masses will tend to decrease these SNR values, while spins may increase or decrease them.

Even for non-optimal configurations, the presence of an MBHB coalescence
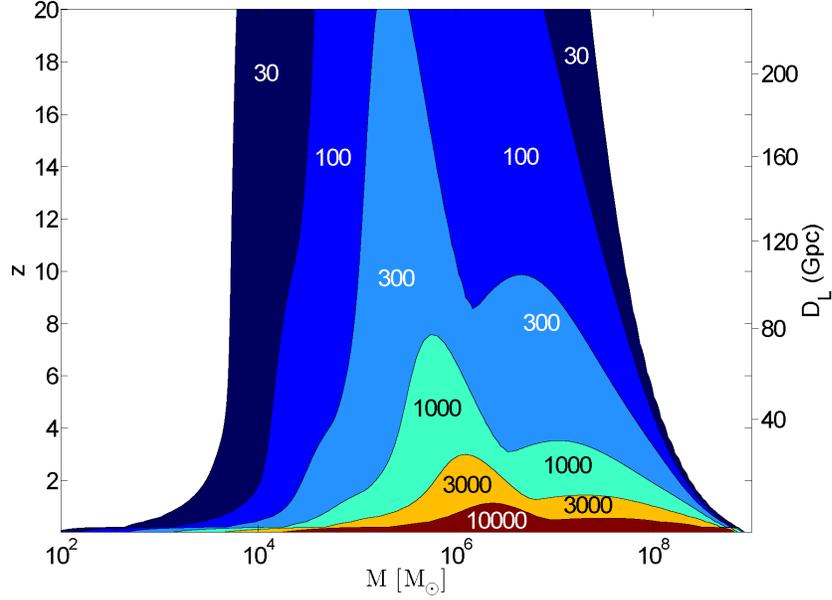
Figure 4.8: SNR contour plot with mass and redshift dependence for LISA. Note that MBHBs with masses $M > 10^7 M_\odot$ may not coalesce within a Hubble time [77].

in the LISA data stream can dominate all the anticipated noise sources. Fig. 4.9 shows a simulation of LISA's response to the merger of equal-mass nonspinning black holes with total mass $M = 10^5 M_\odot$ located at redshift $z = 15$, and oriented so that LISA lies in the system's equatorial plane, where the radiation is weakest. The SNR for a signal from such a source will be $\sim 200$, averaged over sky positions and polarizations (see Fig. 4.8).
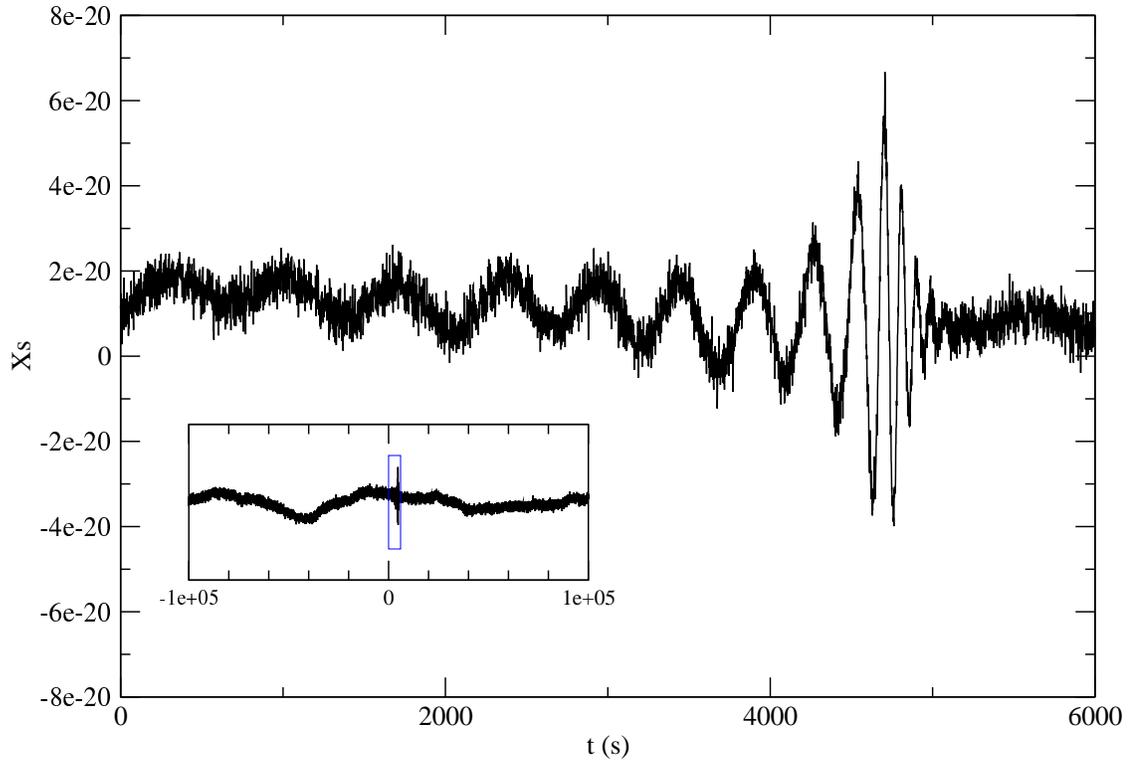
Figure 4.9: Simulated LISA data stream showing LISA's response to a system of two equal-mass black holes ($M = 10^5 M_\odot$) located at redshift z=15 observed on the system's equatorial plane. The quantity plotted is an unequal arm Michelson interferometer observable "X" [75]. The LISA response and instrumental noise are realized using the LISA Simulator [79, 80], and colored noise was added to represent the unresolvable galactic binary foreground with the spectrum used in Ref. [76]. The inset shows the signal over a longer duration where low-frequency noise is evident.

### 4.2.3    Unequal Mass and the Significance of Higher Modes

It has long been known that the peak of signal-to-noise ratio (SNR) along the $\eta$ axis of parameter space occurs for equal mass binaries due to the decrease

in the power content of the emitted waves as $\eta$ deviates from 0.25. However, the higher frequency throughout the evolution of an unequal mass coalescence compared to an equal mass system the same amount of time before merger might lead one to believe that unequal mass systems will yield smaller parameter uncertainties. Furthermore, the presence of subdominant modes in the unequal mass cases brings about the possibility that those modes might break degeneracies among the other parameters that describe the system, degeneracies that are left untouched in the equal mass case. However, the increased number of cycles and presence of higher modes is counterbalanced by the decrease in SNR as mass ratio increases, such that for nonspinning binaries, parameter accuracy generally decreases as we move away from the equal mass case [81]. Therefore, we will focus on questions regarding waveform power and SNR, particularly during the merger, with the assumption that these questions are the most relevant both for detectability and parameter estimation.

In order to compare our data with any prior PN-inspired expectations, we must first generate a complete waveform. We proceed in a similar fashion to what we did in the equal mass case, by first truncating the numerical $h_{22}$ waveform at a point after the initial burst of transient noise has passed. We then attach a PN $h_{22}$ waveform with phasing determined by numerical integration of an expansion of $\dot{\omega}$ in powers of $\omega$, just as was done in the preceding section, the only difference being that the joining point is simply $\sim 150M$ after the start of the numerical simulation, rather than at a point determined by rigorous relative error analysis, which is due for the most part to the relative brevity of some of the unequal mass runs. We
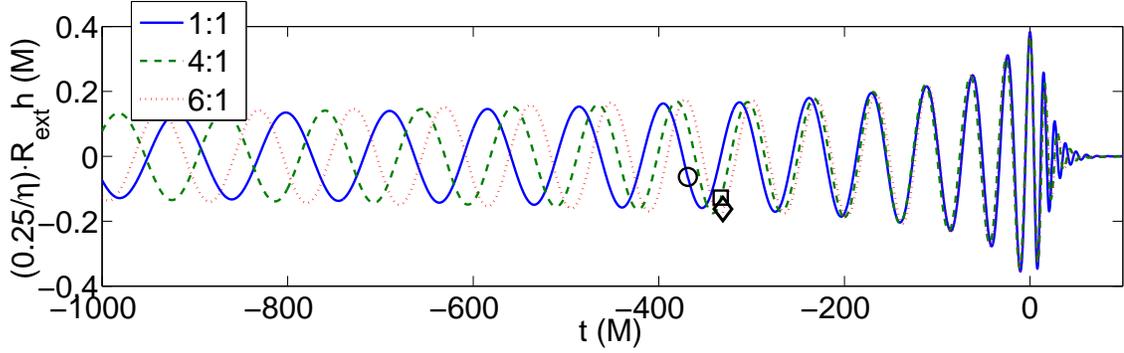
Figure 4.10: Waveforms generated by combining numerical data and PN waveforms. The circle, square, and diamond show where two parts were tied together for the equal mass, 4:1, and 6:1, respectively. Note the striking overlap of the numerical waveforms for all three cases for the final $\sim 6$ cycles leading up to merger. Although the relative frequencies during inspiral and at ringdown required some frequency crossover among the three simulations, the prolonged agreement was unexpected, and is the subject of current investigation.

also use the 3PN amplitude in this analysis, unlike in the equal mass-only analysis, but we add an *ad hoc* 3.5PN term with a constant coefficient chosen to make the amplitude continuous at the matching point, thereby minimizing artifacts in the Fourier transform. Fig. 4.10 shows the end result for three cases, the equal mass, $q = 1/4$, and $q = 1/6$, where $q \equiv m_1/m_2$. To avoid confusion, we will refer to the unequal mass runs as ratios, i.e. the $q = 1/4$ run will be the 4:1 run. The circle, square, and diamond show where the PN and numerical waveforms were tied together for the equal mass, 4:1, and 6:1 runs, respectively.

