ABSTRACT

Title of dissertation: NOVEL TECHNIQUES FOR
SIMULATION AND ANALYSIS
OF BLACK HOLE MERGERS

William Darian Boggs, Doctor of Philosophy,
2011

Dissertation directed by: Professor Manuel Tiglio
Department of Physics

This dissertation consists of three research topics from numerical relativity: waveforms from inspiral mergers of black hole binaries, recoils from head-on mergers of black holes, and a new computational technique for error-reduction. The first two topics present research from journal articles that I coauthored with my colleagues in the NASA Goddard Numerical Relativity research group.

Chapter 2 discusses a heuristic model of black hole binary mergers and the waveforms produced by them, based on simulations of nonspinning black holes. The gravitational radiation is interpreted as the result of an *implicit rotating source* that generates the radiation modes as the source multipoles rotate coherently. This interpretation of the waveform phase evolution provides a unified physical picture of the inspiral, plunge, and ringdown of the binaries, and it is the basis of an analytic model of the late-time frequency evolution.

Chapter 3 presents a study of kicks in head-on black hole mergers, emphasizing the distinct contributions of spin and mass ratio, as well as their combined effects,

to these radiation-induced recoils. The simpler dynamics of head-on mergers allow a more clear separation of the two types of kick and a validation of post-Newtonian predictions for the spin scaling of kicks.

Finally, Chapter 4 presents a technique I developed to improve the accuracy of the field evolution in numerical relativity simulations. This "moving patches" technique uses local coordinate frames to minimize black hole motion and reduce error due to advection terms. In tests of the technique, I demonstrate reduction in constraint violations and in errors in the orbital frequency derived from the black holes' motions. I also demonstrate an accuracy gain in a new diagnostic quantity based on orbital angular momentum. I developed this diagnostic for evaluating the moving patches technique, but it has broader applicability. Though the moving patches technique has significant performance costs, these limitations are specific to the current implementation, and it promises greater efficiency and accuracy in the future.

Novel Techniques for Simulation and Analysis of Black Hole Mergers

by

William Darian Boggs

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
2011

Advisory Committee:
Professor Manuel Tiglio, Chair/Advisor
Dr. John Baker
Professor Dieter Brill
Professor Peter Shawhan
Professor M. Coleman Miller

Preface

Unless otherwise noted, f represents a scalar function of one or more real variables. In the context of general relativity, $f_{,\mu}$ indicates the partial derivative of f with respect to the coordinate μ. In discussions of numerical analysis, $f^{(n)}$ indicates the nth derivative of f. For indices from the beginning of the Latin alphabet, tensor components represent the tensor independent of basis and convey the number of arguments and covariance or contravariance of each argument (*abstract index notation*). Indices from the Greek alphabet indicate spacetime tensor components in a particular basis, with the indices running from 0 (*time*) to 3. Indices from the portion of the Latin alphabet starting with i indicate purely spatial tensor components with the indices running from 1 to 3. Roman font indicates four-dimensional tensors (g_{ab}), while italic font indicates the three-dimensional counterparts (g_{ab}) of those tensors. A metric without indices represents the determinant of the metric ($g = \det g_{ab}$), while any other matrix without indices represents the trace of that matrix ($U = \operatorname{tr} U_{ab}$). Following the Einstein summation convention, the same letter appearing as a superscript index and a subscript index in an expression indicates summation over the range of the index ($T^i S_i = \sum_{i=1}^{3} T^i S_i$). Finally, following a common convention in general relativity, the speed of light in vacuum and the universal gravitational constant are both set to 1, $c \equiv G \equiv 1$. As a result, lengths, times, and masses have the same dimension, which is mass, another common convention in general relativity.

Dedication

I dedicate this dissertation to Hailey, who kept a child's curiosity alive in me, Rebecca, who shared her strength with me, David and Nickie, who believed in me even when I did not, Tina, who kept me questing despite the temptation to settle, and Barry, who taught me to love the "life of the mind".

Acknowledgments

Without my wife Rebecca and my daughter Hailey, this dissertation would not exist. I can never thank them enough for their love, support, and patience - thank you, my loves. My other friends and family have given me unlimited love, support, and patience, as well. I am grateful to all of you: Tina, Barry, Wayne, David, Nickie, Barb, Peggy Ann, and Ray. Most of my graduate career was spent at NASA Goddard Space Flight Center, and I am thankful for the opportunity given me by the Numerical Relativity group there. I am the physicist I am today because of you: Breno, Beany, Jim, Joan, John, and Sean. At the University of Maryland, I am grateful to Alessandra Buonanno and Manuel Tiglio for serving as my official advisors at different times. Your patience with a remote student is much appreciated. Also at the University of Maryland, I am grateful to Ted Jacobson and Michael Coplan for your assistance at a moment when I thought my graduate studies would end. In my life I have been blessed by numerous wonderful teachers, to all of which I owe much gratitude. In particular thank you to Beckie Gibson, Gene Cunningham, Alan Kaplan, Frieda Coleman, Kaye Boone, Gary Christy, and Ellen Rust. Finally, I am thankful to my "family" at Village Baptist Church, including Dave Thompson, a NASA astrophysicist who introduced me to the Numerical Relativity group at Goddard and offered me support on the several occasions when matters personal and academic overwhelmed me.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

$\tilde{A}_{ij}$ traceless conformal extrinsic curvature

α lapse

β^i shift

$\widetilde{\Gamma}^i$ conformal connection function

g_{ij} physical spatial metric

g determinant of g_{ij}

$\tilde{g}_{ij}$ conformal spatial metric

$\tilde{g}$ determinant of $\tilde{g}_{ij}$

K_{ij} extrinsic curvature

K trace of K_{ij}

M total mass

M_{AH} total apparent horizon mass

ϕ conformal factor

x_p^i puncture position

∂_0 $\partial_t - \mathcal{L}_\beta$

ADM   Arnowitt, Deser, Misner

AMR   adaptive mesh refinement

BSSNOK   Baumgarte, Shapiro, Shibata, Nakamura, Oohara, Kojima

FMR   fixed mesh refinement

GR   General Relativity

NR   Numerical Relativity

ODE   ordinary differential equation

PDE   partial differential equation

PN   Post-Newtonian

RK4   Fourth-order Runge-Kutta

Chapter 1

Numerical Relativity

On November 2, 2005, "Numerical Relativity 2005: Compact Binaries", a workshop on numerical relativity cosponsored by NASA Goddard Space Flight Center and the Center for Gravitational Wave Physics at Penn State University, began with coffee and tea, and colleagues catching up on matters professional and personal, like most such meetings. Soon it was clear the meeting was a momentous occasion in a momentous year for the numerical relativity community, as the numerical relativity research groups from University of Texas at Brownsville and NASA Goddard in succession presented the results of successful three-dimensional simulations of the inspiral, plunge, and ringdown of two black holes. Frans Pretorius had announced the successful simulation of a black hole merger on July 4, 2005. As the end of 2005 approached, the black hole binary challenge of general relativity, which had eluded relativists for decades, was conquered.

I sat in the conference room, a second-year graduate student not fully appreciating the years of effort leading up to this point and the importance of the moment. Also, I could not know the rapid progress in the field that lay ahead after years of struggle, progress to which I would contribute as a member of the Goddard numerical relativity group. In this dissertation, I present three contributions I made. Chapter 1 provides a brief introduction to numerical relativity with background

material for the following chapters. Chapter 2 discusses a heuristic model of black

mergers and the waveforms produced by them, based on simulations of nonspinning

black holes, with an emphasis on the unified interpretation of phase to which I con-

tributed. Chapter 3 presents a study of recoil kicks in head-on black hole mergers I

carried out with Bernard Kelly and the rest of the Goddard group, emphasizing the

distinct contributions of spin and mass ratio. In Chapter 4, I present a technique I

developed to improve accuracy in calculations of the evolved fields, based on a sim-

plification of the black hole dynamics and an auxiliary coordinate frame. Beyond the

common basis in numerical relativity the three investigations share, they also share

a common philosophy: simplicity. The heuristic model of Chapter 2 uses a simpli-

fied representation of a black hole binary, the kick study in Chapter 3 investigates

the simpler head-on merger that approximates the plunge following the inspiral of

a black hole binary, and the technique in Chapter 4 is motivated by a simplification

of the merger dynamics. In all three cases, simplicity provides a starting point for

investigating the complex.

1.1   Numerical Relativity from General Relativity

General relativity describes the interaction of space and time with matter and

energy. Often it is described as a theory of gravity, to put it in a historical context

as both the successor to Newton's theory of gravity and the extension of Einstein's

theory of special relativity. To eliminate conflict between the instantaneous action of

Newton's theory of gravity and the finite transmission time of all effects postulated

by special relativity, Einstein proposed his theory of general relativity, which replaces the force of gravity acting on matter by the distortion of a unified spacetime by matter and energy, to explain the motion of matter and energy through spacetime. Gravity is no longer a force separate from the other fundamental forces, but rather a manifestation of the inherent nature of the spacetime in which the other forces act.

Einstein's equations,

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

relate the curvature (distortion) G_{ab} of four-dimensional spacetime to the matter and energy T_{ab} present in it. Curvature is a departure from Euclidean geometry, and it is meant to suggest curved surfaces, such as spheres, in three-dimensional Euclidean space, surfaces on which the geometry is non-Euclidean. The curvature in Einstein's equations is defined within the four-dimensional spacetime. The definition does not require a higher dimensional space, as an intuitive notion of curved two-dimensional surfaces in three-dimensional Euclidean space does. The equations are second-order partial differential equations for the four-metric g_{ab}, which defines distances in spacetime. The equations are not written with respect to a particular reference frame. This property of general covariance includes the unification of space and time, but the implications are broader. There is no preferred reference frame, and two frames are equally valid despite the acceleration of either with respect to the other.

Numerical relativity simulates spacetimes based on Einstein's equations. While

the exact solutions of Einstein's equations assume symmetries, such as time-invariance, spherical or axial symmetry, or homogeneous energy distributions, and therefore represent very specialized physical scenarios, numerical relativity is able to investigate more general solutions, including those representing more astrophysically relevant scenarios, such as systems of black holes and neutron stars (generically known as *compact bodies*).

Mergers of compact bodies are a primary focus of numerical relativity because of their importance for physics, astrophysics, and astronomy. These mergers should produce gravitational waves, propagating distortions of spacetime, which are caused by the non-axially symmetric, time-varying mass distributions of the systems. The direct detection of such waves is an important untested verification of general relativity, and the goal of a worldwide search at detectors such as LIGO, Virgo, and GEO 600, as well as the space-based LISA mission in development by NASA and the European Space Agency (ESA).

Beyond the importance of the detection itself, the gravitational waves carry information about their sources, and therefore they offer a new way to investigate astrophysical processes and astronomical sources. The waves also carry energy and momentum away from their sources, which produces a decaying orbit (*inspiral*) and eventual plunge for a compact binary. The resulting black hole continues to radiate gravitational waves as it "sheds" energy in a process called *ringdown*, until it is axially symmetric. If the linear momentum carried away by the gravitational waves is asymmetric, the remnant black hole has a net linear momentum, which results in a significant velocity in some cases. Numerical relativity provides predictions of

and insights into the waves and their sources. Though complementary approaches like post-Newtonian and perturbation analyses also study these phenomena, only numerical relativity can analyze the transitional phase from inspiral to ringdown. Thus numerical relativity makes crucial contributions to physics, astrophysics, and astronomy.

1.2 Simulation

Numerical relativity treats the solution of Einstein's equations as a constrained initial value problem and uses the tools of numerical analysis to obtain the solution. In this case, the initial data are not freely specifiable, but rather they are obtained by solving constraint equations. These initial data, which describe spacetime at an initial time, are evolved forward in time like other physical systems in simulations. The evolution requires a choice of a time coordinate since general relativity describes spacetime without reference to particular time and space coordinates. This choice of time is referred to as a slicing or foliation of spacetime, since it slices four-dimensional spacetime into three-dimensional "slices" and labels the slices sequentially by a time coordinate t.

Simulating spacetime in numerical relativity divides into four parts conceptually, though the steps are interrelated. We must choose a formulation of Einstein's equations, specify a coordinate system, solve the constraint equations for the initial state of the system (the *initial data solve*), and evolve the system along the time dimension, starting from the initial data of the system. In addition, we must select

appropriate methods from numerical analysis for the initial data solve and evolution, as well as making choices specific to those methods such as the computational grid or the set of basis functions.

Before proceeding, I mention some issues related to the simulation on the grid. The first is the dimensions of the quantities considered. As mentioned in the Preface, from the start I have adopted the convenient practice that is common in discussions of general relativity of setting c and G to 1, which immediately causes them to disappear from all expressions and equations. Also it equates the dimensions of space, time, and mass. Again following a common convention for general relativity, I express all three in units of mass.

In numerical relativity simulations following these conventions, all quantities are implicitly expressed relative to a computational mass unit, which is often written M. Therefore, a length of 1.5 is $1.5M$. Similarly, a time of 10 is $10M$, and a mass of 3 is $3M$. Quantities derived from the fundamental quantities of mass, length, and time, such as energy, momentum, and frequency, have units of M^ξ where the exponent ξ is determined by the derived quantity's relationship to the fundamental units. The computational M is a "bookkeeping" unit, since often it is more natural to express all quantities in terms of a unit like the total mass of the system. In that case, after computing that natural unit in terms of the computational M, we rescale all quantities in terms of the fundamental unit, according to the power of M in which each quantity is expresssed. Then M refers to the fundamental unit. which is usually a mass.

Finally, simulations with different resolutions are compared according to the

finest resolution present in each simulation. Though this does not fully describe the resolution of the grid, the finest resolution is a fundamental characteristic of a simulation since it establishes a ceiling on the simulation accuracy. Also, in comparisons of simulations, the grids often have identical grid structures, up to an overall scaling factor based on the ratio of the finest resolutions. In that case, comparing the finest resolutions is the same as comparing the resolutions of the whole grid.

1.2.1 Formulations

To begin, we must choose the formulation of Einstein's equations. Though in all cases spacetime is sliced and then evolved forward from slice to slice along some time direction, the different formulations of Einstein's equations differ in the slices and temporal parameters they use, as well as the quantities that are directly simulated. (Derived quantities such as gravitational radiation are more similar since they are driven by physical questions.) These differences are important because not all formulations are "well-behaved". Some formulations admit unstable solutions that allow errors to grow rapidly.

The key requirements for well-behaved formulations are *well-posedness* and *hyperbolicity*. In a well-posed system, the solutions are bounded by an exponential function that does not depend on the initial data. That exponential function limits the solutions' growth for all initial data. This guarantees that small differences in initial data result in small changes in the solutions, establishing a bound on the error. Hyperbolicity describes the similarity between a system of evolution equations and

the simple wave equation, with *strongly hyperbolic* and *weakly hyperbolic* referring to systems that are respectively more or less similar to the wave equation.

A precise definition of hyperbolicity and the types of hyperbolicity requires rewriting the system in matrix form and analyzing the eigenvalues and eigenvectors of the matrix that specifies the spatial derivatives. If the matrix has real eigenvalues and eigenvectors that span the space, the system is strongly hyperbolic. A weakly hyperbolic system has real eigenvalues, but it does not have a complete set of eigenvectors. Therefore the well-posedness of the system cannot be established, and it may have solutions with arbitrarily large propagation speeds that lead to instability.

The most common formulations of Einstein's equations in the numerical relativity community are BSSNOK (Baumgarte, Shapiro, Shibata, Nakamura, Oohara, and Kojima) [84, 19, 97] and generalized harmonic coordinates [69]. Both formulations foliate spacetime into spacelike slices, level sets of a time coordinate,[1] but the generalized harmonic formulation maintains the covariance of Einstein's equations and transforms them into wave equations for the four-metric coefficients g_{ab}. This distinguishes it from BSSNOK and other common formulations of general relativity.

1.2.1.1   3+1 Decomposition, ADM, and BSSNOK

BSSNOK is a "3+1" formalism descended from the Arnowitt-Deser-Misner (ADM) formulation [8, 113]. A "3+1" decomposition of Einstein's equations separates time from space, in contrast to generalized harmonic coordinates which remain

---

[1]More precisely, the gradient of this coordinate has an extremum for a future-pointing (and therefore timelike) vector.

covariant. In both types of formulation, the data are evolved from each spatial slice to the next, starting with the slice that has the initial data, but in the generalized harmonic formulation the time coordinate is treated like the spatial coordinates in the equations. A "3+1" formulation decomposes spacetime into a series of three-dimensional slices labeled by the real parameter t, which serves as the time coordinate. (See Fig. 1.1.) The slicing and the spacetime metric g_{ab} induce an intrinsic geometry on each slice described by the spatial metric, g_{ab}. The covariant derivative D_a, the connection $\Gamma^a{}_{bc}$, the Riemann tensor $R^d{}_{abc}$, the Ricci tensor R_{ab}, and the Ricci scalar R are defined the same on the slices as they are in spacetime, except that g_{ab} is substituted for g_{ab}.[2] Because general relativity is covariant, we are free to choose how to slice spacetime and how to assign spatial coordinates to points on the slices.

The lapse α determines the slicing by specifying how quickly time evolves in the direction orthogonal to the slice at each point. The shift β^i specifies how the coordinates on the slices change in time. Together, the lapse and shift constitute the gauge, and our freedom to choose them is gauge freedom. Fig. 1.2 shows the lapse and shift at a point on a slice with coordinates x_0. n^μ is a unit vector normal to the slice at x_0. These coordinates refer to a different location on the next slice because of the shift. The combination $t^\mu = \alpha n^\mu + \beta^i$ is the time vector field for the evolution, along which the fields are evolved. In differential geometry, the Lie derivative of a tensor along a vector field v^a ($\mathcal{L}_\mathbf{v}$) is the change in the tensor produced by an infinitesimal displacement along v^a. Therefore the time derivative for the evolution

---

[2]Sec. A.2 of the appendix provides explicit expressions for these quantities.

9
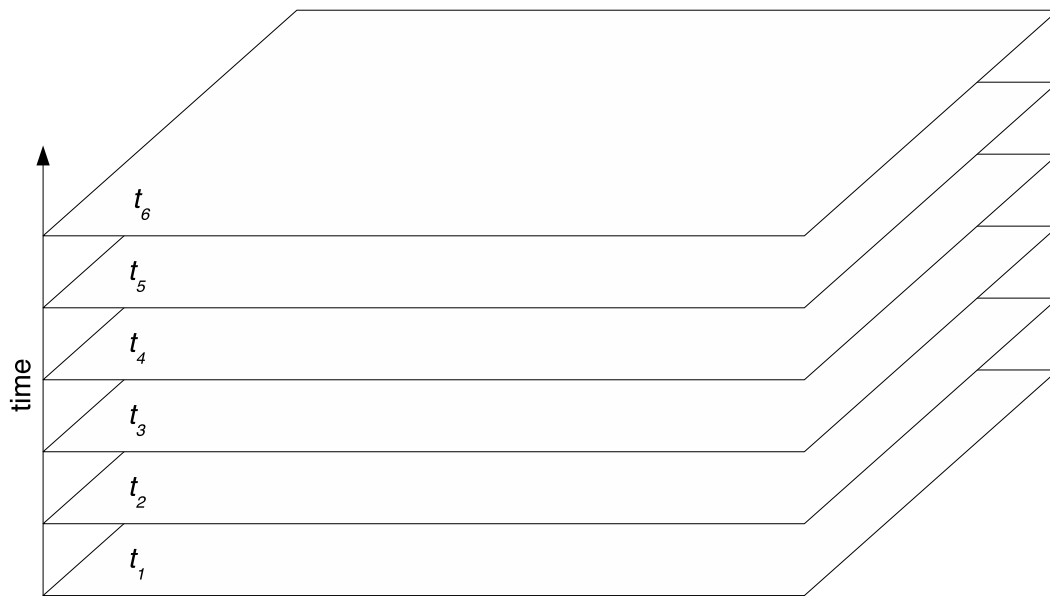
Figure 1.1: 3+1 decomposition of spacetime. Time (t) advances vertically. One spatial dimension has been suppressed.
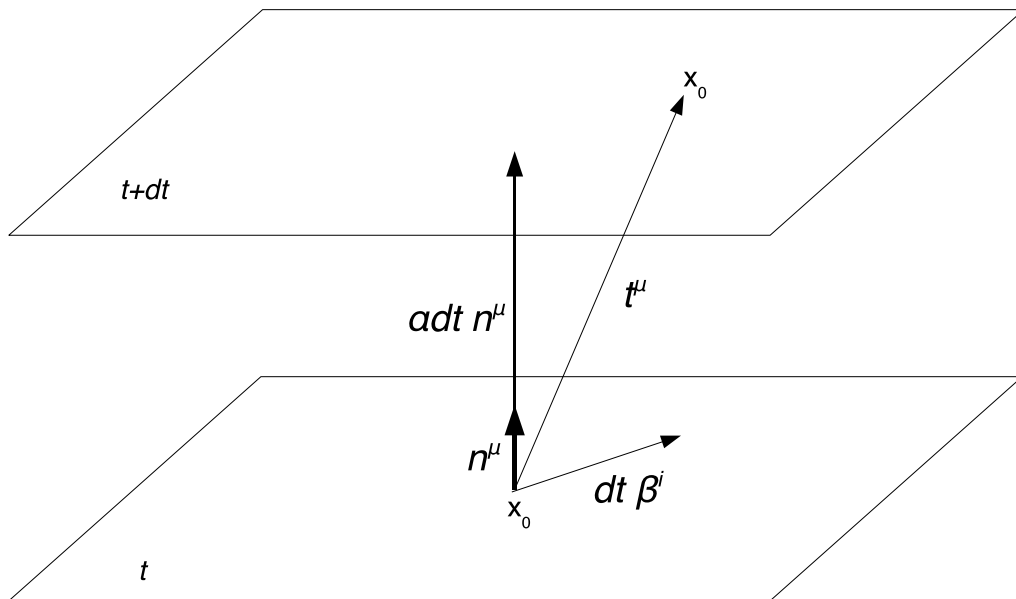
Figure 1.2: Gauge for a 3+1 decomposition of spacetime. n^μ is the unit normal vector of the slice. t^μ is the time evolution vector, and t is the time coordinate.

is the Lie derivative along t^μ $\mathcal{L}_{\mathbf{t}}$.

A "3+1" decomposition separates Einstein's equations into constraint equations which constrain the data on each slice and evolution equations that describe the evolution of the data from one slice to the next. The decomposition also rewrites the spacetime metric g_{ab} in terms of the spatial metric g_{ab} and the extrinsic curvature K_{ab}, which specifies how the spatial slice is embedded in the four-manifold. Also, it is proportional to the Lie derivative of the spatial metric along the normal to the slice, $K_{ab} = -\frac{1}{2}\mathcal{L}_{\mathbf{n}} g_{ab}$. Einstein's second-order-in-time equations for the spacetime

metric g_{ab} become a system of first-order-in-time partial differential equations for the spatial metric g_{ab} and the extrinsic curvature K_{ab}. With the four-dimensional manifold decomposed into time slices and Einstein's equations transformed to obtain evolution equations and constraint equations, the simulation of the spacetime is an initial value problem starting from the data on the initial slice (initial data).

The data must satisfy the constraint equations on all the slices. In the continuum limit, the evolution equations preserve the constraints, so that solutions of the constraint equations continue to satisfy the constraint equations as those solutions evolve and it is not necessary to solve the constraint equations on each slice. For simulations at finite resolutions the evolution equations do not preserve the constraints exactly. *Constraint violations*, which are discussed later in this chapter, measure the failure of the solutions to satisfy the constraint equations and reflect a departure from physics as described by Einstein's equations. Usually it is sufficient to solve the constraints initially and then evolve the data forward to subsequent slices using the evolution equations, while monitoring the constraint violations.

The ADM equations are a "3+1" Hamiltonian formulation of general relativity originally proposed by Arnowitt, Deser, and Misner [8].[3] The original ADM equations describe the spatial metric and its canonical conjugate momentum π^{ij}. York later rewrote the equations in the form used today in the numerical relativity community [113], which I will refer to as the "ADM" equations. In this formulation

---

[3]For an excellent derivation of the ADM equations, see [100]

the constraint equations are

$$R + K^2 - K_{ij}K^{ij} = 16\pi\rho \,, \tag{1.2}$$

and

$$D_j(K^{ij} - g^{ij}K) = 8\pi j^i \,. \tag{1.3}$$

$\rho = T_{ab}n^a n^b$ is the energy density at each point measured by an observer at rest on the slice at that point, and $j^a = -g^{ab}n^c T_{bc}$ is the momentum density as measured by the same observer. So Eqs. (1.2) and (1.3) are called the Hamiltonian and momentum constraint equations, respectively. In cases where there is no matter or energy, such as black hole binaries, the right-hand side of the constraint equations are zero since $T_{ab} = 0$ in vacuum. Since I am considering only black hole simulations, I will use the vacuum form of all equations.

The ADM evolution equations, which describe the change in time of the metric and extrinsic curvature, are

$$\partial_0 g_{ij} = -2\alpha K_{ij} \,, \tag{1.4}$$

and

$$\partial_0 K_{ij} = -D_i D_j \alpha + \alpha(R_{ij} + K_{ij}K - 2K_{ik}K^k{}_j) \,, \tag{1.5}$$

where $\partial_0 \equiv \partial_t - \mathcal{L}_\beta$.[4] First we solve the constraint equations for g_{ab} and K_{ab} on the initial spatial slice, the initial data solve. Then we use the evolution equations

---

[4]Eqs. (1.4) and (1.5) are written using the ∂_0 time derivative for conciseness. Because $\mathcal{L}_\mathbf{t} = \mathcal{L}_{\alpha\mathbf{n}+\beta} = \mathcal{L}_{\alpha\mathbf{n}} + \mathcal{L}_\beta$, ∂_0 is the time derivative orthogonal to the spatial slice $\mathcal{L}_{\alpha\mathbf{n}}$. See Sec. A.1 of the appendix for the $\mathcal{L}_\beta$ terms and elaboration on the additive property of a Lie derivative across its vector field.

to evolve the data onto future slices. This dissertation focuses on the evolution.

The ADM formulation of general relativity as an initial value problem provides an intuitive picture of the spatial metric and its evolution. Unfortunately, it is only weakly hyperbolic, and it is not stable, so it can only be evolved for short times. The BSSNOK formulation [19, 97, 84] is a more stable formulation of general relativity for time evolution. It was originally developed by Nakamura, Oohara, and Kojima [84], and later refined by Shibata and Nakamura [97] and Baumgarte and Shapiro [19]. The improved stability of BSSNOK was demonstrated empirically by Baumgarte and Shapiro [19] and investigated more formally by Alcubierre *et al.*[3]. Later, Sarbach *et al.*demonstrated that the system is strongly hyperbolic [92], in support of the original empirical stability results.

Rather than evolve g_{ij} and K_{ij}, BSSNOK evolves $\tilde{g}_{ij}$, ϕ, $\tilde{A}_{ij}$, and K, which are related to g_{ij} and K_{ij} by

$$g_{ij} = e^{4\phi}\tilde{g}_{ij}\,, \tag{1.6}$$

$$\phi = \frac{1}{12}\log g\,, \tag{1.7}$$

$$K_{ij} = e^{4\phi}\left(\tilde{A}_{ij} + \frac{1}{3}\tilde{g}_{ij}K\right)\,, \tag{1.8}$$

and

$$K = g^{ij}K_{ij}\,. \tag{1.9}$$

From these defining equations, it follows that $\tilde{g} = 1$ and $\tilde{A} = 0$. BSSNOK evolves

an auxiliary variable $\widetilde{\Gamma}^i$,

$$\widetilde{\Gamma}^i \equiv \tilde{g}^{jk}\widetilde{\Gamma}^i_{jk} \qquad (1.10)$$

$$= -\tilde{g}^{ij}_{,j}, \qquad (1.11)$$

where $\widetilde{\Gamma}^i_{jk}$ is the conformal connection associated with $\tilde{g}_{ij}$. The identity on the second line of Eq. (1.10) is proven in Problem 7.7f of [68].

In the definitions of the BSSNOK variables, we see three differences between the BSSNOK and ADM formulations. Eqs. (1.6) and (1.8) are conformal transformations of the physical spatial metric and extrinsic curvature. The initial factor in both equations is a conformal factor, which can be rewritten using Eq. (1.7) as $e^{4\phi} = e^{\frac{1}{12}\log g} = \Psi^4$ where $\Psi = \frac{1}{12}g$. Often ϕ is referred to as the conformal factor. g is the volume element on the spatial slices, since it is the determinant of g_{ij}. So we can see that the conformal transformation separates changes to the magnitude of the physical spatial volume element from non scalar distortions of the metric and extrinsic curvature. The "$\sim$" over a variable indicates that the variable has been conformally scaled. From Eqs. (1.8) and (1.9) we see that $\tilde{A}_{ij}$ is also traceless, so it is called the conformal traceless extrinsic curvature, and K is evolved separately. Finally, all second derivatives of the conformal metric $\tilde{g}_{ij}$ appearing in the evolution equations are reduced to a simple Laplace operator in the Ricci tensor, since the other second derivatives are rewritten as first derivatives of the auxiliary conformal connection variable $\widetilde{\Gamma}^i$.

From Eqs. (1.6)-(1.9) and Eq. (1.10) and the ADM evolution equations Eq. (1.4) and (1.5)

the evolution equations for the BSSNOK variables are

$$\partial_0 \tilde{g}_{ij} = -2\alpha \tilde{A}_{ij} \,, \tag{1.12}$$

$$\partial_0 \phi = -\frac{1}{6}\alpha K \,, \tag{1.13}$$

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

$$\alpha \left(K \tilde{A}_{ij} - 2\tilde{A}_{ik}\tilde{A}^k_j \right) \,, \tag{1.14}$$

$$\partial_0 K = -D^i D_i \alpha + \alpha \left(\tilde{A}_{ij}\tilde{A}^{ij} + \frac{1}{3}K^2 \right) \,, \tag{1.15}$$

and

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

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

$$2\alpha \left(\tilde{\Gamma}^i{}_{jk}\tilde{A}^{jk} + 6\tilde{A}^{ij}\partial_j\phi - \frac{2}{3}\tilde{g}^{ij}\partial_j K \right) \,, \tag{1.16}$$

where again $\partial_0 \equiv \partial_t - \mathcal{L}_\beta$.[5] **TF** indicates the trace-free portion of the quantity in square brackets.