In the case of equal mass, we could immediately move on to calculating, for instance, sky-averaged SNRs, since the only nonzero mode is the $(2,2)$. However, since other modes are excited for unequal mass mergers, we must either include them as well, or demonstrate that they are negligible for our purposes. We can compare the relative contribution of $h_{2,2}$ to the overall $h$, where $h$ is given by

$$h = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{\ell} h_{\ell m}(t,R) \, _{-2}Y_{\ell}^{m}(\theta,\phi). \tag{4.6}$$

As a measure of the relative contribution, we use

$$\mathcal{M} = \frac{\langle h_{2,2} | h_{\text{all modes}} \rangle}{\sqrt{\langle h_{2,2} | h_{2,2} \rangle \langle h_{\text{all modes}} | h_{\text{all modes}} \rangle}} \tag{4.7}$$

where $\mathcal{M}$ is the "match". The value on the right hand side can be viewed as the fraction of the SNR for which the $(2,2)$ mode is accountable. The "$\langle \cdot | \cdot \rangle$" that we apply are simple, unweighted inner products, but could equally well be weighted by the noise spectral density of a given detector. Fig. 4.2.3 shows a typical comparison for the 6:1 case, which should have the strongest higher harmonics among the cases studied here. The $(2,2)$ mode accounts for 0.9845 of the total signal as measured by $\mathcal{M}$ for white strain noise. Also, we show a frequency-based comparison for the 6:1 in Fig. 4.2.3, where the $(2,2)$ mode can be seen to dominate until well into ringdown, which is denoted by a vertical dashed line. It should be noted that this is an analysis of power and phase agreement over just the late inspiral and merger, and that the degree of overlap is only adequate for detection even with this constrained range of application. The degree of overlap might be worse if we compared a long inspiral

110

attached to the merger with and without ignoring extra modes.

## 4.3   Testing Expectations for Unequal Masses

Now that we have established that the $(2,2)$ mode is sufficient at these mass ratios for investigating power-related observables, we can explore how closely the Fourier amplitude of the signal follows the expected behavior for the late inspiral and subsequent different behavior for the merger. Namely, we know that the inspiral scales as $\sqrt{\eta}$ to leading order, and that it was assumed in [65], based on the prediction for total radiated energy in the test particle limit [82], that the merger scales as $\eta$. Fig. 4.12 appears to validate that assumption, as the top panel demonstrates the (not surprising) $\sqrt{\eta}$ scaling of the inspiral ($M\omega \lesssim 0.08$) for the equal mass, 2:1, 4:1, and 6:1 cases, and the bottom panel shows the clear $\eta$ dependence for the merger ($M\omega \gtrsim 0.08$).

Since we have established that $h(t, r, \theta, \phi) \approx h_{2,2}(t, r)$, at least to a fidelity necessary for detection, we can now apply the same procedure for calculating sky-averaged SNR that was done earlier for the equal mass case, rather than having to explicitly calculate the SNR at N sky locations explicitly and average the result. The procedure, then, is to make the above approximation, then calculate the sky- and waveform-polarization-averaged SNR using eq. (4.1), As was mentioned, the SNR falls off as $\eta$ deviates from 0.25, i.e. equal mass. Therefore, to show the impact of changing $\eta$, the 6:1 case provides the clearest distinction among the cases studied. Fig. 4.13 demonstrates the significant decrease in SNR with such a significant
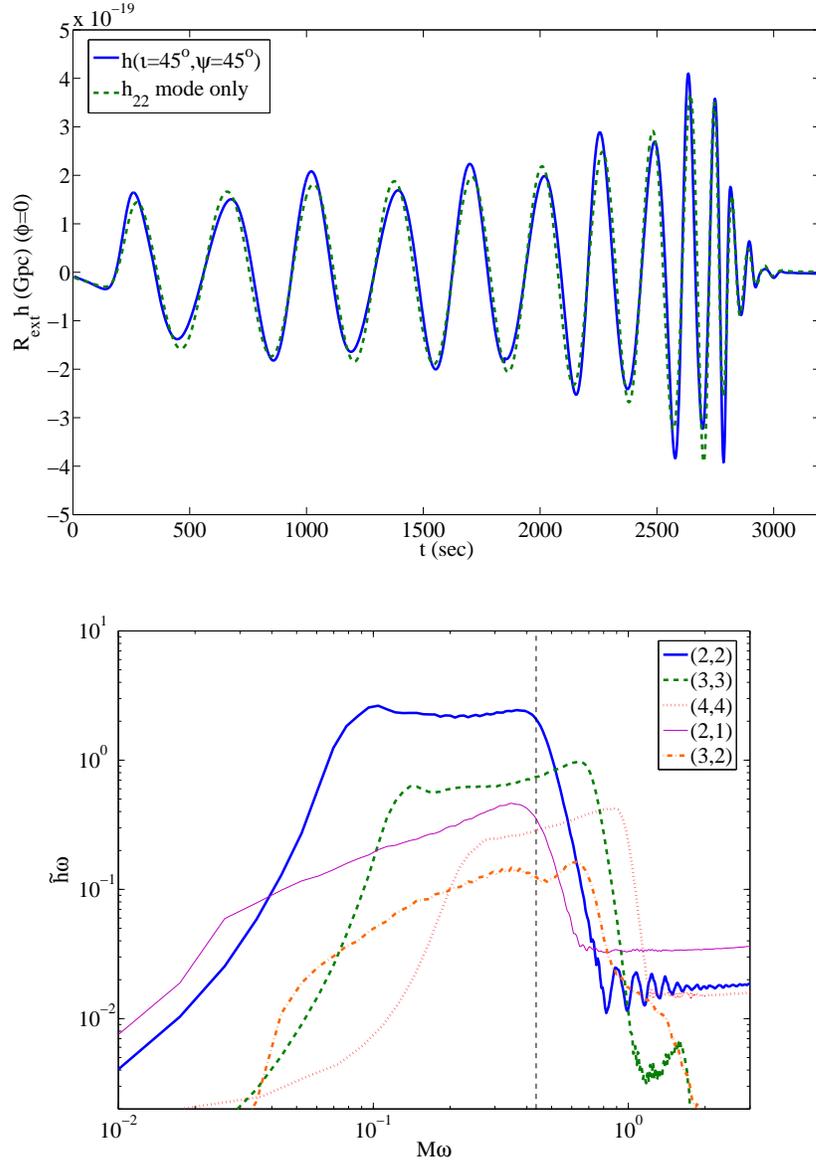
Figure 4.11: Time series and Fourier series representations showing a typical case of the $(2,2)$ component constituting the vast majority of the overall power content of the waveform. The case being shown is a $10^6 M_\odot$ SMBHB at a distance of 10 Gpc. To get a true sky average we would need to calculate the signal all over the sky for unequal masses. $\mathcal{M}$ defined by eq. (4.2.3) for this case is 0.9845. This is further demonstrated in the bottom panel, where the $(2,2)$ dominates until well into the ringdown, designated by a dashed vertical line.
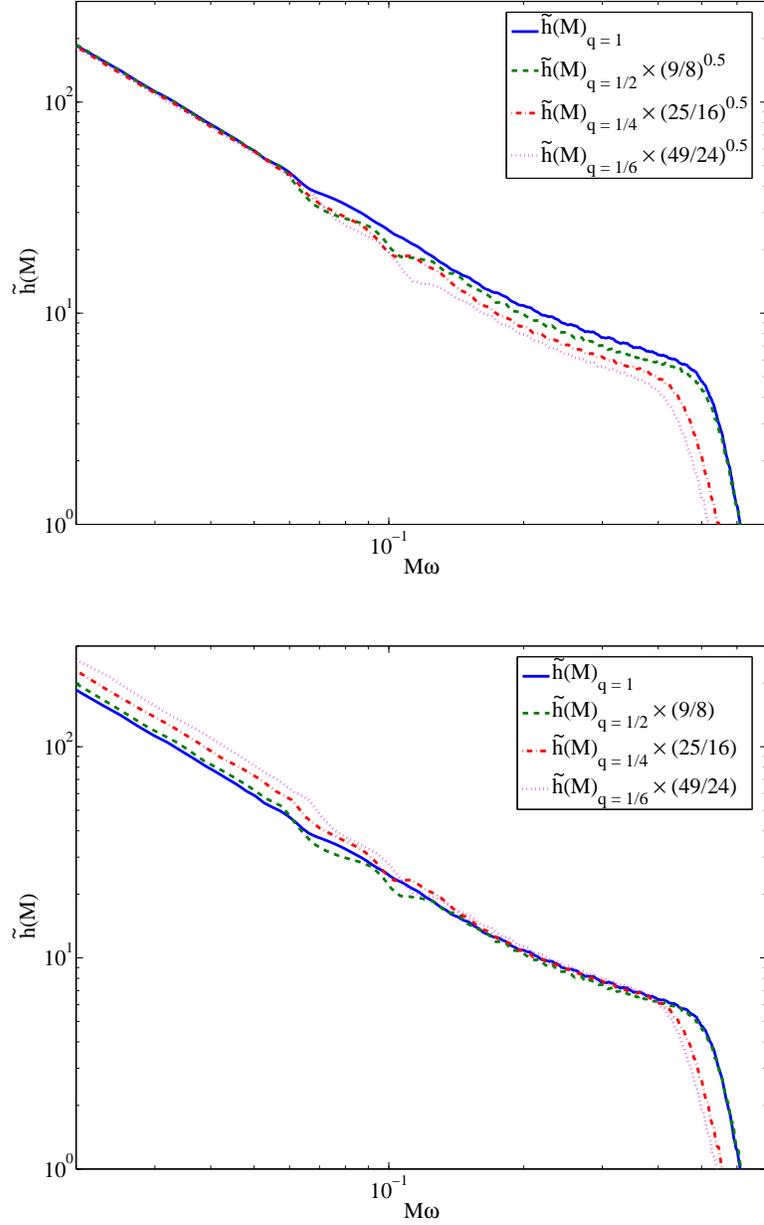
Figure 4.12: Scaling of the Fourier amplitude for the different mass ratio cases $q = 1$, $1/2$, $1/4$, and $1/6$. The amplitudes are scaled by $\sqrt{\eta}$ in the top panel to show the anticipated agreement of the inspiral, and by $\eta$ in the bottom panel to show the agreement during the merger. The $\eta$ scaling was assumed in the SNR calculations in [65] based on the prediction for total radiated energy in the test particle limit [82], and appears to be an excellent approximation.

deviation from equal mass. Both panels show contour lines for both the equal mass and the 6:1 cases, with the top panel showing contours corresponding to Advanced LIGO, and the bottom panel corresponding to LISA. The relative SNR in each panel does not scale as a constant across all masses, as one would expect if either only the inspiral or only the merger was contributing, but rather scales as $\sqrt{\eta}$ for low masses where the inspiral matters most, and $\eta$ for higher masses where the merger contributes the majority of SNR.