In addition to the differences in the BSSNOK variable definitions above, there are two important elements of the BSSNOK formulation are present in the evolution equations, Eqs. (1.12)-(1.16). In the K evolution equation Eq. (1.15), the Hamiltonian constraint Eq. (1.2) has been used to substitute for the spatial Ricci scalar R. Similarly, the momentum constraints Eq. (1.3) have been used to replace the divergence of $\tilde{A}^{ij}$ in the $\widetilde{\Gamma}^i$ evolution equation Eq. (1.16). This substitution is particularly important since without it, BSSNOK is unstable [19, 3].

---

[5]The BSSNOK variables except $\widetilde{\Gamma}^i$ are tensor densities. See Appendix B for an explanation of tensor densities and for the $\mathcal{L}_\beta$ terms. All terms of $\partial_t \widetilde{\Gamma}^i$ are given above explicitly, since $\widetilde{\Gamma}^i$ is neither a tensor nor a tensor density.

1.2.1.2   Harmonic

Though all the simulations presented in this dissertation use the BSSNOK formulations, I will briefly describe the generalized harmonic coordinates formulation because of its contribution to numerical relativity. Choquet-Bruhat used the wavelike nature of the formulation's equations to prove uniqueness and existence theorems for Einstein's equations [44]. Pretorius used generalized harmonic coordinates for the first successful simulation of a black hole binary inspiral and merger [90]. The Caltech-Cornell collaboration uses generalized harmonic coordinates with spectral methods to produce high-accuracy simulations of black hole binary coalescence lasting many orbits [93].

The generalized harmonic coordinates formulation specifies the spacetime coordinates x^μ according to the wave equation,

$$\Box_g x^\mu = H^\mu, \tag{1.17}$$

where $\Box_g \equiv g^{ab}\nabla_a\nabla_b$ is the wave operator for curved spacetime. Through the source functions H^μ we determine the coordinates. Using the coordinates specified by Eq. (1.17), Einstein's equations are rewritten as modified wave equations for the components of the metric. The system of equations is strongly hyperbolic, and like most strongly hyperbolic systems, it is well-posed. *Generalized* refers to these source functions. The original harmonic formulation had no source functions ($H^\mu = 0$). The freely specifiable source functions give greater control of the gauge.

1.2.2   Gauge

After choosing the formulation, we must choose the gauge because of the covariance of general relativity. This choice of time and space coordinates affects the stability, accuracy, duration, and values in a simulation. It includes the initial coordinates and the evolution equations for the coordinates. As discussed above, the choice of source functions H^μ constitutes the gauge choice in the generalized harmonic coordinates formulation. For BSSNOK, the evolution equations for the gauge must be specified. The gauge should:

- Avoid coordinate singularities.

- Avoid physical singularities.

- Evolve stably.

- Approach flat space in the weak field far from the black holes.

1.2.2.1   Lapse

To avoid coordinate singularities we need the coordinates to be well-behaved. $\tilde{g}$ must be positive definite and finite to prevent the spatial metric and its inverse from diverging. Also $\tilde{A}_{ij}$ must be nonsingular, and ϕ must be finite and non-negative. To avoid physical singularities, the lapse should approach zero in the strong field region where the physical singularities lie. Moving away from the strong field, the spacetime should approach flat space, and the lapse should approach 1, so that t becomes the proper time of the distant laboratory frame where physical quantities

should be measured. Therefore, the lapse will be a function of the spatial coordinates on each slice, as well as a function of time.

The obvious first choice for a lapse condition is $\alpha \equiv 1$ through space for all times. This choice of lapse is called *geodesic slicing* since with this choice of α, the coordinate time is the same as the proper time of observers on timelike geodesics, in free fall. Physical intuition shows that this a poor choice for black hole spacetimes, since all the points on our grid will fall onto the black holes in a finite time. Not only will the simulation quickly become trivial as the physical extent shrinks, but also the simulation will crash as the variables become singular. For example, an observer at the event horizon of a Schwarzschild spacetime will reach the physical singularity after evolving for a time equal to πM. Geodesic slicing also causes coordinate singularities as the grid points coincide on the slices (*focusing*).

In the case of geodesic slicing the volume elements at the grid points go to zero. *Maximal slicing* keeps these volume elements constant. This requires that

$$\nabla_\mu n^\mu = 0 \,, \tag{1.18}$$

and

$$K = 0 \,, \tag{1.19}$$

where the second line comes from the definition of K_{ij}. So maximal slicing requires that $K = \partial_t K = 0$, which in turn requires

$$D^2 \alpha = K_{ij} K^{ij} \alpha \,, \tag{1.20}$$

from the K evolution equation Eq. (1.15) and the Hamiltonian constraint equation Eq. (1.2). Though maximal slicing maintains the volume elements and avoids the

19

physical singularities, Eq. (1.20) is an elliptic equation that must be solved at each iteration, which is computationally expensive.

Today most groups evolving BSSNOK use some form of *1+log* slicing, so named because the version found through empirical studies of lapse conditions took the form $\alpha = 1 + \log g$ [6]. 1+log slicing belongs to the *Bona-Massó* family of hyperbolic slicing conditions

$$\partial_t \alpha = -\alpha^2 f(\alpha) K \,, \tag{1.21}$$

with different choices of $f(\alpha)$ producing different slicing conditions [22]. Choosing $f = 2/\alpha$ produces 1+log slicing. Because Eq. (1.21) is a hyperbolic evolution equation, it is evolved like the physical BSSNOK variables, a much simpler procedure than solving the elliptic equation Eq. (1.20) as required for maximal slicing, and yet 1+log slicing has favorable singularity-avoiding properties like maximal slicing. The simulations in this dissertation use a modified 1+log slicing condition

$$\partial_t \alpha = -2\alpha K + \beta^j \partial_j \alpha \,. \tag{1.22}$$

The second term is an advection term that is added to prevent zero-speed modes accumulating around the black holes. Zero-speed modes are characteristic modes of the evolution system that do not propagate. They remain on the grid and contribute to constraint violations.

1.2.2.2 Shift

For the shift, which specifies how the spatial coordinates change between slices, the two main objectives are preventing grid points from falling into the black holes

and reducing distortion of the spatial metric. Both can be achieved by a *Gamma freezing* shift condition $\partial_t \widetilde{\Gamma}^i = 0$. This condition minimizes the change in the shape of the volume elements.[6] The Gamma freezing condition results in a system of three coupled elliptic equations.

To use the Gamma freezing condition, we would need to solve the system at each iteration, which would be computationally expensive. As an alternative, Alcubierre *et al.*suggested the *Gamma driver* shift condition

$$\partial_t^2 \beta^i = \alpha^2 \xi \partial_t \widetilde{\Gamma}^i \,, \tag{1.23}$$

a hyperbolic driver condition that mimics the Gamma freezing condition through a rapid evolution to a quiescent state [5]. $\xi = \xi(x^i, \alpha) > 0$ controls the evolution. Since the initial proposal of the Gamma freezing condition, a damping term $(\propto \partial_t \beta^i)$ has been added to the evolution equation to prevent strong oscillations in the shift. Also an advection term has been added that causes any zero-speed constraint violation modes occurring at the black holes' initial locations to propagate to the outer boundary of the simulation. The simulations in this dissertation use a "shifting" shift condition that includes both additional terms,

$$\partial_t \beta^i = \frac{3}{4} \widetilde{\Gamma}^i + \beta^j \partial_j \beta^i - \eta \beta^i \,, \tag{1.24}$$

where $\eta > 0$ is a parameter that controls the damping term [104].

---

[6]See [2] for the connection between the Gamma freezing condition and the more stringent *minimal distortion* shift condition originating from the work of Smarr and York [99, 98] which minimizes the time variation of the conformal metric $\tilde{g}_{ij}$ over the spatial slice.

1.2.3 Initial Data

With a formulation and a gauge, we turn to the initial data. Because of the constraint equations, the initial data are not freely specifiable. Considering g_{ij} and K_{ij} which are symmetric, there are 12 degrees of freedom, but the constraint equations limit 4 of these, so there are only 8 degrees of freedom. The degrees of freedom specify the properties of the spacetime, such as a single nonspinning black hole, a black hole binary of a certain mass ratio, etc. In general the 12 quantities don't divide easily into constrained and free data. Considering the form of the constraint equations, we must rewrite them into a form where we can specify the free quantities and solve for the four constrained quantities. There are different approaches to that process including the York-Lichnerowicz conformal decomposition, the Bowen-York method (a particular case of the York-Lichnerowicz approach), the conformal thin sandwich approach [112], the quasicircular method [41, 18], and quasiequilibrium methods [48].

1.2.3.1 Bowen-York Initial Data

The York-Lichnerowicz conformal decomposition rewrites the spatial metric in terms of a conformal factor, $g_{ij} = \Psi^4 \bar{g}_{ij}$ [67, 111, 112]. Though similar to the BSSNOK approach to evolution, the conformal factor is not the same. The conformal metric is specified, as well as K and a symmetric trace-free tensor ($\bar{M}_{ij}$) that provides part of the conformal extrinsic curvature $\bar{A}_{ij}$. From these specified parameters, the York-Lichnerowicz method solves for Ψ and $\bar{A}_{ij}$, which in turn provide g_{ij}

and K_{ij}. For nonspinning, non-boosted black holes, the solution is exact for both and reduces to the Brill-Lindquist initial data. Bowen-York initial data [23] add linear momentum and spin to the black holes. It starts with the York-Lichnerowicz conformal decomposition, and assumes the space is conformally flat and that $K = 0$. If $\bar{M}^{ij} = 0$, there is an exact solution for the conformal trace-free extrinsic curvature

$$\bar{A}_{ij} = \frac{3}{r^2} \left[n_i P_j + n_j P_i + n_k P^k (n_i n_j - \delta_{ij}) \right] + \frac{3}{r^3} \left[\epsilon_{kim} n_j + \epsilon_{kjm} n_i \right] S^m n^k . \quad (1.25)$$

n^i is the unit vector pointing away from the origin, and P^i and S^i are the linear and angular ADM momenta of a black hole at the origin. For N black holes, $\bar{A}_{ij}$ is the sum of N terms of the form Eq. (1.25), each term representing one of the black holes with r defined as the coordinate distance from that black hole and P^i and S^i defined as the linear and spin angular momentum of the black hole. In this case the momenta are parameters, but they do correspond to momenta associated with each black hole.

For a single black hole with no spin and no linear momentum, the Bowen-York initial data represent a slice of fully extended Schwarzschild spacetime, an Einstein-Rosen wormhole, which is two asymptotically flat spacetime regions connected by a throat at or within the event horizon of each region. This initial slice is two three-dimensional regions from which an open three-ball bounded by the event horizon has been removed and which are connected by the shared event horizon. For more general Bowen-York initial data of N black holes with spins and linear momenta, the slice consists of a common asymptotically flat region plus N separate asymptotically flat regions, with a throat connecting each of the N separate regions to the common

23

region.

With $\bar{A}_{ij}$ given by the generalization of Eq. (1.25), we solve for the conformal factor. In doing this, the boundary conditions are important. The outer boundary condition is determined by the requirement of asymptotic flatness, which means that the metric and extrinsic curvature should go to their flat space values. Since Ψ becomes singular at the coordinate locations of the black holes, those points must be removed from the solution domain. The missing points are called *punctures*. To avoid the punctures, we need an inner boundary condition. One approach is to demand an isometry that identifies points inside each throat with points outside the throats. Such a condition provides inner boundary conditions, but it also requires additional terms in the conformal extrinsic curvature to create the isometry.

The more common approach for Bowen-York initial data is the *puncture* approach pioneered by Brandt and Brügmann [26]. In this approach, the conformal factor is written as $\Psi = \Psi_{BL} + u$ where Ψ_{BL} is the Brill-Lindquist portion of the conformal factor given by

$$\Psi_{BL}(\vec{r}) = \sum_i^N \frac{m_i}{2r_i}. \tag{1.26}$$

r_i is the coordinate distance to black hole i. The Brill-Lindquist conformal factor describes two nonspinning black holes initially at rest with $u = 1$. In that case, the bare mass of ith black hole m_i, is the contribution of that black hole to the total ADM mass of the system. In the general case, the bare masses are parameters that, with the coordinate separations and the momenta, determine the total ADM mass. Ψ_{BL} is the singular portion of the conformal factor, and u is the remaining

nonsingular portion, which is solved for numerically. In solving for u, it is no longer necessary to exclude the punctures, and therefore there is no need for inner boundary conditions. u is calculated with an elliptic solver such as `AMRMG` [28] or `TwoPunctures` [7]. All the simulations in this dissertation use puncture initial data, Bowen-York initial data solved by the puncture approach, from one of these two solvers.

1.2.4 Evolution

Since the evolution equations are hyperbolic partial differential equations (PDEs), evolving the spacetime metric requires methods to both evaluate the spatial derivatives and perform the time integration of the equations.

1.2.4.1 Spatial Derivatives

In the numerical relativity community the most common method to evaluate spatial derivatives is finite differencing. For a function $f = f(x)$, finite difference methods replace f' and higher derivatives, at a point x_0, with difference quotients, linear combinations of f at points neighboring x_0 divided by the spacing of the neighboring points. The formulas for f' and higher derivatives can be derived from Taylor polynomials $T(x)$ of the function f,

$$f(x) = T_n(x) + R_n(x), \tag{1.27}$$

where

$$T_n(x) = f(x_0) + \sum_{k=1}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k, \tag{1.28}$$

and

$$R_n(x) = \frac{f^{(n+1)}(\varepsilon)}{(n+1)!}(x - x_0)^{n+1}. \tag{1.29}$$

$T_n(x)$ is the nth-order Taylor polynomial, and $R_n(x)$ is the nth-order residual term. The derivatives of f up to $f^{(n)}$ must be continuous in $[x_0, x]$ and $f^{(n+1)}$ must exist in (x_0, x). ε lies in (x_0, x). Since we want to approximate derivatives at points on a discrete regularly spaced grid using other points from that grid, we can rewrite this as

$$f_j = f_0 + \sum_{k=1}^{n} \frac{f_0^{(k)}}{k!}(jh)^k + O(h^{n+1}), \tag{1.30}$$

where j indexes grid points relative to x_0, h is the space between adjacent points, $f_0 = f(x_0)$, $f_j = f(x_0 + jh)$, $f_0^{(k)} = f^{(k)}(x_0)$, and $O(h^{n+1})$ is a term proportional to h^{n+1}. Using a linear combination of Eq. (1.30) for different j to eliminate all $f^{(k)}$ terms except $f^{(m)}$, we create a stencil for $f^{(m)}(x_0)$ from f at x_0 and neighboring grid points with an error proportional to h^{n+1-m}.

The other commonly used approach to evaluating spatial derivatives is spectral methods. With spectral methods, a function $f(x)$ is written as a linear combination of a set of *basis functions* $f_i(x)$, such as the Fourier series. Because the basis functions are analytic, and their derivatives are known, evaluating the derivative of f, is a simple algebraic operation. So if

$$f(x) = \sum_i a_i f_i(x), \tag{1.31}$$

then

$$f'(x) = \sum_i a_i f_i\prime(x). \tag{1.32}$$

26

Spectral methods have exponential convergence, but they also require very careful treatment of boundary conditions and subdomains, so finite difference methods have generally preceded spectral methods in reaching milestones in the general problem of black hole binary mergers.

1.2.4.2 Time Integration

The most common approach to time integration in the numerical relativity community is the *method of lines* using the fourth-order Runge-Kutta (RK4) algorithm for integrating ordinary differential equations (ODEs). Given a PDE

$$\partial_t u = F(u, \partial_x u, \partial_x^2 u, t) \,, \tag{1.33}$$

where u is one or more functions that constitute the system, and F is a function of u and its spatial derivatives, the method of lines rewrites the original PDE as an ODE given by

$$\frac{d}{dt} u = f(u, t) \,, \tag{1.34}$$

where f represents the function that results from the spatial derivative treatment (such as finite differencing or spectral methods). Then an integration method such as RK4 solves the resulting ODE, Eq. (1.34). The method of lines changes the PDE into an ODE by making the spatial derivative evaluations independent of the integration, and since ODEs are simpler to solve than PDEs, in general, the method of lines simplifies the integration.

With the RK4 algorithm, given Eq. (1.34) and the solution at t_n, $u_n \equiv u(t_n)$,

the solution at the next time step $t_{n+1} = t_n + \Delta t$ is

$$u_{n+1} = u_n + \left(\frac{k_1 + 2k_2 + 2k_3 + k_4}{6} \right) \Delta t \,, \qquad (1.35)$$

where

$$k_1 = f(u_n, t_n) \,, \qquad (1.36)$$

$$k_2 = f(u_n + k_1 \tfrac{1}{2}\Delta t, t_n + \tfrac{1}{2}\Delta t) \,, \qquad (1.37)$$

$$k_3 = f(u_n + k_2 \tfrac{1}{2}\Delta t, t_n + \tfrac{1}{2}\Delta t) \,, \qquad (1.38)$$

and

$$k_4 = f(u_n + k_3 \Delta t, t_{n+1}) \,. \qquad (1.39)$$

RK4 is an explicit time integration method since u_{n+1} is an explicit function of u_n. Precise cancellations in the sum of k terms results in an error proportional to $(\Delta t)^5$ at each timestep. Because the number of timesteps in a simulation is inversely proportional to the size of the timestep, the integration error for the simulation is fourth-order.

Implicit integration methods provide an alternative to explicit methods such as RK4. In such methods, u_{n+1} is given by an implicit equation involving u_{n+1}, u_n, and possibly earlier values of u. A simple example is

$$u_{n+t} = u_n + f(u_{n+t}, t_{n+1})\Delta t \,. \qquad (1.40)$$

Solving such implicit equations is more complicated and time-consuming than explicit methods, and in general implicit methods cannot be used with the method of lines since the spatial discretization is intimately connected to the time integration with implicit methods. The advantage of these methods is that they allow for larger

timesteps without loss of stability. We do not apply implicit methods in our simulations, but the possibility of taking larger timesteps partially motivates my work on the moving patches technique, which is discussed in Chapter 4.

1.2.4.3 Boundary Conditions

We must specify boundary conditions at the boundaries of the simulation domain. In the case of simulations using finite differencing, the boundary conditions allow us to evolve the data at the outer edge of the grid. The boundary conditions provide data for the finite difference stencil points that lie beyond the grid. When interior points are removed from the domain of a numerical relativity simulation, a technique called *excision* that is discussed below, we also need boundary conditions at the excision boundary, an inner boundary.

At the outer boundary, numerical relativity simulations often use the radiative boundary condition. The radiative boundary condition assumes outgoing radiation at the boundary, a Sommerfeld-type boundary condition. For a scalar field u in three dimensions the condition is

$$U_{,r} + U_{,t} = 0 \,, \tag{1.41}$$

where r is the distance from the wave source and $U \equiv ur$ [109]. This condition matches the outgoing radiation of a field u with a mean value of 0, and it produces no inbound radiation for a linear wave with spherically symmetric data and a spherical boundary.

For spacetime, the radiative boundary condition describes variation about a

flat-space metric. Though this is an artificial condition, it approximates outgoing radiation for boundaries sufficiently far from the strong field region.[7] Since the flat-space values of the spatial metric and extrinsic curvature are δ_{ij} and 0, respectively, the radiative boundary condition becomes

$$(\partial_t + \partial_r)r(g_{ij} - \delta_{ij}) = 0 \,, \tag{1.42}$$

and

$$(\partial_t + \partial_r)rK_{ij} = 0 \,. \tag{1.43}$$

These conditions are simple to implement, and given the wavelike nature of the fields at the outer boundary, the radiative boundary condition approximates them well. However, it does not satisfy the constraint equations, so the outer boundary introduces nonphysical solutions that propagate into the simulation domain. For this reason, the outer boundaries must be located far enough from the wave zone and strong field regions that such error does not have sufficient time in a simulation to reach those regions. An alternative approach is to compactify the coordinates so that the edge of the finite grid is an infinite coordinate distance from the origin. This can be thought of as applying an additional metric to the grid for the purpose of calculation. Such a metric is conformally flat, with a conformal factor that becomes large at the edge of the grid. Unfortunately compactification scatters outbound radiation back into the grid, so it introduces errors as well.

---

[7] "Sufficiently far" can be estimated as the distance at which the field values of an analytic approximation of the simulation spacetime are sufficiently close to flat space. Since the main concern is error introduced by the artificial boundary condition, an empirical judgment based on observation of error originating from the outer boundary is often more useful.

1.2.4.4 Moving Punctures

Once Bowen-York initial data are solved by the puncture technique, the data can be evolved. In the initial data case, the conformal factor Ψ is separated into a singular part $\Psi_{BL}(x^i)$ and a nonsingular part $u(x^i)$, which is solved numerically. To continue using the puncture technique for the evolution of the data, the puncture locations must be kept constant since Ψ_{BL} is an analytical function of position. To accomplish this, the shift must go to 0 at the puncture locations, so $\beta^i(t=0)=0$ and $\dot{\beta}^i \propto \Psi^{-n} = 0$ for $n > 0$ at the puncture locations. While this gauge choice does keep the puncture coordinates constant, it also forces the grid to "stretch" to accommodate the physical motion of the black holes. The decreasing separation of the binary requires that the physical spatial metric between the punctures go to 0 through the conformal factor Ψ or the conformal spatial metric $\tilde{g}_{ij}$. The result is either the evolved ϕ must go to $-\infty$ or the inverse conformal spatial metric $\tilde{g}^{ij}$ must diverge. In either case, this creates large field gradients, which in turn lead to large numerical errors and instability. The tangential motions of the punctures cause a twisting of the grid that also causes large errors and instabilities.

In response to these challenges, the numerical relativity groups at NASA Goddard Space Flight Center and University of Texas Brownsville independently developed the moving punctures approach [34, 14]. By allowing the punctures to move on the grid, each group evolved the merger and ringdown of a black hole binary. Though the groups took slightly different approaches, the procedure had the same two elements in each case. First, the conformal factor was no longer decomposed

into a singular analytical part and an evolved nonsingular part. Second, each group used a shift gauge condition that did not vanish at the punctures, but instead, the shift at the punctures was responsible for moving the punctures on the grid. The Goddard group evolved the BSSN $\phi = \log \Psi$ variable as before, but without splitting out the singular part, while the Brownsville group evolved a new variable $\chi \equiv \Psi^4 = e^{-4\phi}$. The groups used slightly different *Gamma driver* conditions, but in both cases, $\partial_{t^2} \beta^i \propto \partial_t \widetilde{\Gamma}^i$ and there was no power of Ψ in the shift evolution equations driving the shift to 0 at the punctures. The moving punctures approach has become the established choice for numerical relativity puncture evolutions. All of the simulations discussed in this dissertation used the moving puncture approach.

1.2.5 Tools and Techniques

1.2.5.1 Excision

The region of spacetime bounded by an event horizon is causally disconnected from the rest of spacetime, and it contains the physical singularity of a black hole. So one way to avoid the complications created by the singularity in numerical relativity simulations is to *excise* a region of the simulation domain within the event horizon that includes the singularity. If the boundary conditions at the excision boundary are handled properly, the excision will have no effect on the physical results in the simulation domain. This idea is attributed to Unruh by Thornburg [102], and it was first implemented successfully by Seidel and Suen [96]. Though it eliminates the problem of the singularities, it is computationally complex for general simu-

lation domains. Excision was simplified by Alcubierre and Brügmann with their introduction of simple excision [4]. They found that it is unnecessary to impose boundary conditions at the excision boundary since all the characteristic fields are ingoing within the event horizon, which includes the excision region by construction. The introduction of moving punctures significantly reduced the use of excision, since there was no need to excise singularities with this technique, and excision is computationally expensive. Excision is still used in the numerical relativity community, most notably by the Caltech/Cornell collaboration and when the initial data include an excision region.

1.2.5.2   FMR

Large field gradients near the black holes, error propagating inward from the outer boundary of the simulation grid, and finite computational memory create a need for multiple resolutions in numerical relativity simulations. The large field gradients present in the strong field region near the black holes require high resolution in that region. Accurately resolving the gravitational radiation in the wave zone far from the black holes requires significantly less resolution. The region between the strong field region and the wave zone requires an intermediate resolution sufficient to minimize dissipation of the waves as they propagate through it. Placing the outer boundary of the grid far from the wave zone delays contamination from error propagating in from the outer boundary. While in principle the high resolution of the strong field region could be used for the entire grid, this would be very wasteful

of finite computational memory, especially for longer simulations where the outer boundary is farther from the wave zone. Fixed mesh refinement (FMR) divides the simulation grid into zones and refines the grid in each zone to match the resolution requirements of the physical region of the grid. The zone resolutions are fixed at the beginning of the simulation. Adaptive mesh refinement (AMR) also divides the grid into zones of different resolution, but it allows the portion of the grid included in each refinement zone to change during the course of the simulation. This allows high resolution around each black hole without an unnecessarily high resolution region between the black holes. In simulations of widely separated black holes, this allows more efficient use of memory. The high resolution zones track with the black holes. AMR allows the zones outside the strong fields to vary as the gauge evolves in those regions.

1.3   Implementation

HAHNDOL is the finite difference numerical relativity code written by the NASA Goddard group. It solves for the initial data as specified by input parameters, and it evolves the resulting data according to the BSSNOK equations using gauge conditions specified by the input parameters. The PARAMESH [74, 89, 88, 1] library of Fortran 90 modules and subroutines provides the simulation grid as specified by the application, HAHNDOL. The library abstracts the details of the grid as well as the parallelization of the grid, and HAHNDOL implements the physics of numerical relativity and methods of numerical analysis on the grid.

1.3.1 The PARAMESH Grid

PARAMESH creates a simulation grid of the size and dimension (1, 2, or 3) specified by the application, manages the field data and the information about the grid, and carries out all of the tasks of maintaining, updating, and modifying the grid. It decomposes the simulation grid into blocks and distributes the blocks over the available processors, which allows parallel processing by the application using PARAMESH. Also, the library handles the refinement and derefinement of the grid.

PARAMESH further divides each block into cells on a Cartesian grid. The blocks are logically identical, which means that for each dimension, the number of cells is the same for all blocks. Each block consists of interior cells and guard cells that border the interior cells on the block faces. Each face has the same number of layers of guard cells. HAHNDOL uses three levels, currently. These guard cells contain data from neighboring blocks, which allows local calculations on interior cells of finite differences and other quantities that require data from neighboring cells. PARAMESH maintains the coordinates and size of each block and uses this coordinate information to determine the neighbors of each block, which it uses to fill the guard cells of the block. At the boundaries of the simulation grid, PARAMESH uses boundary conditions supplied by the application to fill the guard cells.

PARAMESH provides multiple resolutions on the grid for the application, through FMR and AMR. It refines and derefines the blocks as specified by the application. For any spatial region, the highest resolution blocks covering the region are *leaf* blocks. When the application indicates that higher resolution is required for a block,

PARAMESH refines the block by creating 2^D child blocks of half the size and twice the resolution where D is the dimension of the grid. Then PARAMESH fills the child blocks from the original block (*prolongation*). The child blocks become leaf blocks, and the original block becomes a *parent* block. Conversely, when the application indicates that blocks should be derefined, PARAMESH fills the parent blocks from the indicated leaf blocks (*restriction*), changes those parent blocks to leaf blocks, and deactivates the child blocks. PARAMESH only does this when all the child blocks of a parent block are marked for derefinement. The refinement and derefinement of blocks provides true AMR to the application, instead of repeated FMR.

PARAMESH uses a common interpolator for three operations that are necessary when the simulation grid contains multiple resolutions. In that case the grid consists of multiple refinement levels. The refinement level of a block is the number of refinements that generated the block, and all of the blocks of the same refinement level comprise a refinement level of the grid, a region of common resolution. PARAMESH interpolates to fill the guard cells of blocks on refinement boundaries, where different refinement levels meet. It also interpolates in prolongation for block refinement and in restriction for block derefinement. The interpolator uses Lagrange polynomials of an order specified by the application. Because HAHNDOL uses three levels of guard cells at present, it specifies sixth-order polynomials for the PARAMESH interpolator. This results in sixth-order error from PARAMESH interpolation to match the sixth-order finite difference error in HAHNDOL, as discussed below.

1.3.2 The HAHNDOL Application

1.3.2.1 Numerical Analysis

HAHNDOL uses standard numerical analysis techniques. It uses sixth-order centered finite differencing for spatial derivatives except those appearing in advection terms. For these, it uses upwind differencing, in which the finite difference stencil includes more points on the side away from the black hole. This produces fifth-order error. For time integration, HAHNDOL uses the method of lines with RK4 integration, which results in error that is fourth-order in the timestep. In practice, the finite difference errors are much larger than the time integration errors, and so they produce the leading order error. HAHNDOL also includes its own sixth-order interpolator that is separate from the PARAMESH interpolator. Similar to the finite difference operators, the interpolator uses Eq. (1.30) to create an interpolation stencil that has sixth-order error. As I will discuss in Chapter 4, the moving patches technique uses the HAHNDOL interpolator intensively. To prevent introducing lower-order error from the interpolator in the implementation of the technique, I increased the order of the HAHNDOL interpolator from fourth-order to sixth-order.

HAHNDOL uses a number of other techniques common in the numerical relativity community. Among these are constraint damping and artificial dissipation. In the limit of infinitely high resolution, quantities proportional to Δx go to 0. Assuming that the constraints converge, adding multiples of the constraints to any of the evolution equations should have no effect in the continuum limit. However, for finite resolutions, such additions do change the behavior of the evolution in a

convergent manner. Constraint damping refers to the technique where terms are added to reduce the constraint violations. HAHNDOL uses the form of constraint damping suggested by Duez *et al.*[42]. Again using additional terms that scale with the resolution, and therefore vanish in the continuum limit, HAHNDOL also uses artificial dissipation, which reduces high-frequency noise in the evolved fields. The artificial dissipation terms are finite difference approximations of high-order spatial derivatives scaled by the time step Δt, with the order of the spatial derivatives chosen sufficiently large to prevent introduction of lower-order error to the evolution. HAHNDOL uses Kreiss-Oliger dissipation [64].