## 4.4   Parameter Estimation with LISA

Questions of detectability are the primary interest for LIGO, but for LISA, detectability of the signals is typically not in question. For LISA, the goal is to be able to accurately extract parameters. All of the work on parameter estimation using LISA that has been done to date has focused exclusively on the inspiral waveform and neglected the merger. This is not without justification given that, until recently, only the inspiral and some aspects of the ringdown waveforms were known, and, perhaps more importantly, the merger is such a fleeting event that, despite its very high SNR, it occurs too fast for LISA to move significantly in its orbit, so the Doppler modulation due to LISA's motion around the Sun, which helps to break parameter degeneracies during the inspiral, is absent for the merger. However, mergers, though brief, contain substantial power and span roughly an order of magnitude higher in frequency beyond the end of the inspiral. If the added information from the merger breaks degeneracies between observables that can be matched well locally, then

114
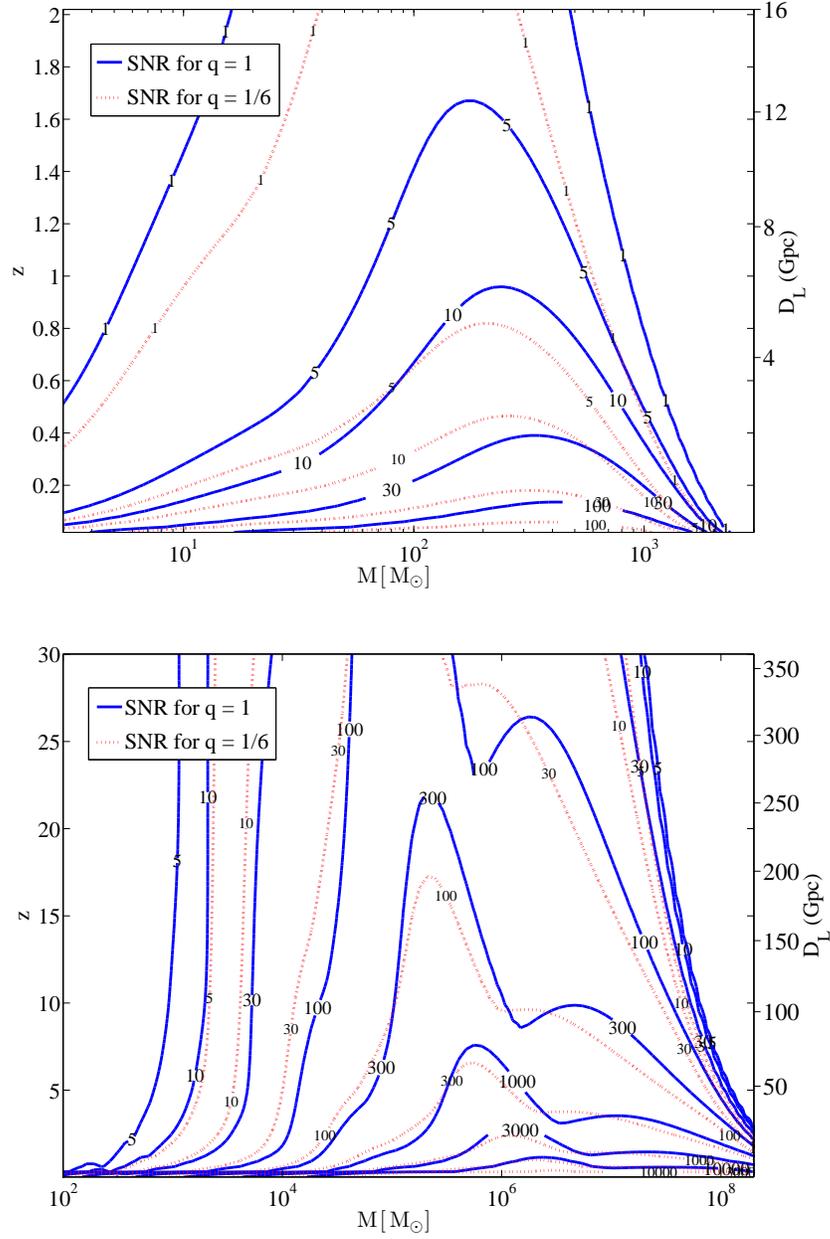
Figure 4.13: SNR contours for Advanced LIGO in the top panel and LISA in the bottom panel. The solid lines with larger value markers correspond to $q = 1$ for both figures, while the dotted lines with smaller markers correspond to $q = 1/6$. In both cases, the late inspiral-merger phase constitutes the majority of the SNR [65, 2], so that the $\sqrt{\eta}$-to-$\eta$ scaling transition demonstrated in Fig. 4.12 is apparent as one moves to larger masses, thus emphasizing the merger more so.

the smaller uncertainty in those observables can affect their covariance with other observables, thereby propagating improved accuracy. Furthermore, at high frequencies, where the wavelength of the gravitational radiation becomes comparable with the arm length of the interferometer, the short wavelength approximation breaks down, and the gravitational wave transfer function begins to vary non-trivially with frequency (see Fig. 4.14). This admits the possibility of non-negligible interaction of the incident waveform with the varying transfer function given the right mass and redshift combination. Whether the transfer function might act as a surrogate for the Doppler modulation is an open question, but may prove less critical if mergers help break degeneracies in general, since the transfer function only comes into play for a limited mass range.

To approximate the measurement accuracy that LISA can achieve, one approach we can take is the Fisher matrix formalism. If you have a waveform, $h(\lambda^i)$, embedded in a signal, $s$, so that the noise, $n$ is given by $n = s - h$, then the probability that a signal contains a waveform with the parameter set $\tilde{\lambda}^a$ is given by

$$p(\tilde{\lambda}^a|s) \propto e^{-(h(\tilde{\lambda}^a)-s|h(\tilde{\lambda}^a)-s)/2} \tag{4.8}$$

where $(\cdots|\cdots)$ is a noise weighted inner product [83]. The "maximum likelihood" set of parameters, $\hat{\lambda}^a$, are those that maximize $p$. Errors in the $\hat{\lambda}^a$ set of parameters can be assessed by expanding $p$ around $\hat{\lambda}^a$, such that

$$p(\tilde{\lambda}^i|s) \propto e^{-\Gamma_{ab}\delta\lambda^a\delta\lambda^b/2} \tag{4.9}$$

where $\delta\lambda^a \equiv \tilde{\lambda}^a - \hat{\lambda}^a$. The Fisher information matrix, which is the centerpiece of
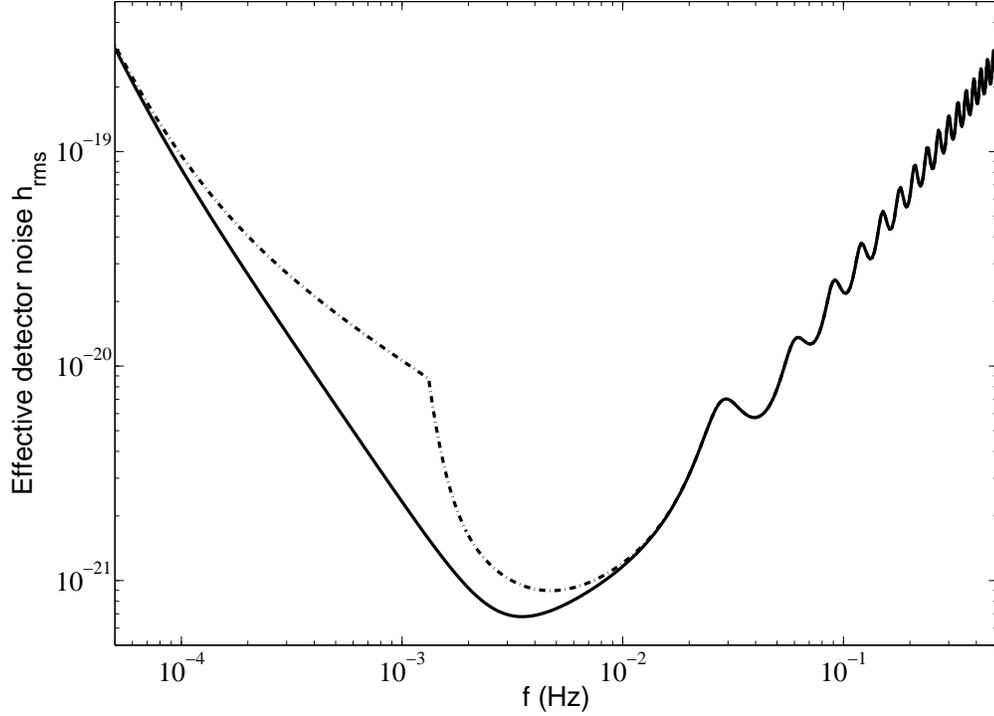
Figure 4.14: Effective strain spectral density of the noise in LISA. This data is not strictly noise, however, as it includes a bumpy, upward trending component at high frequency which is due to the inclusion of (and division by) the signal response, which remains flat until $\sim 0.01$ Hz, but then begins to trend downward and shows its nontrivial structure. The response was calculated using the LISA Sensitivity Curve Generator [72, 73].

our subsequent analysis, is defined to be

$$\Gamma_{ab} \equiv \left( \frac{\partial h}{\partial \lambda^a} \,\middle|\, \frac{\partial \mathbf{h}}{\partial \lambda^b} \right), \tag{4.10}$$

where $a$ and $b$ are parameter indices. To lowest order in an expansion in $\mathrm{SNR}^{-1}$,

117

the covariance matrix for the errors is just the inverse of the Fisher matrix:

$$\Delta\lambda^a \, \Delta\lambda^b \equiv \langle\delta\lambda^a\delta\lambda^b\rangle = \left(\Gamma^{ab}\right)^{-1}\left[1 + \mathcal{O}(\text{SNR}^{-1})\right], \qquad (4.11)$$

where "$\langle\cdots\rangle$" means "expectation value" and $\Delta\lambda^a$ is the standard deviation of parameter $a$. The covariance matrix is symmetric, with the off-diagonal terms giving the covariance between parameters, and the diagonal terms giving the variance of each parameter.