1.3.2.2 Puncture Tracker

Finding a black hole in a simulation is not simple. Our lapse conditions produce spatial slices that avoid the physical singularity because of the instability that occurs if the singularity is on the grid. So the intuitive choice of the physical singularity for the black hole location does not work. Instead, we use the puncture position as the black hole location, keeping in mind that it is gauge-dependent and that we must be careful in how we interpret it.

HAHNDOL includes a puncture tracker. Based on the empirical observation that in the neighborhood of a puncture,

$$\beta^i(x^i) = k^2(x^i - x_p^i) + \beta^i(x_p^i),\qquad(1.44)$$

where k is a real constant, and x_p^i is the puncture location, the puncture tracker uses the current puncture location and the shift at that location to predict the puncture

location at the next iteration using an Euler integration,

$$x^i_{p,n+1} = x^i_{p,n} - \beta^i(x^i_{p,n})dt \,. \tag{1.45}$$

Though the integration is first-order in time, the calculated puncture position is driven back to the "true" puncture position by the first term in Eq. (1.44). The "true" puncture position is the point that represents the other side of the wormhole. (Once the puncture simulation begins, the puncture no longer represents the second asymptotic infinity, but rather the most distant points of the region within the throat [53, 55, 54, 27]).

1.3.2.3 Apparent Horizon Finder

In general relativity, an event horizon is the boundary of a region which no null geodesics leave. For black hole spacetimes, often this is described as the boundary of the region from which light cannot escape, with the outgoing light rays on the event horizon "frozen" to the boundary. In numerical relativity simulations, when there is no analytic expression for the metric, establishing the location of the event horizon would require evolving to $t = \infty$ to establish the region from no null geodesics leave. So in numerical relativity simulations, the related concept of the *apparent horizon* is used instead. First, consider a single black hole and two neighboring outgoing null geodesics (for which the spatial component of the velocity points away from the black hole). If the geodesics escape from the black hole, the separation between the geodesics on the spatial slices increases as time increases. If the geodesics fall into the black hole, the separation decreases, and if they are

marginally trapped, the separation remains constant. More formally, at time t_0, for a closed, two-dimensional surface σ around the black hole, if the tangent vectors to outgoing null geodesics converge, the null geodesics are trapped, and the volume bounded by them is decreasing. However, if the bounded volume is constant, the null geodesics are marginally trapped. This property is measured by *expansion* which is the trace of the projection into σ of the covariant derivative of the tangent vector,

$$\theta = g^{\mu\nu}(\perp_\sigma)^\alpha_\mu(\perp_\sigma)^\beta_\nu \nabla_\alpha u_\beta \,, \tag{1.46}$$

where $\perp_\sigma$ is the projection operator and u^μ is the tangent vector. So a marginally trapped surface is a $\theta = 0$ level set. For more general spacetimes, including black hole binaries, there can be more than one marginally trapped surface. The apparent horizon (AH) is the outermost marginally trapped surface. For any physically reasonable spacetime, the apparent horizon does not lie outside the event horizon, so the apparent horizon provides a conservative estimate of the event horizon. In particular, it provides a conservative boundary for the simulation domain region that is causally disconnected from the outer simulation domain, which is interpreted as the universe outside a black hole. This is necessary in numerical relativity simulations that use excision since the excision boundary must be inside the apparent horizon. Also, the apparent horizon can be used to measure the mass and spin of individual black holes. HAHNDOL locates and characterizes apparent horizons using the Thornburg apparent horizon finder [103].

For spacetimes representing multiple black holes, the horizon mass measures

the mass of each black hole.[8] From the apparent horizon area A_{AH} and the spin S of the black hole, we calculate the horizon mass m with the Christodoulou formula [40],

$$m^2 = m_{irr}^2 + \frac{S^2}{4m_{irr}^2} \, , \tag{1.47}$$

where $m_{irr} = \sqrt{A_{AH}/16\pi}$. With the horizon mass definition, we can define the mass ratio q of a black hole binary by

$$q \equiv \frac{m_1}{m_2} \, , \tag{1.48}$$

and the symmetric mass ratio by

$$\eta \equiv \frac{m_1 \, m_2}{(m_1 + m_2)^2} \, . \tag{1.49}$$

1.4 Analysis

1.4.1 Observables

Because general relativity is covariant, we must be careful in interpreting tensor (including vector) components, since those components depend on the particular coordinate system we are using. Such quantities are gauge-dependent, and therefore it is more difficult to identify physical effects from the effects of the particular

---

[8]In this case, the apparent horizon of a black hole must be defined as the outermost marginally trapped surface *surrounding the black hole.* This definition is not rigorous since it requires a location for the black hole. Also, as the black holes approach, a common outermost marginally trapped surface forms, and this definition breaks down. In practice, the black holes can be associated with puncture locations or other meaningful coordinate positions, and until the common horizon forms, we can consider horizons of the individual black holes.

gauge. When possible, it is better to consider *observables* that do not depend on gauge (scalars) or where the gauge is clearly defined.

1.4.1.1 Radiation: ψ_4, $\dot{h}$, and h

General relativity predicts the existence of gravitational waves, distortions of spacetime that propagate through spacetime, as electromagnetic waves propagate through vacuum, and gravitational radiation is among the most important results of numerical relativity simulations. Gravitational radiation is time-varying curvature of spacetime that propagates away from a source (such as a black hole binary) and which diminishes in amplitude inversely with distance from the source. In this way, it is similar to electromagnetic radiation.

In vacuum, the Weyl tensor C_{abcd}, which is the trace-free part of the Riemann tensor R_{abcd}, fully describes the curvature of spacetime ($R_{abcd} = C_{abcd}$ in vacuum). It can be decomposed into five complex scalars, the *Weyl* scalars (ψ_0 - ψ_4), using the Newman-Penrose formalism. When so decomposed, the outgoing radiation component is ψ_4. The Newman-Penrose formalism decomposes tensors, such at C_{abcd}, with respect to a *null tetrad*, a set of four null (zero-length) basis vectors. This tetrad consists of two real vectors, ℓ^a and n^a, and two complex vectors, m^a and $\bar{m}^a$ (the conjugate of m^a) which satisfy the orthogonality conditions

$$\ell^a m_a = n^a m_a = \ell^a \bar{m}_a = n^a \bar{m}_a = 0\,. \tag{1.50}$$

Often the vectors are normalized by the conditions

$$\ell^a n_a = 0\,, \tag{1.51}$$

and

$$m^a \bar{m}_a = 0 \,, \tag{1.52}$$

a convention which I follow in this dissertation. The Weyl scalars are five components of C_{abcd} in the tetrad basis which fully specify the ten independent components of C_{abcd}. If we impose the three additional requirements that ℓ^a lies along outgoing null geodesics, n^a lies along ingoing null geodesics, and m^a and $\bar{m}^a$ lie along orthogonal spatial directions, ψ_4 which is given by

$$\psi_4 \equiv C_{abcd}\, n^a \bar{m}^b n^c \bar{m}^d \,, \tag{1.53}$$

is the component of C_{abcd} that represents outgoing radiation, a wave propagating along outgoing geodesics with an inverse fall-off in amplitude along the geodesics [101]. This is the gravitational radiation which we are interested in.

ψ_4 is calculated directly from the evolved data in a numerical relativity simulation. It measures spacetime curvature. Gravitational wave detectors measure strain $h \equiv \Delta L / L$, where L is a characteristic length of the detector and ΔL is the change in length as a gravitational wave encounters the detector. Following the derivation in Schnittman *et al.*[95] and Chap. 35 of [82], we connect h with ψ_4 via $R_{abcd} (= C_{abcd}$ in vacuum) by the relation

$$\psi_4 = -(\ddot{h}_+ - i\ddot{h}_\times) \,, \tag{1.54}$$

where h_+ and $h_\times$ are the two orthogonal polarizations of the strain. This relation assumes that h is written in the transverse-traceless gauge, a choice of coordinates in which h is a purely transverse wave with no time component. It also assumes that

ψ_4 is computed with a tetrad satisfying the three additional tetrad requirements described above.

While in some cases this tetrad is specified by coordinate expressions, such as the Kinnersley tetrad for small perturbations of Kerr spacetime, in general the correct tetrad is not known exactly for numerical relativity simulations, and a tetrad must be computed, and it only approximates the correct tetrad. Some tetrads, such as the quasi-Kinnersley tetrad, better satisfy the additional tetrad requirements, but for 3+1 simulations, the symmetric tetrad,

$$\ell^a \equiv \frac{1}{\sqrt{2}}(\hat{\tau} + \hat{r})^a, \tag{1.55}$$

$$n^a \equiv \frac{1}{\sqrt{2}}(\hat{\tau} - \hat{r})^a, \tag{1.56}$$

$$m^a \equiv \frac{1}{\sqrt{2}}(\hat{\theta} + i\hat{\varphi})^a, \tag{1.57}$$

and

$$\bar{m}^a \equiv \frac{1}{\sqrt{2}}(\hat{\theta} - i\hat{\varphi})^a, \tag{1.58}$$

where $\hat{\tau}$ is the future-pointing unit vector normal to the spatial slice at all points, and $\hat{r}$, $\hat{\theta}$, and $\hat{\phi}$ asymptotically approach the standard spherical polar basis vectors as the distance from the origin goes to infinity, offers greater efficiency due to its computational simplicity. All the simulations in this dissertation use the symmetric tetrad to calculate ψ_4.

To better understand the structure of ψ_4, we decompose it into components in a suitable basis. Because the gravitational field is a rank-2 tensor, a natural basis is

the set of spin-weight -2 spherical harmonics $_{-2}Y_{lm}$ [85, 101]. In terms of this basis,

$$\psi_4 = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{m=\ell} \psi_{4,\ell m} \, _{-2}Y_{\ell m} \,, \tag{1.59}$$

with $\psi_{4,\ell m}$ the contribution of the (ℓ, m) harmonic to ψ_4. With such a decomposition, we can more easily compare the radiation from different simulations. Also, because each mode of the radiation corresponds to a multipole moment in a post-Newtonian description of the source, the radiation has a simple modal structure in this basis. h and $\dot{h}$ can be decomposed into components in the same basis in an analogous manner.

1.4.1.2 Mass and Momentum

A natural observable to consider for a spacetime is the total energy (mass) of a spacetime, which is the energy density from Einstein's equations integrated over all space. Using the Hamiltonian constraint Eq. (1.2) this becomes

$$M_{ADM} = \lim_{r\to\infty} \frac{r^2}{16\pi} \oint \hat{r}^k \left(D_j^0 h_k^j - D_k^0 h \right) d\Omega \,, \tag{1.60}$$

where $\hat{r}^i$ is the unit normal vector pointing away from the origin and $d\Omega$ is the solid angle element. Note that this definition of the total energy of the spacetime, the *ADM mass* [8], requires the integral be evaluated at spatial infinity. For simulations with boundaries at a finite distance, this is not possible, and instead the integral is integrated at $r \geq 50\, M_{ADM}$, where the fields are sufficiently weak to approximate the flat-space fields are infinity. Richardson extrapolation of M_{ADM} to spatial infinity from the values calculated at several radii can mitigate the error introduced by such a finite radius evaluation.

Like the total energy, the total momentum of the system can be defined using the momentum density from Einstein's equations and Eq. (1.3) [8]. It takes a form similar to M_{ADM},

$$P^i_{ADM} = \lim_{r \to \infty} \frac{r^2}{8\pi} \oint \hat{r}^j \left(K^i_i - \delta^i_j K \right) d\Omega \,. \tag{1.61}$$

From the standard definition of angular momentum $(\vec{r} \times \vec{p})$, the ADM angular momentum is defined in a similar way [8] by

$$J^i_{ADM} = \lim_{r \to \infty} \frac{r^2}{8\pi} \oint_S \hat{r}^l \epsilon^{ijk} x_j \left(K_{kl} - \delta_{kl} K \right) d\Omega \,, \tag{1.62}$$

where ϵ^{ijk} is the Levi-Civita tensor.

Though the ADM quantities defined above remain constant, gravitational waves carry energy and momentum away from a radiating system, such as a black hole binary. The radiated energy and momentum can be calculated directly from ψ_4 using

$$\frac{dE}{dt} = \lim_{r \to \infty} \frac{r^2}{16\pi} \oint \left| \int_{-\infty}^{t} \psi_4 dt' \right|^2 d\Omega \,, \tag{1.63}$$

and

$$\frac{dP^i}{dt} = \lim_{r \to \infty} \frac{r^2}{16\pi} \oint \hat{r}^i \left| \int_{-\infty}^{t} \psi_4 dt' \right|^2 d\Omega \,. \tag{1.64}$$

Like the ADM energy and momentum, the radiated power and momentum flux are evaluated at spatial infinity. The time integrals from ∞ to t reflect the effect of all past values of ψ_4 on the present flux. The waves also carry away angular momentum. In Cartesian coordinates the expressions for the angular momentum flux in terms of ψ_4 for $\dot{J}^x$ and $\dot{J}^y$ use derivative operators that account for the spin weight of gravitational waves. Because the simulations for this dissertation

only involve angular momentum flux in the z direction, I include the much simpler definition of $\dot{J}^z$,

$$\frac{dJ^z}{dt} = -\lim_{r\to\infty} \frac{r^2}{16\pi} \Re \left\{ \oint \left(\int_{-\infty}^{t} \psi_4^* dt' \right) \partial_\phi \left(\int_{-\infty}^{t} \int_{-\infty}^{t'} \psi_4^* dt'' dt' \right) d\Omega \right\} . \qquad (1.65)$$

∂_ϕ is the partial derivative with respect to the azimuthal coordinate of spherical polar coordinates. The flux only includes the real portion of the integrated expression.

1.4.2 Error

Numerical relativity simulations provide approximate solutions to Einstein's equations, and therefore the results inevitably include error. To assess the quality of the results, we must estimate the magnitude of the error. To improve our simulations, we need to understand the sources of the error. This is particularly important for the discussion of moving patches in Chapter 4.

1.4.2.1 Error Sources

For any application of numerical analysis, different types of error result from the representations of the numbers, the expressions, and the equations, as well as the implementation of the numerical methods. *Round-off error* that results from representing real numbers in the finite memory found in digital computers is the most fundamental error source in the implementation of numerical analysis methods. While rational numbers can be stored as two real numbers, irrational numbers must be maintained symbolically, which is highly impractical, or truncated eventually. Beyond this, digital computers are nearly exclusively binary, which means that

only integer powers of two can be represented exactly. Such fundamental issues of numerical analysis are beyond the scope of this dissertation, so I mention them only.

Discretization error occurs when a continuous problem is replaced by a discrete problem that approximates it (*discretization*), since the solution of the discrete problem does not solve the original continuous problem. For example, a differential equation $y' = f(x)$ could be changed into the algebraic equation $D[y(x_i)] = f(x_i)$ for the points x_i of a grid by replacing the derivative of y by a finite-difference operator D and evaluating f at the grid points, but the solution $y(x_i)$ of the algebraic equation would generally not be a solution of the original differential equation, though it might approach it as the grid resolution increases, since f is evaluated at a finite number of discrete points.

When a function, operator, or procedure is approximated by a smaller number of iterations or terms, the difference between the original quantity and the approximation is the *truncation error*. Returning to the example of a finite differencing, a finite difference operator approximates the derivative of a function using Taylor polynomials of the function. In this case, the truncation error results from truncating the Taylor series of the function. For spectral methods, truncation errors result from the finite number of basis functions used to represent the fields. In our finite difference simulations, several error sources originate from truncation errors. These include finite differencing, RK4 integration, and interpolation by PARAMESH and HAHNDOL. As discussed previously, the truncation errors from finite differencing and both interpolators are sixth-order in the grid resolution, while the RK4 integrator is fourth-order in the timestep.

The other major type of error is errors that result directly from the physical model, both in the assumptions and the implementation of it. An example of this "nonphysical error" is the radiative boundary condition. It is an approximation of the true physics since it assumes purely outgoing radiation in all the fields at the grid boundary, and it does not satisfy the constraints. Therefore it introduces error from the outer boundary that propagates inwards. Another source of nonphysical error is evaluating observables at finite radii. These errors occur because the observables, such as ψ_4, the ADM mass and momenta, and the radiated energy and momenta, are defined in the limit that the observation points are infinitely far from the strong fields ($R \rightarrow \infty$). So all are subject to errors that scale as R^{-n} for some positive integer n. In fact, there may be error terms for multiple n, but in practice one term usually dominates, and that is often the leading order term. A final example of nonphysical error sources is dissipation. As discussed above, artificial dissipation introduces error that is higher-order in the grid spacing. In addition, the simulation grid produces dissipation that results in nonphysical error. The amplitude of waves propagating on the grid diminishes because of the finite resolution. The reduction depends on the wavelength, with shorter wavelengths reduced more down to the *Nyquist limit*, twice the grid spacing, below which shorter wavelengths cannot be resolved on the grid. In terms of frequency, the Nyquist limit is an effective low-pass filter. Because this differential dissipation of the radiation is nonphysical, it introduces error to the simulation.

1.4.2.2 Convergence Testing

In a numerical simulation of a quantity φ, convergence is the approach of the computed value f to the true value φ. A simulation using discrete time steps or a spatial grid of discrete points is convergent if

$$f_a - \varphi = \epsilon h_a{}^n \tag{1.66}$$

for a constant ϵ, where h_a is the discretization parameter, the time step or grid spacing, and n is the order of convergence. A larger value of n indicates more rapid convergence and therefore a more accurate simulation. When both time and space are discrete, the simulation has separate orders of convergence in time and in space.

Eq. (1.66) provides a definition of convergence for a simulation. It does not describe the error sources present in the simulation. A particular contribution to the simulation error may be comprised of terms of different orders. For example, the finite difference stencils result from eliminating lower-order terms from the Taylor expansion at a point by combining field values in the neighborhood of the point. In the case of the HAHNDOL sixth-order approximations of the first derivatives, the linear combination eliminates the f'' through $f^{(6)}$ terms. While the lowest-order error term is sixth-order, higher-order terms do contribute to the error. Usually these terms are insignificant, but if the resolution is too coarse, the higher-order terms become larger than the leading-order terms, and the convergence order increases. Therefore over-convergence often indicates under-resolved phenomena.

In general, different error sources scale with resolution at different orders, and so while each error source may have a clear convergence order determined by the

methodology, the aggregate convergence depends on the relative sizes of the different error terms, the ϵ coefficients in Eq. (1.66) for the various error sources. It is possible that higher-order error sources can dominate lower-order sources, and we find this is the case for the finite difference and integration errors in HAHNDOL simulations. The sixth-order finite difference errors are much larger than the fourth-order RK4 errors.[9] As in the case of individual error contributions, the aggregate convergence also varies when the resolution becomes too coarse.

For these reasons, we analyze the convergence of various quantities in our simulations to assess the quality of the simulation and to evaluate if the resolution is sufficient. While we could use the highest resolutions possible, efficiency demands we use resolutions where the various error sources are comparable in size. Also, if the error is dominated by terms of a specific order, and we determine that order, we can estimate the absolute size of errors.

In general, the true value φ and the constant ϵ, are not known in advance, so Eq. (1.66) is not directly useful, but if we have results from three resolutions, we can determine the order of convergence using Eq. (1.66) at the three resolutions (three-point convergence). After differencing the three resolutions of Eq. (1.66) in two pairs, forming the ratio of the differences and solving for one of the differences we have

$$|f_a - f_b| = \left| \frac{h_a{}^n - h_b{}^n}{h_b{}^n - h_c{}^n} \right| |f_b - f_c|, \tag{1.67}$$

---

[9]Though the integration occurs in time and the differencing occurs in space, the timesteps scale with the spatial resolution in our simulations. Since the scaling factor is the same for simulations at different resolutions, the two errors sources affect convergence in both time and space.

51

where $[a, b, c]$ label the resolutions. The value of n for which Eq. (1.67) is true is the order of convergence. Three-point convergence tests are often used to assess quantities derived from the evolved fields, such as the real and imaginary components or phase and amplitude of ψ_4.

Convergence testing is based the suggestion by Richardson [91] that a finite-difference solution for a quantity φ can be expressed as an expansion in the discretization parameter. Following Richardson's suggestion, we can estimate the error of a simulation based on the results from multiple resolutions. In the case where φ is unknown, we perform a three-point convergence test using Eq. (1.67). If the test shows that nth-order terms dominate the error, we difference Eq. (1.66) for two of the resolutions, and solve for the error term coefficient ϵ,

$$\epsilon = \frac{f_a - f_b}{h_a^n - h_b^n} \, , \tag{1.68}$$

and the error estimate is ϵh_c^n for the results from resolution h_c, which could be h_a, h_b, or a third resolution. We can use a similar procedure to estimate φ. Collectively, these estimation procedures are known as *Richardson extrapolation*. It can be used for any quantity for which the error is given by a power series in some parameter.

1.4.2.3   Constraint Violations

Though for many quantities φ is not known *a priori*, we know that the evolved fields should satisfy the constraint equations in the continuum limit. Therefore we define constraints

$$C_H \equiv R + K^2 - K_{ij}K^{ij} - 16\pi\rho \, , \tag{1.69}$$

and

$$C_{M^i} \equiv D_j(K^{ij} - g^{ij}K) - 8\pi j^i \,, \tag{1.70}$$

from the Hamiltonian constraint equation and the momentum constraint equations, Eqs. (1.2) and (1.3). Nonzero values of the constraints are *constraint violations*, and they quantify the departure of a solution from the physics, a measure of error.

Since ideally the constraint violations are zero, we know the true value, and we can perform two-point convergence tests of the constraint violations using two different resolutions. to determine the order of convergence (two-point convergence). Substituting C_H or C_{M^i} at two different resolutions for f_a in Eq. (1.66), forming the ratio of the resulting equations, and solving for the results at one resolution, we have

$$C_a = \left(\frac{h_a}{h_b}\right)^n C_b \,, \tag{1.71}$$

where C_a represents the constraint at resolution a. Again, the order of convergence is the value of n for which Eq. (1.71) is true.

1.4.2.4 Norms

To assess constraint violations in simulations of three-dimensional data in time, it is simpler to reduce a constraint violation at each time to a single datum using norms. For a function φ approximated numerically by f in a simulation, we define

three norms

$$f_{L1}(t) \equiv \frac{1}{N} \sum_{i=1}^{N} |f_i(t) - \varphi(t)| \ , \tag{1.72}$$

$$f_{L2}(t) \equiv \left[\frac{1}{N} \sum_{i=1}^{N} (f_i(t) - \varphi(t))^2 \right]^{\frac{1}{2}} \ , \tag{1.73}$$

and

$$f_{L\,\mathrm{inf}}(t) \equiv \max_{i=1}^{N} |f_i(t) - \varphi(t)| \ , \tag{1.74}$$

where in each case i labels the grid points of which there are N. Eq. (1.72) is the mean average of absolute error, while Eq. (1.73) is a norm sensitive to regions of larger error due to its root-mean-square form. Eq. (1.74) provides a ceiling on the absolute error. While these norms can be computed for any quantity φ, most often we use them for constraint violations, where $\varphi = 0$.

1.4.2.5   Puncture Track Angular Momentum

Finally, I present here a diagnostic related to Newtonian angular momentum that I developed for the moving patches technique, since it is not specific to moving patches, and it can be useful for simulations in general. I will discuss results of this error measure in Chapter 4.

Early in a binary simulation, any eccentricity in the system produces oscillations in the orbital frequency that dominate the secular frequency evolution and complicate comparisons of frequency across resolutions. This eccentricity results from angular momentum that exceeds the angular momentum of a quasi-circular orbit at the binary's separation. In the Newtonian limit, the oscillations in the

orbital frequency from eccentricity remain, but the angular momentum does not os-cillate since it is conserved. This suggests that angular momentum would be a better diagnostic of the system's dynamics. So from the puncture tracks I define a quantity $\ell_{PT} \equiv r^2 \omega$ which is the Newtonian orbital angular momentum of a unit mass in a circular orbit. The virtues of ℓ_{PT} are that we can calculate it from the puncture tracks and that it is not affected by eccentricity in the Newtonian limit. We can use it to estimate the relative accuracy of the orbital angular momentum L from different simulations. To better connect ℓ_{PT} to L, I will include $1PN$ post-Newtonian effects using (2.9b) from [62],

$$\mathbf{L_{PN}} = \mathbf{L_N} \left\{ \frac{1}{2} v^2 (1 - 3\eta) + (3 + \eta) \frac{m}{r} \right\} , \qquad (1.75)$$

where $\mathbf{L_N}$ is the Newtonian angular momentum, v is the velocity in the reduced system, r is the separation in the reduced system, and η is the symmetric mass ratio given by Eq. (1.49).

So the modified error measure, which I call the puncture track angular momentum to avoid confusion, is

$$\mathbf{L_{PT}} = \mathbf{L_N} + \mathbf{L_{PN}} . \qquad (1.76)$$

Note that I have not included spin-orbit corrections, since I originally this diagnostic for systems in which the total spin is 0. In the future the definition could be extended to include spin. I will discuss the behavior of L_{PT} in Chapter 4 and use it to evaluate the moving patches technique.

Chapter 2

A Simple Interpretation of Phase Evolution in Nonspinning

Black-Hole Binary Mergers

The material in this chapter was published in [11] in 2008, including work by me in collaboration with the NASA Goddard Numerical Relativity group. This chapter summarizes the results of the original article emphasizing the portions to which I contributed significantly.

2.1   Overview

Non-eccentric black hole binaries are described by seven parameters: the mass ratio of the black holes and the three spin components of each black hole. We expect vanishing eccentricity for astrophysical black hole binaries, due to radiation of angular momentum, by the time they reach the separations found in numerical relativity simulations. The first successful simulations of the inspiral, merger, and ringdown of a black hole binary [29, 90, 34, 14, 13] involved equal mass black holes with no spins. Several investigations of the accuracy of the equal-mass waveforms and applications of the waveforms to data analysis followed [12, 30, 16, 17, 25]. At this point the numerical relativity community began to investigate more general parameter sets for black hole mergers, beginning with unequal-mass binaries. The NASA Goddard Numerical Relativity group, to which I belong, made an early con-

tribution to the investigation of such mergers and the influence of mass ratio on the waveforms produced by them [11]. We showed an underlying simplicity that dominates the waveforms, despite the complexity and nonlinearity of the system of equations describing the binaries, and how the waveforms varied with mass ratio. That simplicity suggests a heuristic description of the binaries' dynamics, the *implicit rotating source* or *IRS* model.

We performed a series of simulations of nonspinning black hole binary mergers for four different values of the mass ratio q: 1, 2, 4, and 6. Each simulation is labeled with the mass ratio as $q : 1$. All of the simulations included at least four orbits of inspiral before merger. We find excellent agreement in the "rotational phase" as calculated from the strain-rate for different modes in the $q = 4$ case and across mass ratios. This agreement suggests an implicit rotating source that generates the waveforms. Einstein's equations are coupled, nonlinear, partial differential equations describing the metric, a four-dimensional continuous tensor field. However, for much of the inspiral, a nonspinning black hole binary approximates a rigid rotator of slowly decreasing length. The strong curvature of the physical fields is sufficiently localized that in the wave zone, the IRS approximation describes the dynamics well. Note that this is even simpler than the post-Newtonian description. Because of the equality of the $\ell = m$ quasinormal modes, the IRS heuristic description holds through merger and ringdown.

For a well-separated black hole binary with small velocities, post-Newtonian analysis approximates the binary as a pair of point particles. In this case the weak field is a perturbation to flat space, so in fact, in the post-Newtonian analysis, the

gravitational radiation in the wave zone behaves as if there were a simple two-particle system in the source region. Taking this a step further, from the gravitational radiation we can infer a single quasi-stationary rotating massive body which we call an *implicit rotating source* or *IRS*. The mass distribution generates gravitational radiation from the quadrupole and higher moments as they rotate. In the post-Newtonian analysis, the (ℓ, m) modes of the radiation have frequencies $m\Omega$ to at least 2PN order, where Ω is a common shared frequency that describes the rotation of the IRS as a whole. Ω is proportional to the symmetric mass ratio η (Eq. (1.49)). The modes correspond to the source moments, and so from the radiation, it appears that the moments rotate synchronously, and on orbital timescales, the IRS behaves as a rigid rotator. It is quasi-stationary because the separation of the binary decreases as it radiates energy, so the IRS is more accurately a rigid rotator of slowly decreasing length.

The IRS goes beyond an *ad hoc* description. It provides a uniform physical description from inspiral to ringdown. Building on the qualitative description of the source during inspiral, we introduce an analytic model for the phase evolution of the waveforms through merger, based on a continuous, monotonically increasing rotational frequency. We find that, with appropriate scaling, the peak rate of change of waveform frequency is the same across $\ell = m$ modes and across mass ratios. Based on this model, we establish a relationship between frequency and amplitude evolution. Finally we use this extended IRS model in an analytic waveform model based on the *pseudo-4PN* (p4PN) EOB model [32]. Though my focus will be on the IRS phase model, I include the results built on it because of their importance and

to establish the relevance of the phase description.

2.2 Simulations

The simulations were run with HAHNDOL, starting with puncture initial data. For the initial data, the constraints were solved using AMRMG, a multi-grid elliptical solver that produces second-order error [28]. Though this is a lower convergence order than the time integration and finite differencing, the magnitude of the error is much smaller for AMRMG than for the other two sources, so the convergence order of the simulation is determined by the evolution operations. The initial momenta for the punctures were determined according to Kidder [62], a 2PN estimate. The evolutions used the moving puncture approach. The conventional BSSNOK variables were evolved, except that $e^{-2\phi}$ was evolved in place of ϕ itself. This quantity goes to 0 at the punctures, which eliminates a source of singularities and therefore instabilities in the simulations. The simulations used the 1+log lapse and Gamma-freezing gauge conditions suggested by [104], the constraint damping terms suggested by [42], and the dissipation terms suggested by [64, 59]. The simulation grids were adaptively refined by PARAMESH, based on a criterion that the product of the square root of the Coulomb scalar [20, 33] and the grid spacing remain roughly constant. These simulations used fifth-order interpolation at the refinement boundaries on the grid. Spatial derivatives were evaluated with sixth-order finite differencing except for advection terms which used second-order lopsided upwinding.