For this investigation, since we wished to use the most accurate waveform possible, we attached T4PN and our high resolution run at a time, $t = -166M$, which corresponds to the point where the numerical waveform has reached $M\omega = 0.1$, so that we know, as was pointed out in the preceding chapter, that our accumulated phase error will be $\sim 0.3$ rads over more than 30 orbits leading through merger and ringdown, 0.05 rads being due to inspiral error from the T4PN, and the other 0.25 rads being due to our numerical simulation from the late inspiral starting at $M\omega = 0.1$, through the merger and ringdown (see Fig. 3.21).

In order to carry out a meaningful assessment of the statistical errors, as mentioned, we need to generate a TDI observable (or set of observables) to run through the analysis pipeline. To make a set of TDI observables, we used components from the software package *Synthetic LISA* [84], which includes a full model of the dynamics of the instrument, the application of TDI, and the response to incident gravitational wave signals. Specifically, *Synthetic LISA* generates a set of observables referred to as the Michelson unequal-arm $X$, $Y$, and $Z$ TDI observables [84]. However, this set of TDI observables is not linearly independent as we would prefer,

so we then convert the $X$, $Y$, and $Z$ to a linearly independent set we refer to as the pseudo- $A$, $E$, and $T$ variables.Using nearly the same combination employed in eq. 18 of [85] to construct the original $A$, $E$, and $T$ variables from the 1st generation TDI observables $\alpha$, $\beta$, and $\gamma$, we construct the pseudo- $A$, $E$, and $T$ via the relations

$$\bar{A} = \frac{Z - X}{2\sqrt{2}} \tag{4.12}$$

$$\bar{E} = \frac{X + Z - 2Y}{2\sqrt{6}} \tag{4.13}$$

$$\bar{T} = \frac{X + Y + Z}{2\sqrt{3}} , \tag{4.14}$$

where a bar indicates a "pseudo" quantity. The only difference we apply is an overall factor of $\frac{1}{2}$ to all three formulae to make $\bar{A}$ and $\bar{E}$ agree with $A$ and $E$ in the low frequency limit.

In order to be consistent with our usage of a different set of TDI variables, we must also apply the corresponding noise response. Following the procedure provided in [85, 86] we generate a set of expressions for the pseudo- $A$, $E$, and $T$ in terms of the proof mass and optical path fractional-frequency-fluctuation spectral densities, denoted $S_{pm}$ and $S_{op}$, respectively, which are given by

$$S_{\bar{A},\bar{E}} = 2\sin^2(\Phi)2[3 + 2\cos(\Phi) + \cos(2\,\Phi)]S_{pm} + (2 + \cos(\Phi))S_{op} \tag{4.15}$$

$$S_{\bar{T}} = 8\sin^2(\Phi)\sin^2(\Phi/2)(4\sin^2(\Phi/2)S_{pm} + S_{op}) , \tag{4.16}$$

where $\Phi \equiv \frac{\omega L}{c}$ and $L$ is the average of the arm lengths, expressed as a light-travel time. These expressions are the analog of eqs. 67 and 68 of [86] and eqs. 19 and 20 of [85] for the noise response of the original $A$, $E$, and $T$ variables, and we have verified that our script for processing the procedure detailed therein duplicates those results

[87] (accounting for a typo which we believe appears in [85] which, if corrected, would make it consistent with [86] and with our results). One subtlety that arises is that, if we allow perfect TDI cancellation of the laser phase noise, then the response to noise for our choice of observables goes to zero at certain frequencies. We then are left with an undefined or divergent SNR and set of elements in the Fisher matrix. To remedy this, we add a small amount of laser phase noise, the equivalent of allowing slightly unequal arms, or putting a floor on the noise response.

The result is a set of observables which we can now use for parameter estimation studies. We note that while we hope that adding the merger will decrease the statistical uncertainties due to the presence of noise, it will most certainly *increase* the theoretical uncertainties due to having an imperfect template. Furthermore, since it has already been estimated [88] that the theoretical uncertainties from the PN portion of the waveform are already approximately the same magnitude as the statistical errors, we are assured that the theoretical errors of our combined PN-NR waveform are larger that the statistical errors. Therefore, we must be clear that we are not testing actual templates that are sufficiently accurate to extract parameters with the estimated level of certainty. Rather, we are applying the methodology and assessing the level of statistical uncertainty that LISA can reasonably expect to be able to achieve once the theoretical errors have been adequately suppressed.

Since there are fluctuations in the parameter uncertainties depending upon the location in parameter space where you calculate the results, we perform a Monte Carlo simulation, evenly distributing the sampling of all parameters except the total mass and the luminosity distance, which are kept fixed at $M = 2 \times 10^5$ M$_{odot}$ and $D =$
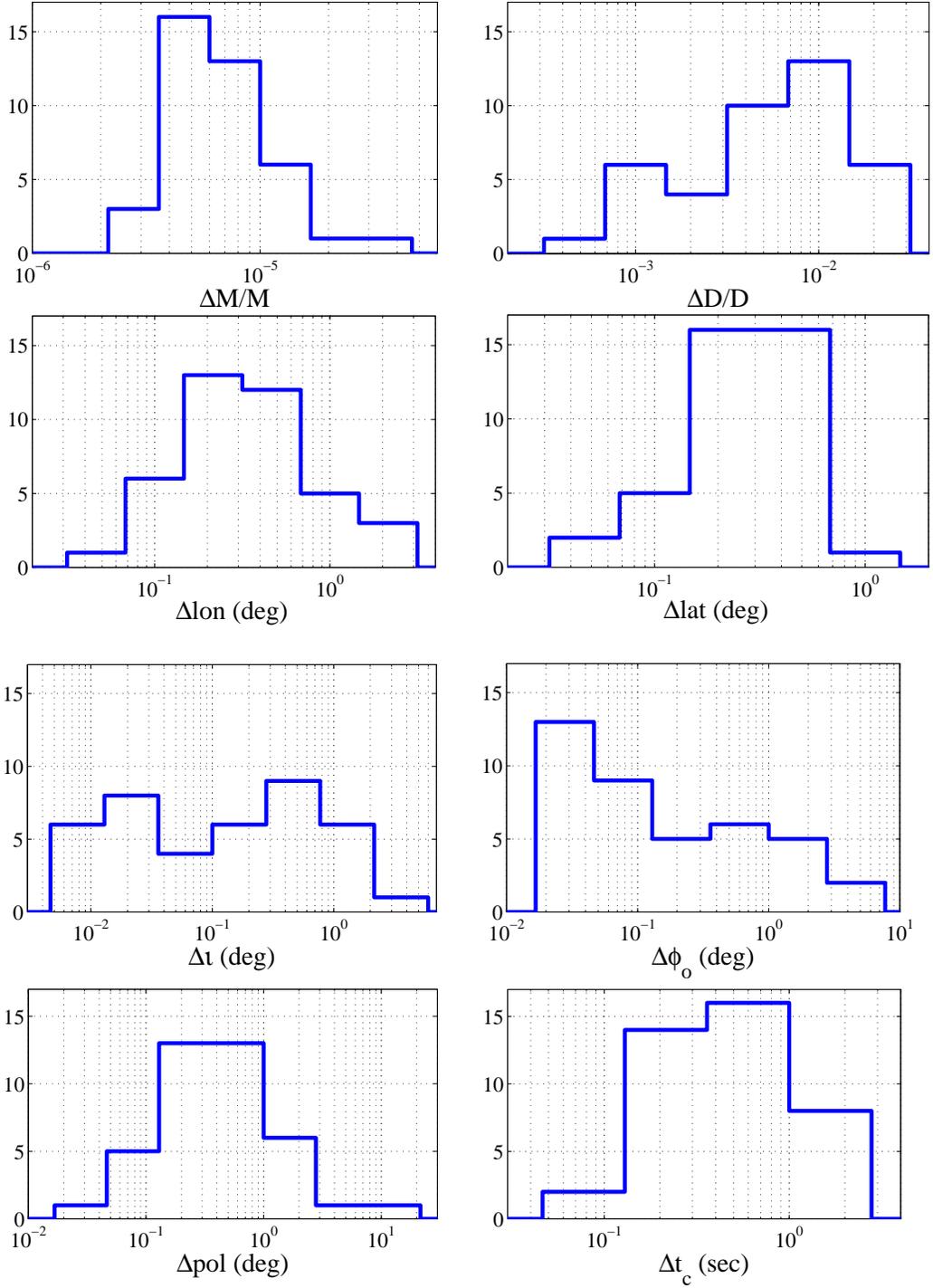
Figure 4.15: Parameter uncertainties for the full inspiral-merger-ringdown waveform calculated using the Fisher matrix. The data shown is the result of 40 Monte Carlo trials. All histogram plots run from 0 to 17, to better allow visual comparison.

| Parameter | $\langle Var(X) \rangle$ |
|:---:|:---:|
| $\Delta M/M$ | 7.84e-6 |
| $\Delta D/D$ | 7.80e-3 |
| $\Delta lat$ | 0.317 deg |
| $\Delta lon$ | 0.495 deg |
| $\Delta \iota$ | 0.368 deg |
| $\Delta \phi_o$ | 0.573 deg |
| $\Delta pol$ | 0.831 deg |
| $\Delta t_c$ | 0.605 sec |

Table 4.1: Comparison of mean fractional variance of all the extrinsic parameters for the full inspiral-merger-ringdown waveform, which combines PN through most of the early time evolution with NR in the late inspiral though merger where it becomes more accurate.