Table 2.1 shows the initial parameters for the simulations. For convenience, I

will refer to the simulations by a nominal mass ratio, the target mass ratio in the initial data. In the simulations, the finest grid spacing varied from $M/32$ to $3M/224$, as shown in the table. For the 4:1 case, the simulations were run at three different resolutions to provide convergence testing, so I will refer to the 4:1 simulations by the resolutions, coarse ($3M/160$), medium ($M/64 = 3M/192$), and fine ($3M/224$). The separations were chosen to produce at least five orbits before merger. The simulations were scaled by the total mass, so for a given total mass scaling of a simulation, the effective resolution for the smaller black hole was equal to the mass ratio, and so larger mass ratios required higher resolutions. The outer boundaries were placed $1000M$ or more from the punctures to prevent outer boundary noise from propagating to the wave zone during the necessary durations of the simulations.

Table 2.2 lists initial masses and angular momentum for each simulation. In addition to M_{ADM}, the total energy of the initial data, the table includes two different calculations of the infinite-separation mass of the system, the total mass of the black holes if they were infinitely far apart. The first is the sum of the apparent horizon masses, which were calculated by AHFINDERDIRECT [103]. The second infinite-separation mass is calculated from M_{ADM} and the binding energy E_b of the binary according to

$$M_{rest} = M_{ADM} - E_b \,. \tag{2.1}$$

The binding energy, which is negative, is an effective-one-body [31] post-Newtonian calculation based on the initial angular momentum J_0. All the simulations use the first infinite-separation mass (M_{AH}) as the mass scale, except the 1:1 simulation,

mass ratio	h_f	$m_{1,p}$	$m_{2,p}$	r_0	P_0	η
1:1	$M/32$	0.4872	0.4872	10.800	0.09118	0.2500
2:1	$3M/160$	0.3202	0.6504	8.865	0.09330	0.2228
4:1	$3M/224$	0.1890	0.7900	8.470	0.06957	0.1601
	$M/64$	0.1890	0.7900	8.470	0.06957	0.1601
	$3M/160$	0.1890	0.7900	8.470	0.06957	0.1601
6:1	$M/64$	0.1338	0.8490	8.003	0.05559	0.1226

Table 2.1: Physical and numerical parameters of the initial data for all the runs presented. $m_{1,p}$ and $m_{2,p}$ are the puncture masses of the two pre-merger holes. r_0 and P_0 are the initial coordinate separation and (transverse) linear momentum, respectively. h_f is the spatial resolution of the highest refinement level for each run. Finally, η is the resulting symmetric mass ratio, as determined from the two holes' horizon masses.

mass ratio	h_f	J_0^2	M_{ADM}	M_{AH}	M_{rest}
1:1	$M/32$	0.9847	0.9907	—	1.0005
2:1	$3M/160$	0.8271	0.9889	0.9989	0.9990
4:1	$3M/224$	0.5893	0.9929	1.0003	1.0004
	$M/64$	0.5893	0.9929	1.0003	1.0004
	$3M/160$	0.5893	0.9930	1.0003	1.0005
6:1	$M/64$	0.4449	0.9942	1.0000	1.0001

Table 2.2: Masses and angular momenta. J_0 is the total initial orbital angular momentum. M_{ADM} is the total energy of the initial data. The total, infinite-separation mass M of the system is measured in two ways – M_{AH}, the sum of the initial (apparent) horizon masses of the two holes, and M_{rest}, the sum of the ADM energy and the binding energy from effective-one-body theory [31].

which was run before AHFINDERDIRECT was implemented in HAHNDOL and therefore uses the second infinite-separation mass (M_{rest}).

2.3 Rotational Phase

From the gravitational radiation, we can calculate the rotational phase of an implied source responsible for the radiation. Post-Newtonian analysis and gravitational wave observation conventionally describe the radiation in terms of the strain. As explained in Chapter 1, spin-weighted spherical harmonics provide an excellent basis in which to decompose the strain $h(t, \vec{r})$. We can isolate the azimuthal varia-

tion by writing the components as

$$Rh_{\ell m}(t) = \begin{cases} H_{\ell m} e^{im\Phi_{\ell m}^{(h)}(t)} & (m > 0) \\ H_{\ell m} e^{im\Phi_{\ell m}^{(h)}(t)} (-1)^\ell & (m < 0) \end{cases} . \tag{2.2}$$

The form of Eq. (2.2) reflects the circular polarization of the radiation. The radial dependence is approximately removed by the leading R which represents the coordinate radius at which the radiation is observed. The oscillatory time variation is contained in the *rotational phase* function $\Phi_{\ell m}^h(t)$, which is a rescaling of the traditional phase description of h in anticipation of the implied source. The h superscript indicates that the phase originates from the strain. For each (ℓ, m) mode, we consider an implied source multipole that generates the radiation mode as that source rotates. The source rotational phase $\Phi_{\ell m}^h(t)$ produces the radiation phase $m\Phi_{\ell m}^h(t)$ due to the m-fold symmetry of the multipole. In the post-Newtonian analysis is valid, $\Phi_{\ell m}^h(t)$ agrees with the orbital phase of the system up to 2PN order, so during the inspiral where the post-Newtonian approximation is valid, the rotational phase should coincide with the orbital phase for any (ℓ, m) mode. The remaining time variation is represented by the $H_{\ell m}(t)$, a slowly varying function of time during the inspiral. $H_{\ell m}$ represents the magnitude of the implied (ℓ, m) source multipole. The additional factor of $(-1)^\ell$ for $m < 0$ maintains the connection between the rotational phase and the phase of the (ℓ, m) waveform component.

Now consider the strain-rate defined as $\dot{h} \equiv dh/dt$. Because HAHNDOL calculates ψ_4 directly from the evolved fields, the strain-rate allows computation of the rotational phase from simulation results more directly. Calculating h from ψ_4

requires two time integrations, and for both of these an integration constant must be set, which introduces additional assumptions. As in the case of the strain, the strain-rate can be decomposed as

$$
R\dot{h}_{\ell m}(t) = \begin{cases} i\,A_{\ell m}e^{im\Phi_{\ell m}(t)} & (m > 0) \\[2ex] -i\,A_{\ell m}e^{im\Phi_{\ell m}(t)}\,(-1)^\ell & (m < 0) \end{cases}.
\tag{2.3}
$$

The decomposition is analogous to Eq. (2.2), except that i is explicitly factored out and Φ is not superscripted. The initial i corresponds to either an additional factoring of the secular amplitude factor or a phase shift in the complex exponential. In what follows, $\Phi_{\ell m}$ is always the rotational phase calculated from the strain-rate, so there is no need for a superscript. A direct differentiation of Eq. (2.2) produces a similar expression and shows why Eq. (2.3) factored out the i. It was anticipating the time differentiation. Comparing the two

$$
A_{\ell m}e^{im\Phi_{\ell m}(t)} = \left[|m|\dot{\Phi}_{\ell m}^{(h)}(t) - i\frac{\dot{H}_{\ell m}}{H_{\ell m}} \right] H_{\ell m}e^{im\Phi_{\ell m}^{(h)}(t)}
\tag{2.4}
$$

for $m > 0$. The $m < 0$ case is identical except for a complex conjugation of the bracketed expression. Because the secular change in $H_{\ell m}$ is so much smaller than the variation from rotation, we can ignore the second term in the brackets. $A_{\ell m} \sim |m|H_{\ell m}\dot{\Phi}_{\ell m}^{(h)}$ holds up to 5PN order, and $\Phi_{\ell m} = \Phi_{\ell m}^{(h)}$ up to 2.5PN order. So we can determine the IRS rotational phase from strain-rate rather than the strain, up to 2.5PN order.

If we decompose the strain-rate in a more conventional manner,

$$
R\dot{h}_{\ell m}(t) = V_{\ell m}(t)e^{i\varphi_{\ell m}(t)},
\tag{2.5}
$$

64

set it equal to Eq. (2.3), and solve for $\Phi_{\ell m}$ we find

$$\Phi_{\ell m}(t) = \begin{cases} \frac{1}{m}\left(\varphi_{\ell m}(t) - \frac{\pi}{2} + 2\pi n_{\ell m}\right) & (m > 0) \\ \frac{1}{m}\left(\varphi_{\ell m}(t) + \frac{\pi}{2} + 2\pi n_{\ell m} + \ell\pi\right) & (m < 0) \end{cases}, \qquad (2.6)$$

where n is an integer. The $\frac{\pi}{2}$ difference between the strain and the strain-rate produces the $\frac{\pi}{2}$ term, and the $(-1)^\ell$ factor in the strain decomposition Eq. (2.2) produces the $\ell\pi$ factor.

The origin of the $2\pi n_{\ell m}$ term is less obvious. It may seem like a trivial phase shift to the strain-rate, but identifying it was the crucial insight that allowed me to compare the rotational phases. When the Bowen-York initial data has propagated past the wave zone and the strain-rate amplitude is large enough to accurately compute the phase, we start computing the accumulated phase in the strain-rate modes. The initial phase has a 2π ambiguity, so adding a multiple of 2π would seem to have no impact, but if it is not an integer multiple of $2\pi m$, it results in a nontrivial change in the rotational phase, a different orientation of the system, when the phase is divided by m, as it is in Eq. (2.6). Therefore, to compare the accumulated rotational phases computed from the different strain-rate modes, we choose $n_{\ell m}$ so that the modes are consistent in the orientation.

There are m distinct values of the rotational phases that produce the same value for the $\dot{h}_{\ell m}$ phase, even after eliminating trivial shifts. To eliminate the m-fold phase degeneracy of the (ℓ, m) mode, we compare it to the $(\ell'm')$ mode, where m and m' are relatively prime, and choose the smallest $n_{\ell m}$ that minimizes $|\Phi_{\ell m} - \Phi_{(\ell'm')}|$, the difference in the rotational phases determined from the two modes. Since the

amplitude of the $(2, 2)$ mode is larger than all other $\ell > 0$ modes, we expect it to have the smallest relative error. So we compare it to another large amplitude mode that is relatively prime in m and determine n_{22} as explained above. Then we compare all other modes to $(2, 2)$ to determine the remaining $n_{\ell m}$ values. Choosing the $n_{\ell m}$ to minimize the rotational phase differences may seem arbitrary, but in fact after the degeneracy is eliminated, it is no more arbitrary than omitting the full-cycle phase shift terms.

2.4   Similarity of Phase Across Modes

For a simulation, agreement in the rotational phases calculated from different modes suggests that the implicit multipoles rotate coherently, which is what we would expect if the multipole moments described a rigid rotator. Conversely, if the radiation were due to a rigid rotator source, we would expect that the phase of a particular (ℓ, m) mode would be $m\Phi$, where Φ is the rotational phase of the rotator.

In the 4:1 case, the rotational phase $\Phi_{\ell m}$ computed from different $\ell = m$ modes agrees to within $\pi/60$ as seen in the left panel of Fig. 2.1 which shows the rotational phase from several different modes extracted at $R = 45M$. The Φ_{lm} are aligned using the $n_{\ell m}$ procedure described above. The time axis has been shifted so that the peak of $\dot{E}$ occurs at $t = 0$. The plot also includes Φ_{21} and Φ_{32}, as well as the phase of the punctures' separation vector. The puncture phase tracks the $\ell = m$ rotational phases' evolutions well, while Φ_{21} and Φ_{32} both have noticeable offsets and different slopes than the $\ell = m$ phases. The right panel of Fig. 2.1 shows

the differences of the phases in the left panel from Φ_{22}, and here the deviations of the $(2,1)$ and $(3,2)$ modes are more clear, as is a phase shift of ~ 0.2 in the puncture phase. The $(2,1)$ strain-rate has the longest wavelength of the modes shown, and since the wavelength of the mode sets the scale for finite R effects, it is not surprising that the $(2,1)$ mode shows the impact of the finite extraction radius more than the other modes since its effective extraction radius is smaller by a factor of at least 2 compared to the other modes. To improve the phase calculation for the $\ell \neq m$ modes, the right panel includes Richardson extrapolations of these modes (indicated by *) which I performed using the waveforms extracted from the $R = 45M$ and $R = 90M$ extraction radii. (see Eq. (1.68)) The extrapolated phases are consistent with the other phases until approximately $50M$ before $t = 0$.