168 Gpc (corresponding to $z = 15$), respectively. The results from our Monte Carlo run, shown in Fig. 4.15, yield predicted uncertainties which compare favorably with available results in the literature (e. g. [81]) for cases which studied the parameter uncertainty from equal mass, nonspinning systems, but which only included the PN waveform, truncated at some point near ISCO. Our own investigations of the relative performance of the full inspiral-merger-ringdown waveform compared to truncated waveforms intended to represent a PN-only attempt have borne out that

the inclusion of the late inspiral-merger that NR provides substantially improves the predicted parameter accuracy. However, there is ambiguity in how one treats the PN-only contribution, and the hard divisions between the inspiral and merger phases that were once anticipated have turned out to be non-existent.

Ideally, perhaps, we might use a PN method until ISCO and compare, but a time series will need to be windowed to avoid introducing artifacts in the Fourier transform that might dominate the parameter estimates. The stationary phase approximation (SPA), which is already posed in frequency and so avoids windowing, will nonetheless likely be inaccurate, since its requirement that the frequency change adiabatically is invalid well before ISCO. In addition, since our investigation of the full waveform involves using *Synthetic LISA* to apply the detector response to an input time series, using the SPA would mean applying a separate model of the response that takes frequency input directly, which is not ideal for side-by-side comparison (although we are implementing this for future investigations). Therefore, without having performed a comprehensive study of the dependence of parameter uncertainties on the various possible treatments of a waveform without numerical relativity content, we refrain for the moment from presenting a direct comparison, stating only that the merger waveform provided by numerical relativity appears to significantly improve the predicted measurement accuracy, with the level of that improvement depending on the divisions of "inspiral" and "merger", or "PN contribution" and "NR contribution", and the particular type and treatment of PN inspiral chosen.

We note that the total phase error estimate of $\sim 4$ rads found in Chapter 3

is larger than the initial phase, $\Delta\phi_o$, and the polarization phase, $\Delta pol$, which are the two phase parameters that should be most directly affected by errors in the template phase, and which are both generally several tenths of a radian. Both error estimates from Chapter 3 do not take into account a detector response (and therefore a limiting range of frequency over which to integrate), so the result for phase error in Chapter 3 will certainly be an overestimate, while the result for fractional amplitude in Chapter 3 is a localized measure of the maximum disagreement, and is therefore also an overestimate, since the correct measure would effectively find the average fractional amplitude weighted by the detector response, rather than the maximum value. However, given the observations in [88], we would expect that all our theoretical uncertainties will ultimately exceed the statistical uncertainties with the current state-of-the-art PN and any addition of a merger from numerical simulations, in agreement with our estimates.

Consequently, theoretical errors appear to be the current limitation to accuracy. Despite that, it is not unreasonable to expect the shape of the nearby parameter space around the correct values of the luminosity distance, polarization phase, and initial orbital phase to be the same or similar as it is around the theoretical values which are incorrect. Furthermore, we do not expect the theoretical errors to have a net bias. Therefore, by performing the Monte Carlo simulation and sampling throughout parameter space, we have been able to construct distributions for the errors, and if the perturbations of those statistical errors caused by the theoretical errors do not introduce a bias in the statistical uncertainties, the expectation values for the statistical uncertainties will not be shifted by the presence of theoret-

ical error. We therefore have reason to expect that the predicted statistical errors may be reasonable approximations, and may not change significantly once improved templates are applied.

Chapter 5

Summary and Conclusion

5.1   Review of Results

In this work, we analyzed simulations of equal-mass nonspinning BHBs start-ing in the late inspiral regime and covering approximately an additional factor of three in frequency before the merger-ringdown. We carried out runs at three resolutions, $h_f = 3M/64, 3M/80$, and $M/32$. Our runs start with relatively low eccentricity and show good convergence and conservation properties. We have demonstrated the stability and accuracy of our simulations over the course of seven orbits. We also showed the value of using frequency (rather than time) to set a reference for the purpose of comparing results between runs as well as with the PN approximation. In recasting phase vs. frequency we have found particularly good agreement, not only between the runs but also with PN predictions.

We made use of the p4EOB model [12] in order to provide a novel method for removing eccentricity from runs. The previous method of fitting a fourth order polynomial for the frequency trend allowed for the possibility that the polynomial would "wiggle", locking on to some of the eccentricity and effecting the sinusoidal part of the fit that would ideally represent the eccentricity. However, using the non-eccentric p4EOB for the frequency trend ensures that all the oscillatory content will be left behind for the sinusoidal part of the fit. Furthermore, the p4EOB gives

a natural, physically motivated method of time alignment by aligning the peak radiation amplitude with the peak of the p4EOB frequency prediction, which is roughly the p4EOB prediction for the location of the light ring. We also look at the frequency from the puncture track, aligning it by time-shifting based on the peak radiation amplitude as well as the light travel time to the extraction radius, for which we simply use the coordinate value for $R_{ext}$ and find excellent alignment. The eccentricity from the two measurements agrees very well, as we would expect.

We proceed to analyze the puncture track phase and frequency predictions, similarly to what we had previously done for the radiation [1]. For this study, we propose a new metric for comparing the phase error, which we refer to as the maximized phase error, which is constructed not to permit cancellation of accumulated phase error. We use this to assess the relative error of the two methods, and we find that, looking at both the high resolution and Richardson-extrapolated cases, the radiation phase outperforms the puncture track phase over the frequency range that we observe, although only by $\sim 0.5$ rads in both cases. Since the puncture track phase is a gauge quantity, even if it yielded a more accurate result in this one case, that would certainly not guarantee its continued accuracy throughout parameter space. However, the fact that our improvements in moving puncture gauges appear to also improve the utility of the gauge phase as a measurement tool is a point of interest. The likelihood that the puncture phase will ever become more accurate through some perfect choice of gauge is small, and, in fact, if it did happen it seems likely that the community would be hesitant to accept a gauge quantity as a measurement. The gauge that makes it possible would likely have to be understood

127

in an analytic sense to be accepted. For the moment, however, the puncture phase has been shown to be comparable, though slightly worse, but its measurement has value as a check of the radiation phase, or as a tool, combined with the radiation phase, to try and estimate sources of error, since each method has some different error sources which impact their value.

We have also matched the numerical gravitational waveforms to the T4PN inspiral. We have demonstrated that care must be taken when extrapolating waveforms using extraction radius, and that it yields poor results for the late inspiral and merger. We have tested our accuracy using extrapolation in resolution only, combined the numerical merger with a PN inspiral, and found that the resulting waveform has less than 3/4 cycle of accumulated phase error over its entire frequency band.

Using this waveform, we calculated the SNRs for LIGO, Advanced LIGO, and LISA. Our results confirm the importance of the merger-ringdown signal, which yields the highest values of SNR for the majority of equal-mass signals [65, 56]. We also show the SNR for the late inspiral regime, which numerical simulations are now beginning to address. The late inspiral dominates the SNR for LIGO and Advanced LIGO for the lower mass ($\lesssim$ a few $\times 10 M_\odot$) stellar BHBs, and the SNR for LISA generated by MBHBs with $M \sim 10^5 M_\odot$. Contour plots of SNR as a function of $z$ and $M$ show that Advanced LIGO can achieve SNR $\gtrsim$ 10 for IMBHBs out to nearly $z \sim 1$, and that LISA can observe MBHBs at SNR $> 100$ out to the earliest epochs of structure formation at $z > 15$.

We looked at several unequal mass waveforms, and established the level of im-

portance of the presence of higher modes once we deviate from equal mass. We also calculated Fourier amplitudes, and used them to confirm the predicted scaling of the merger power with mass ratio. Since the predicted scaling came from extrapolating the test mass limit to the comparable mass case, we were by no means assured that our answer would agree with the predicted scaling, but we found that it does to a high degree of fidelity.

Lastly, we explored the parameter accuracies that LISA can hope to attain from our waveforms, in particular the impact of including mergers. We do note, however, that the fractional accuracy in the luminosity distance is smaller than the fractional accuracy of the waveform's amplitude as estimated in Chapter 3, and the total phase error estimate of $\sim 4$ rads found in Chapter 3 is larger than the initial phase, $\Delta \phi_o$, and the polarization phase, $\Delta pol_o$, which are the two phase parameters that should be most directly affected, and which are both generally several tenths of a radian. These estimates are consistent with [88], which finds that for PN inspiral waveforms (i. e. without the merger contribution from numerical relativity) yields approximately equivalent sizes of theoretical and statistical uncertainties. Adding the merger increases the theoretical uncertainty and decreases the statistical uncertainty, so that the result that including the merger yields larger theoretical errors than statistical errors for the parameters is not surprising.

## 5.2 Perspective

This work is an accumulation of our ongoing effort to try and take the output from the fields of PN and NR, combine them in a reasonable way, and begin applying them to real questions of detectability and parameter accuracy. When this work began, there was no such material in the literature, but as the work has continued, the body of literature has grown to include several of the results in this paper which have already been published, as well as critical contributions from other groups.

For instance, in a relatively short time span, we used the T1PN phase to approximate the duration of a numerical simulation that would be required to validate the PN phase approximation in the strong field regime [64]. The technical hurdles required to perform such a lengthy simulation were overcome, and a run of sufficient length and accuracy to validate the PN phase was achieved (the results of which are presented in Chapter 3 and can also be found in [1]). The PN amplitude, which is more difficult to validate due in part to the larger impact of numerical dissipation on the measured amplitude than on the phase of the numerical waveforms. However, the necessary level of accuracy was achieved by the numerical relativity group located in Jena, Germany, and they were able to conclusively demonstrate the accuracy of the PN amplitude in the strong field regime [61], as we had done for phase, thereby providing a complete validation of PN gravitational waveforms. Finally, subsequent work by the group at Caltech, using completely different (and considerably more accurate) numerical methods than either our group or Jena (they use a generalized harmonic evolution system instead of BSSN, they employ spec-

tral, rather than finite, differencing, and they use dual coordinates frames, described in [89]) yielded exceedingly good agreement between several PN variants and the numerical simulations until very late in the inspiral [59] (Caltech did not, at that time, have the capability to successfully evolve through merger). In particular, they found that the T4PN phase, which we had chosen as our benchmark for validating PN based on previous observations regarding its good qualitative agreement [56], agreed to within 0.05 rads with their simulation over a range of 30 orbits leading into the late inspiral just before merger.

Already, the field has moved on to spinning binaries which, depending on their configuration, undergo massive "kicks", meaning they are imparted with linear momentum which is transferred via conversion from the radiation. Work continues on attempting to make the codes accurate and efficient enough to do ever-larger mass ratios, and to combine mass ratios with spin in order to be capable of evolving BHB systems which are truly generic, and hopefully ever-more astrophysical. On the PN front as well, progress in numerical accuracy and the desire to test the agreement with PN undoubtedly played a role in prompting work on the derivation of the current highest order expression for quadrupole radiation amplitude, which was presented in [63] after it had already been applied in [59] for the purpose of providing a comparison with numerical results. Across many different groups presently active in numerical relativity, there has been an extremely fruitful crossover of talent with members of the PN community, with the two sides sharing expertise, helping to drive one another's progress, and generating a torrent of useful analysis from the new results being turned out from numerical simulations.

## 5.3   Directions for Future Work

With the data already available, a number of projects can and should be undertaken in the near future. In the near future, we plan to assess the dependence of our uncertainty estimates on our treatment of inspiral-only cases, in an effort to fairly isolate the contribution from numerical relativity, meaning not only the merger but also the late portion of the inspiral that numerical relativity can solve more accurately than PN can. It has been shown that the most accurate parameter estimates in PN studies come from the case of highly spinning BHBs, since the spin-induced precession breaks degeneracies [90]. Therefore, a study similar to the one performed in Chapter 4, but using highly spinning waveforms, should yield an estimate of the absolute best accuracy we can expect from LISA [1], and would also inform us as to whether the improvement seen in Chapter 4 is universal or a special case.

Another investigation would be the parameter accuracies that Advanced LIGO would be capable of if the merger were included. In this case, SNRs may be too low to apply the Fisher matrix formalism with confidence (although the approach will still be informative as a first estimate), so a better approach would be to use Markov Chain Monte Carlo, injecting the waveform and matching to a template to

---

[1] This assumes the addition of spin does not diminish the contribution of the merger. We have no reason to expect it to, other than the possibility of covariance with the 6 new parameters describing the spins, and in fact for cases like spin-orbit interaction moving ISCO in and extending the merger, there is reason to expect that the additional late inspiral-into-merger time will decrease the statistical uncertainty.

calculate the likelihood at each point in the chain. In this case, one could either restrict the study to extrinsic parameters and use the numeric waveform itself as a template (with the analytic extrinsic scalings), as was done in Chapter 4, or else one could use a full analytic approximation, such as the p4EOB, to calculate the waveform for any value of the intrinsic or extrinsic parameters, in order to maximize the likelihood over both.

There is much work to be done, but with the first GW detection looming around the corner, it is an exciting time for the budding field of gravitational wave astronomy. Every time a new window has been opened onto the universe, like the advent of radio astronomy over a half century ago, our understanding of universe has taken a giant leap forward. Measuring BHBs, in particular, represents an opportunity to study the most extreme environment in nature. However, even more exciting than the prospect of measuring signals from BHBs is finding signals we did not expect, or signals with unexpected behavior, because there may lie new physics, and an opportunity to more fully understand our universe.

# Appendix A

## Post-Newtonian Approximations for Gravitational Waveforms

In the first section of this Appendix, we provide the Taylor PN equations that are implemented throughout the text. Then, in the second section, we present the relevant details of the p4EOB model, which was originally presented in [12], and which we use for comparisons with the gravitational wave phase and frequency measured from our numerical simulations.

## A.1  Adiabatic Taylor-expanded PN

As discussed in [59], there are multiple ways in the Taylor PN formalism to calculate the orbital phase. Following the nomenclature in [59], we implement two methods in this work, the T1 and T4 approximants. Both approximants are rooted in the energy balance equation,

$$\frac{dE}{dt} = -F \, , \tag{A.1}$$

relating the energy loss of the BHB to the outbound gravitational flux. To get the orbital phase, we first divide eq. (A.1) by $dE/d\omega$, yielding an ODE of the form $\dot{\omega}(\omega)$ (assuming $F = F(\omega)$). However, the form of this expression depends on the procedure applied, and may be the most significant distinction between T1 and T4. In the T1 method, $\dot{\omega}(\omega)$ is reexpanded in terms of $\omega$. This reexpanded expression can then be solved analytically to yield an expression of the form $\omega(t)$, which is truncated at 3.5PN order, and finally is integrated, and again truncated, to yield

the T1 orbital phase through 3.5PN,

$$
\phi = \phi_0 - frac1\nu \left\{ \tau^{5/8} + \left( \frac{3715}{8064} + \frac{55}{96}\nu \right) \tau^{3/8} - \frac{3}{4}\pi\tau^{1/4} \right.
$$

$$
+ \left( \frac{9275495}{14450688} + \frac{284875}{258048}\nu + \frac{1855}{2048}\nu^2 \right) \tau^{1/8} + \left( -\frac{38645}{172032} + \frac{65}{2048}\nu \right) \pi \ln(\tau)
$$

$$
+ \left( \frac{831032450749357}{57682522275840} - \frac{53}{40}\pi^2 - \frac{107}{56}\gamma_E + \frac{107}{448}\ln\left(\frac{\tau}{256}\right) \right)
$$

$$
+ \left[ -\frac{126510089885}{4161798144} + \frac{2255}{2048}\pi^2 \right]\nu + \frac{154565}{1835008}\nu^2 - \frac{1179625}{1769472}\nu^3 \right) \tau^{-1/8}
$$

$$
\left. + \left( \frac{188516689}{173408256} + \frac{488825}{516096}\nu - \frac{141769}{516096}\nu^2 \right) \pi\tau^{-1/4} \right\}, \tag{A.2}
$$

where $\gamma_E = 0.577 \cdots$ is the Euler constant, $\phi_0$ is an arbitrary constant, and $\tau \equiv \nu(t_c - t)/5M$. $t_c$ is the so-called "coalesence time", although the approximation will become invalid before reaching $t_c$, since the phase diverges at that time.

Alternatively, one can choose not to reexpand $\dot{\omega}(\omega)$, but rather can solve the ODE numerically to yield a data set for $\omega(t)$. This can then be numerically integrated to yield a data set for the T4 orbital phase as a function of time. The expression for $\dot{\omega}(\omega)$ is given by

$$
\frac{\dot{\omega}}{\omega^2} = \frac{96}{5}\nu(M\omega)^{5/3} \left\{ 1 - \frac{743 + 924\,\nu}{336}(m\omega)^{2/3} + 4\pi(M\omega) \right.
$$

$$
+ \left( \frac{34\,103}{18\,144} + \frac{13\,661}{2\,016}\nu + \frac{59}{18}\nu^2 \right)(M\omega)^{4/3} - \frac{1}{672}(4\,159 + 15\,876\,\nu)\pi(M\omega)^{5/3}
$$

$$
+ \left[ \left( \frac{16\,447\,322\,263}{139\,708\,800} - \frac{1\,712}{105}\gamma_E + \frac{16}{3}\pi^2 \right) + \left( -\frac{56\,198\,689}{217\,728} + \frac{451}{48}\pi^2 \right)\nu \right.
$$

$$
\left. + \frac{541}{896}\nu^2 - \frac{5\,605}{2\,592}\nu^3 - \frac{856}{105}\log\left[16(M\omega)^{2/3}\right] \right](M\omega)^2
$$

$$
\left. + \left( -\frac{4\,415}{4\,032} + \frac{358\,675}{6\,048}\nu + \frac{91\,495}{1\,512}\nu^2 \right)\pi(M\omega)^{7/3} \right\}. \tag{A.3}
$$

It is noteworthy that, by numerically integrating, we have effectively kept higher-

135

order PN terms in the T4 expression that were discarded in the T1 version in order to keep it consistently at 3.5PN order. It may be that, for instance, the effective 4PN content that has been generated is, in fact, the dominant contribution to the correct, yet-to-be calculated 4PN phase, which is one possible explanation for the superior performance observed from the T4 waveform in [56, 2, 59].

Once we have the frequency from either the T1 or T4 method, we can calculate the dominant $l = m = 2$ component of the gravitational waveform up to 3PN order:

$$
\begin{aligned}
h^{22} = -8\sqrt{\frac{\pi}{5}}\frac{G\nu m}{c^2 R}e^{-2i\phi}x\Bigg\{ &1 - x\left(\frac{107}{42} - \frac{55}{42}\nu\right) + 2\pi x^{3/2} - x^2\left(\frac{2173}{1512} + \frac{1069}{216}\nu - \frac{2047}{1512}\nu^2\right) \\
&- x^{5/2}\left[\left(\frac{107}{21} - \frac{34}{21}\nu\right)\pi + 24i\nu\right] + x^3\left[\frac{27027409}{646800} - \frac{856}{105}\gamma_E + \frac{2}{3}\pi^2 - \frac{1712}{105}\ln 2\right. \\
&\left.- \frac{428}{105}\ln x - \left(\frac{278185}{33264} - \frac{41}{96}\pi^2\right)\nu - \frac{20261}{2772}\nu^2 + \frac{114635}{99792}\nu^3 + \frac{428i}{105}\pi\right]\Bigg\},
\end{aligned}
\qquad \text{(A.4)}
$$

where $x \equiv \left(\frac{Gm\omega}{c^3}\right)^{2/3}$. Other modes, some of which may be nonnegligible for unequal mass cases, are given by eqs. (80)-(116) of [63] (ignoring terms of the form $\ln\left(\frac{x}{x_0}\right)$, as we have in eq. A.4, since they are subsequently reabsorbed into the phase at 4PN and beyond, and are therefore neglected).

## A.2  p4EOB: The pseudo-4PN Effective One Body Model

Here we provide the relevant details of the p4EOB model, which was first presented in [12]. Expressions are given in polar coordinates $(r, \phi, \theta, p_r, p_\phi, p_\theta)$ throughout (although in the absence of spin, motion is constrained to the plane, i.e.

$p_\theta = 0$). The EOB effective Hamiltonian for the nonspinning case is given by

$$H_{\text{eff}}(\mathbf{r}, \mathbf{p}) = \mu \, \widehat{H}_{\text{eff}}(\mathbf{r}, \mathbf{p})$$

$$= \mu \sqrt{A(r) \left[ 1 + \mathbf{p}^2 + \left( \frac{A(r)}{D(r)} - 1 \right) (\mathbf{n} \cdot \mathbf{p})^2 + \frac{1}{r^2} \left( z_1 (\mathbf{p}^2)^2 + z_2 \, \mathbf{p}^2 (\mathbf{n} \cdot \mathbf{p})^2 + z_3 (\mathbf{n} \cdot \mathbf{p})^4 \right) \right]},$$

$$(A.5)$$

where $\mathbf{r}$ and $\mathbf{p}$ are the reduced dimensionless position and momentum variables, and

$\mathbf{n} = \mathbf{r}/r$ where we set $r = |\mathbf{r}|$. The EOB effective metric is

$$ds_{\text{eff}}^2 \equiv g_{\mu\nu}^{\text{eff}} \, dx^\mu \, dx^\nu = -A(r) \, c^2 dt^2 + \frac{D(r)}{A(r)} \, dr^2 + r^2 \left( d\theta^2 + \sin^2 \theta \, d\varphi^2 \right), \qquad (A.6)$$

the real Hamiltonian is

$$H_{\text{real}} = M \sqrt{1 + 2\eta \left( \frac{H_{\text{eff}} - \mu}{\mu} \right)} - M, \qquad (A.7)$$

through at least 3PN order, and we define $\widehat{H}_{\text{real}} = H_{\text{real}}/\mu$. The coefficients $z_1, z_2$

and $z_3$ in Eq. (A.5) are subject to the constraint

$$8z_1 + 4z_2 + 3z_3 = 6(4 - 3\eta) \, \eta \,, \qquad (A.8)$$

but are otherwise arbitrary. For the p4EOB model, we set $z_1 = z_2 = 0$, $z_3 = 2(4 - 3\eta)\eta$.

To determine the waveform phase, we will need to find values for the coefficients $A(r)$ and $D(r)$, as well as the radiation reaction force, $F(r, \phi)$. Since the primary purpose of the p4EOB is to maximize agreement with NR waveforms, we introduce a 4PN term to the radial potential $A(r)$ for that purpose:

$$A_T^{\text{p4PN}}(r) = A_T^{\text{3PN}}(r) + \frac{a_5(\eta)}{r^5}, \qquad (A.9)$$

137

where the "T" means Taylor-expanded, and $A_T^{3PN}(r)$ is given by

$$A_T^{3PN}(r) = 1 - \frac{2}{r} + \frac{2\eta}{r^3} + \left[\left(\frac{94}{3} - \frac{41}{32}\pi^2\right)\eta - z_1\right]\frac{1}{r^4}. \tag{A.10}$$

We note that $A_T^{3PN}(r)$, as given, does not have a light ring, which would be a fatal flaw for the p4EOB, as will be explained shortly. However, the Padé approximant of $A_T^{3PN}(r)$ does, and is generally the expression that is used instead for this reason. We therefore Padé approximate eq. (A.9) to get

$$A_{P_4^1}^{p4PN}(r) = \frac{\text{Num}(A_{P_4^1}^{p4PN})}{\text{Den}(A_{P_4^1}^{p4PN})} \tag{A.11}$$

$$\text{Num}(A_{P_4^1}^{p4PN}) = r^3\left[32 - 24\eta - 4a_4(\eta, 0) - a_5(\eta, \lambda)\right] + r^4[a_4(\eta, 0) - 16 + 8\eta] \tag{A.12}$$

$$\text{Den}(A_{P_4^1}^{p4PN}) = -a_4^2(\eta, 0) - 8a_5(\eta, \lambda) - 8a_4(\eta, 0)\eta + 2a_5(\eta, \lambda)\eta - 16\eta^2$$

$$+ r\left[-8a_4(\eta, 0) - 4a_5(\eta, \lambda) - 2a_4(\eta, 0)\eta - 16\eta^2\right]$$

$$+ r^2\left[-4a_4(\eta, 0) - 2a_5(\eta, \lambda) - 16\eta\right]$$

$$+ r^3\left[-2a_4(\eta, 0) - a_5(\eta, \lambda) - 8\eta\right] + r^4\left(-16 + a_4(\eta, 0) + 8\eta\right), \tag{A.13}$$

where $a_5(\eta, \lambda) = \lambda\eta$.

The 3PN Taylor-expanded form of $D(r)$ is

$$D_T^{3PN}(r) = 1 - \frac{6\eta}{r^2} + \left[7z_1 + z_2 + 2\eta\left(3\eta - 26\right)\right]\frac{1}{r^3} \tag{A.14}$$

and the 3PN Padé approximated form is

$$D_{P_3^0}^{3PN}(r) = \frac{r^3}{r^3 + 6\eta r + 2\eta(26 - 3\eta)}. \tag{A.15}$$

In the original work, the $D(r)$ coefficient used was a pseudo-4PN Padé approximant, although ultimately it was found that the pseudo-4PN approximant used is negligi-

138

bly different from eq. (A.15) in its impact on the results, so the version of p4EOB applied in this work in fact uses eq. (A.15) for $D(r)$.

The final component needed for the evolution equations is the reduced radiation reaction force, $\widehat{F}_\phi$, for which we use the Pade-approximant

$$\widehat{F}_\phi \equiv -\frac{1}{\eta\, v_\omega^3}\, \mathcal{F}[v_\omega] = -\frac{32}{5}\,\eta\, v_\omega^7\, \frac{f(v_\omega; \eta)}{1 - v_\omega/v_{\text{pole}}(\eta)}\,, \qquad (A.16)$$

where $v_\omega \equiv \widehat{\omega}^{1/3} \equiv (d\phi/d\widehat{t})^{1/3}$, with all "hats" now indicating reduced quantities, i. e. $\widehat{t} \equiv \frac{t}{M}$. For $v_{\text{pole}}$, we use

$$v_{\text{pole}} = \sqrt{\frac{1 + \eta/3}{3\,(1 - 35 * \eta/36)}} \qquad (A.17)$$

as in [91]. The calculation of $f(v_\omega; \eta)$ follows the procedure in [91], first by expressing it in terms of 7 coefficients:

$$f(v_\omega; \eta) = \left(1 - \frac{1712}{105}\, v^6\, \log\frac{v}{v_{\text{MECO}}}\right)\left(1 + \frac{c_1\, v}{1 + \frac{c_2\, v}{1+\dots}}\right)^{-1}\quad (\text{up to } c_7). \qquad (A.18)$$

where for $v_{\text{MECO}}$ we used

$$v_{\text{MECO}} = \sqrt{\frac{12\,(1 + \eta/3)}{36 - 35\eta}\left(1 - \sqrt{\frac{12 + 8\eta + 3\eta^2}{48 - 27\eta + 3\eta^2}}\right)}\,, \qquad (A.19)$$

which is found by differentiating eq. (46) of [91], which gives the 2PN Pade-approximated energy function, in order to find the maximum, i.e. $\frac{de}{dv}(v_{\text{MECO}}) = 0$. The $c_i$'s are in turn functions of another set of coefficients, $f_i$, given by eqs. (51)-(53) of [91], which in turn are related to the Pade-approximated flux terms at increasing half-PN order by $f_i = \mathcal{F}_i - \frac{\mathcal{F}_{i-1}}{v_{\text{pole}}}$. For $\mathcal{F}_i$ we use the values from eqs. (38)-(41) in [91]. With the flux calculated, we have all the components necessary to evolve the equations of motion,

given by

$$\frac{dr}{d\hat{t}} = \frac{\partial \widehat{H}}{\partial p_r}(r, p_r, p_\phi)\,, \tag{A.20}$$

$$\frac{d\phi}{d\hat{t}} \equiv \widehat{\omega} = \frac{\partial \widehat{H}}{\partial p_\phi}(r, p_r, p_\phi)\,, \tag{A.21}$$

$$\frac{dp_r}{d\hat{t}} = -\frac{\partial \widehat{H}}{\partial r}(r, p_r, p_\phi)\,, \tag{A.22}$$

$$\frac{dp_\phi}{d\hat{t}} = \widehat{F}^\phi[\widehat{\omega}(r, p_r, p_\phi)]\,, \tag{A.23}$$

which will provide us with, among other things, the waveform phase as a function of time. Finally, we can use eq. A.4 to calculate the full gravitational waveform using the EOB frequency.

The aforementioned procedure will fail to yield sensible results at some point near the merger, where the black holes plunge together at relativistic speeds. To complete the p4EOB model, a ringdown stage consisting of a sum of three quasi-normal modes (QNMs) is attached. The BHB eventually merges to form a single perturbed Kerr black hole, with the perturbations damping away exponentially in time, and QNMs will be the signature of such perturbations [92, 93]. In the close-limit approximation [94], the model for the BHB coalescence switches from the two-body description to the one-body description of a single perturbed Kerr hole close to the light-ring location. Following [52], the final black hole mass and spin are used to determine the fundamental frequency and damping time of our QNMs, and we finally match the summed QNMs to our p4EOB inspiral at the light ring. Therefore, using coefficients that fail to predict a light ring would prevent us from completing the model. In this work, no analysis incorporating the ringdown is performed, so we discuss the details here only to present a complete picture of the original p4EOB

140

model.

# Bibliography

[1] J. G. Baker *et al.*, Phys. Rev. Lett. **99**, 181101 (2007).

[2] J. G. Baker *et al.*, Phys. Rev. D **75**, 124024 (2007).

[3] S. T. McWilliams *et al.*, (2008), to be included in the Proceedings of the 12th Gravitational Wave Data Analysis Workshop.

[4] B. W. Carroll and D. A. Ostlie, *Modern Astrophysics* (Addison-Wesley, New York, 1996).

[5] R. A. Alpher, H. A. Bethe, and G. Gamow, Phys. Rev. **73**, 803 (1948).

[6] G. Gamow, Phys. Rev. **74**, 505 (1948).

[7] R. A. Hulse and J. H. Taylor, Astrophys. J. Letters **195**, L51 (1975).

[8] S. Chandrasekhar, Astrophys. J. **97**, 255 (1943).

[9] S. Chandrasekhar, Astrophys. J. **97**, 263 (1943).

[10] S. Chandrasekhar, Astrophys. J. **97**, 54 (1943).

[11] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W. H. Freeman, San Francisco, 1973).

[12] A. Buonanno *et al.*, Phys. Rev. D **76**, 104049 (2007).

[13] J. Weber, Phys. Rev. Lett. **22**, 1320 (1969).

[14] P. R. Saulson, *Fundamentals of Interferometric Gravitational Wave Detectors* (World Scientific, New Jersey, 1994).

[15] B. Brügmann, W. Tichy, and N. Jansen, Phys. Rev. Lett. **92**, 211101 (2004).

[16] F. Pretorius, Phys. Rev. Lett. **95**, 121101 (2005).

[17] M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower, Phys. Rev. Lett. **96**, 111101 (2006).

[18] J. G. Baker *et al.*, Phys. Rev. Lett. **96**, 111102 (2006).

[19] S. Brandt and B. Brügmann, Phys. Rev. Lett. **78**, 3606 (1997).

[20] P. Diener *et al.*, Phys. Rev. Lett. **96**, 121101 (2006).

[21] M. Campanelli, C. O. Lousto, and Y. Zlochower, Phys. Rev. D **73**, 061501(R) (2006).

[22] J. G. Baker *et al.*, Phys. Rev. D **73**, 104002 (2006).

[23] J. M. Centrella *et al.*, J.Phys.Conf.Ser. **78**, 012010 (2007).

[24] R. Arnowitt, S. Deser, and C. W. Misner, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (John Wiley, New York, 1962), pp. 227–265.

[25] T. W. Baumgarte and S. L. Shapiro, Phys. Rep. **376**, 41 (2003).

[26] M. Alcubierre *et al.*, Phys. Rev. D **62**, 124011 (2000).

[27] O. Sarbach, G. Calabrese, J. Pullin, and M. Tiglio, Phys. Rev. D **66**, 064002 (2002).

[28] O. A. Reula, (2004), arXiv:gr-qc/0403007v1.

[29] M. Alcubierre *et al.*, Phys. Rev. D **67**, 084023 (2003).

[30] Y. Zlochower, J. G. Baker, C. O. Lousto, and M. Campanelli, Phys. Rev. D **72**, 024021 (2005).

[31] P. Hübner, Class. Quantum Grav. **16**, 2823 (1999).

[32] M. D. Duez, S. L. Shapiro, and H.-J. Yo, Phys. Rev. D **69**, 104016 (2004).

[33] J. D. Schnittman *et al.*, (2007), arXiv:0707.0301 [gr-qc], to be published in Phys. Rev. D15.

[34] P. Szekeres, J. Math. Phys **6**, 1387 (1965).

[35] C. Beetle, M. Bruni, L. M. Burko, and A. Nerozzi, Phys. Rev. D **72**, 024013 (2005).

[36] J. Baker, M. Campanelli, and C. O. Lousto, Phys. Rev. D **65**, 044001 (2002).

[37] J. Goldberg *et al.*, J. Math. Phys. **8**, 2155 (1967).

[38] Y. Wiaux, L. Jacques, and P. Vandergheynst, J. Comput. Phys **226**, 2359 (2007).

[39] J. Winicour, Living Reviews in Relativity **1**, (1998).

[40] P. MacNeice *et al.*, Computer Physics Comm. **126**, 330 (2000).

[41] `http://www.physics.drexel.edu/~olson/paramesh-doc/`
`Users_manual/amr.html`.

[42] C. Misner, Phys. Rev. D **118**, 1110 (1960).

[43] S. G. Hahn and R. W. Lindquist, Ann. Phys. **29**, 304 (1964).

[44] J. D. Brown and L. L. Lowe, J. Comp. Phys. **209**, 582 (2005).

[45] W. Tichy and B. Brügmann, Phys. Rev. D **69**, 024006 (2004).

[46] J. R. van Meter, J. G. Baker, M. Koppitz, and D.-I. Choi, Phys. Rev. D **73**, 124011 (2006).

[47] M. Hannam *et al.*, Phys. Rev. Lett. **99**, 241102 (2007).

[48] J. D. Brown, (2007), arXiv:0705.1359v3 [gr-qc], to be published in Phys. Rev. D15.

[49] P. C. Peters, Phys. Rev. **136**, B1224 (1964).

[50] C. W. Misner, Class. Quantum Grav. **21**, S243 (2004).

[51] D. R. Fiske *et al.*, Phys. Rev. D **71**, 104036 (2005).

[52] A. Buonanno and T. Damour, Phys. Rev. D **59**, 084006 (1999).

[53] T. Damour, P. Jaranowski, and G. Schäfer, Phys. Rev. D **62**, 084011 (2000).

[54] L. Blanchet, Phys. Rev. D **65**, 124009 (2002).

[55] E. Berti, S. Iyer, and C. M. Will, Phys. Rev. D **74**, 061503(R) (2006).

[56] A. Buonanno, G. B. Cook, and F. Pretorius, Phys. Rev. D **75**, 124018 (2007).

[57] L. Blanchet, Living Reviews in Relativity **5**, 3 (2002).

[58] J. R. van Meter, (2007), private communication.

[59] M. Boylexa *et al.*, Phys. Rev. D **76**, 124038 (2007).

[60] K. G. Arun, L. Blanchet, B. R. Iyer, and M. S. S. Qusailah, Class. Quantum Grav. **21**, 3771 (2004), Erratum: **22**, 3115(E) (2005).

[61] M. Hannam *et al.*, Phys. Rev. D **77**, 044020 (2008).

[62] E. Berti *et al.*, Phys. Rev. D **76**, 064034 (2007).

[63] L. E. Kidder, (2007), arXiv:0710.0614v1 [gr-qc], to be published in Phys. Rev. D15.

[64] S. T. McWilliams and J. G. Baker, in *Sixth International LISA Symposium*, edited by S. Merkowitz and J. Livas (American Institute of Physics, New York, 2006), pp. 110–114.

[65] E. E. Flanagan and S. A. Hughes, Phys. Rev. D **57**, 4535 (1998).

[66] D. N. Spergel *et al.*, Astrophys. J. Supp. Ser. **170**, 377 (2007).

[67] A. Lazzarini and R. Weiss, Technical Report No. LIGO-E950018-02 E 25.03.96, LIGO Scientific Collaboration, (unpublished), http://www.ligo.caltech.edu/docs/E/E950018-02.pdf.

[68] D. Shoemaker, 2006, private communication.

[69] K. Belczynski *et al.*, Astrophys. J. **662**, 504 (2007).

[70] J. M. Fregeau *et al.*, Astrophys. J. **646**, L135 (2006).

[71] M. Campanelli, C. O. Lousto, and Y. Zlochower, Phys. Rev. D **74**, 041501(R) (2006).

[72] S. Larson, W. Hiscock, and R. Hellings, Phys. Rev. D **62**, 062001 (2000).

[73] S. Larson, http://www.srl.caltech.edu/ shane/sensitivity/.

[74] S. M. Merkowitz, in *Sixth International LISA Symposium*, edited by S. Merkowitz and J. Livas (American Institute of Physics, New York, 2006), pp. 133–142.

[75] D. A. Shaddock, M. Tinto, F. B. Estabrook, and J. W. Armstrong, Phys. Rev. D **68**, 061303 (2003).

[76] L. Barack and C. Cutler, Phys. Rev. D **69**, 082005 (2004).

[77] M. Milosavljevic and D. Merritt, Astrophys. J. **596**, 860 (2003).

[78] A. Sesana, F. Haardt, P. Madau, and M. Volonteri, Astrophys. J. **623**, 23 (2005).

[79] N. J. Cornish and L. J. Rubbo, Phys. Rev. D **67**, 022001 (2003).

[80] N. J. Cornish, L. J. Rubbo, and O. Poujade, The LISA Simulator, www.physics.montana.edu/LISA/.

[81] C. Cutler, Phys. Rev. D **57**, 7089 (1998).

[82] S. Detweiler and E. Szedenits, Astrophys. J. **231**, 211 (1979).

[83] C. Cutler and E. Flanagan, Phys. Rev. D **49**, 2658 (1994).

[84] M. Vallisneri, Phys. Rev. D **71**, 022001 (2005).

[85] T. A. Prince, M. Tinto, S. L. Larson, and J. W. Armstrong, Phys. Rev. D **66**, 122002 (2002).

[86] A. Krolak, M. Tinto, and M. Vallisneri, Phys. Rev. D **70**, 022003 (2004).

[87] Baker, 2007, technical note.

[88] C. Cutler and M. Vallisneri, Phys. Rev. D **76**, 104018 (2007).

[89] M. A. Scheel *et al.*, Phys. Rev. **D74**, 104006 (2006).

[90] R. N. Lang and S. A. Hughes, Phys. Rev. D **74**, 122001 (2006).

[91] A. Buonanno, Y. B. Chen, and M. Vallisneri, Phys. Rev. D **67**, 024016 (2003).

[92] M. Davis, R. Ruffini, H. Press, and R. H. Price, Phys. Rev. Lett. **27**, 1466 (1971).

[93] W. H. Press, Astrophys. J. Letters **170**, L105 (1971).

[94] R. H. Price and J. Pullin, Phys. Rev. Lett. **72**, 3297 (1994), R. J. Gleiser et. al., Class. Quant. Grav. **13**, L117 (1996), Phys. Rev. Lett. **77**, 4483 (1996), P. Anninos et. al., Phys. Rev. Lett. **71**, 2851 (1993), J. G. Baker et. al., Phys.

Rev. D **55**, 829 (1997), Z. Andrade and R. H. Price, Phys. Rev. D **56**, 6336 (1997).