2.5 Phase Comparison Across Mass Ratios

Since we have seen the consistency of the rotational phase computed from different modes in the 4:1 case, I will consider only the $(2,2)$ component in comparing the rotational phase across mass ratios. Post-Newtonian analysis suggests that the natural scale for time in such a comparison is the *chirp mass* $\mathcal{M} \equiv M\eta^{3/5}$, rather than the total mass M. That is, to the leading order post-Newtonian term, the rotational phases of the different mass ratios should be the same when time is divided by the chirp mass. Using this time scaling, Fig. 2.2 shows that comparison. The rotational phase is again computed using Eq. (2.6). In addition to scaling time by the chirp mass, the time and phase have been shifted such that at $t = 0$, the chirp
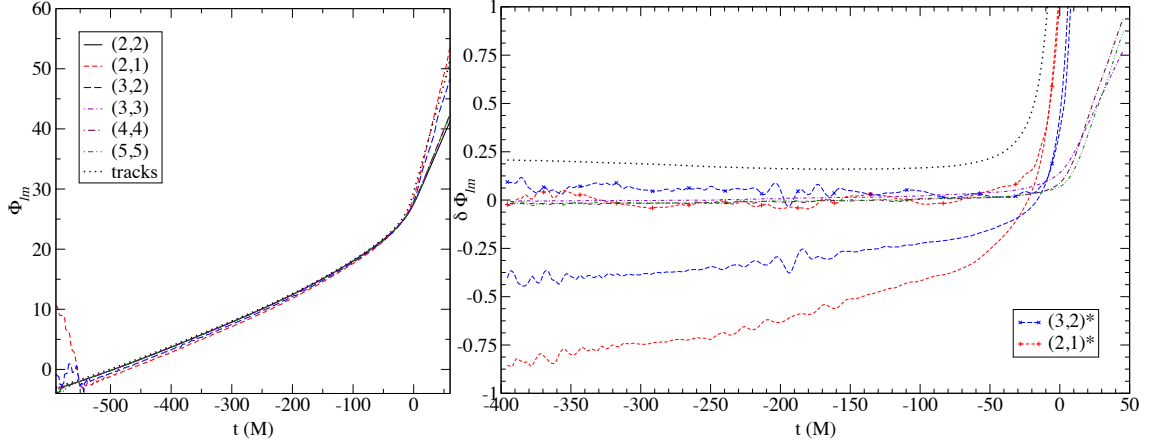
Figure 2.1: Rotational phase $\Phi_{\ell m}$ calculated from puncture tracks and strain-rate modes extracted at $45\ M$ for the 4:1 mass ratio case. The rotational phase is computed from each mode using Eq. (2.6) and from the angle of the vector connecting the two punctures, which remain in the x-y plane. The phase curves are aligned in time so that the peak of $\dot{E}$ occurs at $t = 0$ for all curves. The left panel compares the different calculations of $\Phi_{\ell m}$, while the right panel compares the differences between the phases and the rotational phase from the (2,2) mode. In the right panel, we add Richardson extrapolations of the (2,1) and (3,2) modes based on the $45\ M$ and $90\ M$ extractions of these modes and assuming an R^{-2} error. The Richardson extrapolations are distinguished by asterisks in the figure. The differences are smoothed to better show the trends.
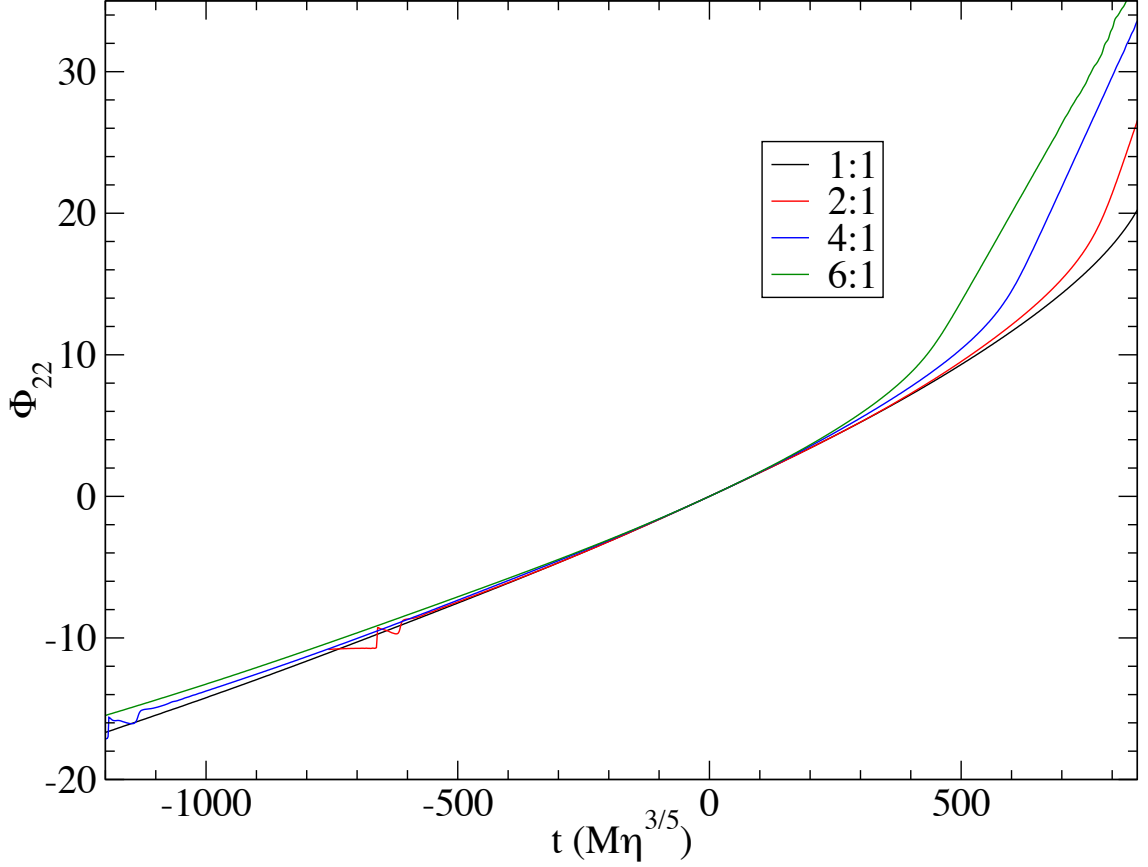
Figure 2.2: Rotational phase Φ_{22} for different mass ratios derived from $\dot{h}_{22}$. The curves are time shifted so that the chirp frequencies $\Omega_{22}\mathcal{M} = \dot{\Phi}_{22}\mathcal{M}$ are equal at $t = 0$.

frequency ω_{chirp} is 0.033 and the rotational phase Φ is 0. The chirp frequency is the rotational frequency multiplied by the chirp mass. (This is the appropriate rescaling of frequency since frequency has dimensions of inverse time. If time is divided by chirp mass, frequency should be multiplied by chirp mass to maintain consistency.) As expected the rotational phases from the different mass ratios are close for several hundred $\mathcal{M}$ approaching merger. The agreement continues for several hundred $\mathcal{M}$ before and after $t = 0$, so the leading-order post-Newtonian approximation provides a rough approximation even through the merger.

2.6 *IRS* Interpretation of Phase Evolution

We have looked at simulations of black hole binaries for mass ratios between 1:1 and 6:1. In particular we have focused on the phase evolution leading up to merger. We find that for a single mass ratio (4:1), different (ℓ, m) modes of the strain-rate show very close agreement up to merger, when the modes' phases are scaled by their m values, the azimuthal index of the mode. This agreement suggests an underlying rotational phase in the system, as if the radiation originates from a rigid rotator (IRS) that decreases in length on orbital time scales. The fields appear to be strongly tied to the black hole point sources, and apparent structural integrity of the IRS continues through merger. In this heuristic description, the synchronous source multipoles describing the mass distribution of the rigid rotator generate the strain-rate modes as the IRS rotates according to the multipoles' azimuthal symmetries. The dominant $(2, 2)$ mode oscillates at twice the rotational frequency because of a π-symmetric mass distribution rotating at the rotational frequency, the $(3, 3)$ mode that appears for unequal-mass binaries oscillates at three times the rotational frequency because of a $2\pi/3$-symmetric mass distribution rotating at the rotational frequency, and in general, the (ℓ, m) mode oscillates at m times the rotational frequency because a $2\pi/m$-symmetric mass distribution rotating at the rotational frequency. A similar phase agreement occurs within the modes of the other mass ratios. Also, the phases of the $(2, 2)$ modes of each of the mass ratios agree closely when time is scaled by the chirp mass as suggested by post-Newtonian analysis, which predicts that the phase scales as the symmetric mass ratio to the

$\frac{5}{3}$ power for early times. This phase agreement continues up through merger, well past the regime for post-Newtonian approximations.

In the original article [11], we extend the IRS model through merger with a quantitative model of frequency evolution in the transition from inspiral to ringdown, based on the source model explained in this chapter. The source rotational frequency is given by

$$g(t) = \Omega_i + (\Omega_f - \Omega_i) \left(\frac{1 + \tanh[\ln \sqrt{\kappa} + (t - t_0)/b]}{2} \right)^\kappa. \qquad (2.7)$$

The form of $g(t)$ provides exponential growth from an initial frequency Ω_i, followed by a transition to an exponential decay to the final frequency Ω_f, which we identify with the fundamental quasinormal-ringing frequency by $\Omega_f \equiv \omega_{QNM}/m$, with m the azimuthal index of a (ℓ, m) mode. This is possible because of the approximate equality of the $\ell = m$ fundamental quasi-normal-ringing frequencies when scaled in this manner. Then the re-parameterization

$$\Omega_i \equiv \Omega_f - \tfrac{1}{2} b \dot{\Omega}_0 \left(1 + \frac{1}{\kappa} \right)^{1+\kappa}, \qquad (2.8)$$

connects the model explicitly with the "chirping" leading up to the peak of radiation which marks the merger. With this substitution, the chirp rate $\dot{g}$, the rate of change of the frequency, peaks at $\dot{\Omega}_0$ at time t_0. Ω_f, $\dot{\Omega}_0$, and t_0, along with κ and b, which characterize the time scales for the exponential growth and decay, constitute the model parameters.

Fitting $g(t)$ to the data for strain-rate mode data via the parameters yields agreement between the model and the data to within a few percent after $t = -20$, for the $(2, 2)$ mode of each mass ratio. The agreement is similar for the significant

71

(ℓ, m) modes of the 4:1 case, excluding the $(3,2)$ mode. While this mode may be subject to an incorrect measurement procedure,[1] and therefore, is not a good reflection of the model accuracy, we do not claim that the model is valid for $\ell \neq m$ modes. Also we find that $\dot{\Omega}_0 M_{\mathrm{f}}/\Omega_{\mathrm{f}} \sim 0.02$ within the fit uncertainties, for all $\ell = m$ modes. The peak chirp rate is consistent across these modes when scaled by the final frequency and mass, which provide natural scaling magnitudes since the chirp rate, the instantaneous time rate of change of the frequency, reflects both the frequency the mode is approaching and the time scale for observing the changes.

This frequency model, based on the IRS description, can be used through merger and into the ringdown phase. Combined with an approximate relationship between the amplitudes and frequencies of the waveform modes, which was established in the original article [11], the frequency model provides a variation of the p4PN EOB model found in [32]. Rather than match the inspiral p4PN waveform to a superposition of the fundamental ringdown mode and a few overtones weighted for continuity with the inspiral waveform, we can match the p4PN waveform to the waveform derived from the IRS frequency model. As discussed in the original article, this p4PN variation matches the numerical results from the nonspinning simulations well, with an accuracy similar to that of the original p4PN EOB.

---

[1]The $(3,2)$ mode is subject to coupling with the much stronger $(2,2)$ mode, and therefore it is sensitive to the surface used to measure the radiation. We use coordinate spheres for the measurements, which are distorted from the areal-radius spheres required for greater accuracy. As a result, the measured $(3,2)$ mode may mix the fundamental $(2,2)$ and $(3,2)$ quasinormal-ringing frequencies.

The IRS heuristic model provides a simple but accurate description of the dynamics of nonspinning black hole mergers that offers both direct predictions of the waveforms from such mergers and comparisons for subsequent simulations. In particular, the IRS-EOB model has been used in studies of parameter estimation by LISA [76] and the detectability and mass ratio degeneracy for observations of moderate mass ratio black-hole binaries [76]. Also, later investigations of unequal-mass mergers have confirmed our results, such as the importance of $\ell > 2$ radiation modes for gravitational wave data analysis [46]. Finally, the generality of the IRS description could be evaluated in simulations involving spin. For non-precessing systems, where the spins are aligned or antialigned with the orbital angular momentum vector, the dynamics are similar to the nonspinning case, as shown in [36], and we may reasonably anticipate a similar interpretation of the radiation.

Chapter 3

Kicks Due to Mass Ratio and Spin in Head-On Black-Hole Binary

Mergers

The material in this chapter was published in [39] in 2007, including work
by me in collaboration with the NASA Goddard Numerical Relativity group. Dale
Choi initiated the research, but Bernard Kelly and I were responsible for most of the
simulations, as well as the analysis, with our coauthors. This chapter summarizes
the results of the original article, emphasizing the portions to which I contributed
significantly. These results were submitted for publication and made public on
`arxiv.org`, following successful simulations of merger kicks due to the effects of
unequal masses and spin, separately, but before inspiral simulations demonstrating
kicks due to the combined effect of unequal masses and spin. Therefore, they served
as an important transition from the original full-GR studies of kicks to the later,
more comprehensive kick studies.

3.1   Introduction

The gravitational radiation emitted by the inspirals of black hole binaries
carries energy, angular momentum, and linear momentum. This radiation is emitted
asymmetrically when the masses are not equal and when the black holes have spins.
When the radiation emission is asymmetrical, the total linear momentum of the

radiation is not zero, and so conservation of linear momentum requires that the binary must have an equal and opposite total linear momentum. During the inspiral phase, the direction of the radiation asymmetry changes as the black holes orbit each other, and over the course of a full orbit of the binary, the net momentum flux in the radiation is near zero. At merger the radiation bursts in a time that is very small compared to the orbital period, and there is a resulting net momentum flux, and so the black hole formed in the merger has a net linear momentum, a recoil "kick". Such recoils can significantly alter the subsequent development of astrophysical black holes, such as supermassive black holes in merged galaxies [77, 24, 51, 75, 110, 107, 66, 78] and intermediate-mass black holes in dense stellar clusters [81, 80, 83, 79, 49, 50, 87].

Since the first successful simulations of a black hole binary inspiral, merger, and ringdown in 2005 [90, 34, 14], there have been numerous studies of the kicks produced in such simulations due to the effects of unequal masses and spins [57, 15, 47, 56, 63, 36, 35, 37, 9, 38, 45, 10, 71, 70, 105, 73]. In this chapter, I consider kicks from head-on mergers because these systems have simpler dynamics than inspiralling systems, and head-on mergers approximate the final plunge of a binary inspiral. The elaborate dynamics of such inspirals complicate establishing the direction of the final kick and its dependence on mass ratio and spin. For nonspinning head-on mergers all motion occurs along the coordinate line defined by the initial positions (*longitudinal*). Spin adds slight velocities perpendicular to this line (*transverse*) as the black holes approach, but the dominant motion still occurs in initial line. In either case, this coordinate line provides a meaningful reference for kick directions.

Finally they provide clean tests of the leading-order post-Newtonian (PN) prediction of orthogonality of spin and mass-ratio kicks and the scaling of spin kicks [62].

In this study of head-on mergers, we investigate the effects of mass ratio and spin, individually and combined, on the resulting kicks. We use a mild mass ratio of $q = 2/3$ and moderate spins (< 0.76) orthogonal to the black holes' motions. In the absence of spin, we observe modest longitudinal kicks for unequal-mass mergers, with velocities of $\sim (2 - 5)$km/s. In mergers of equal-mass holes with spin, the observed kicks are transverse, and they roughly scale with the sum of the Kerr spin parameters $a_1 + a_2$, with transverse velocities $\sim 15 - 30$km/s. We find both types of kick in mergers of unequal-mass black holes, with the direction of each kick type independent of the other. In these cases, total velocities are still $\sim 15 - 30$km/s since the spin kicks are significantly larger than the mass-ratio kicks.

3.2   Initial Data

All of the simulations use puncture initial data [26]. In each case the punctures start at rest. In the nonspinning cases, the constraints are solved exactly by Brill-Lindquist initial data, while the cases that include spin are solved by Bowen-York initial data. In both cases, `AMRMG`, a second-order-convergent multi-grid elliptic solver [28], solves for the initial data on the grid. In all cases

$$a_A \equiv \frac{|S_A|}{m_A},$$ (3.1)

the Kerr spin parameter.[1]

The initial configurations use either equal (EQ) and unequal (NE) black-hole masses with various spins on the black holes. Table 3.1 lists the initial parameters of the seven different cases are, while Table 3.2 provides initial derived quantities. Lengths are scaled by the fiducial mass M (computational). The punctures are located on the y-axis initially, with the center of mass at the coordinate origin. The "proper separation" l of the holes measures the proper distance along the y-axis of the closest parts of the apparent horizons. This is a convenient measure of the black holes, but it is not necessarily the smallest physical distance between the holes in the spatial slice, since the space-like geodesics may not lie on the coordinate axis due to spin effects. We scale all distances such that the horizon mass of the lighter hole m_1 remains constant relative to the grid spacing. This provides a common "effective resolution" across the simulations by maintaining a common level of resolution of the smaller hole. With this restriction, we have attempted to maintain the same proper separation l for the holes as well.

3.3 Methodology

Here we discuss two important parts of the methodology: calculating the thrusts and kicks, as well as our method to reduce the error introduced by the Bowen-York pulse. The original article includes full details of the methodology as well as a discussion of convergence.

---

[1]Sometimes a is defined as the dimensionless Kerr spin parameter, $|S|/M^2$ or S/M^2. Here, we use Eq. (3.1).

Run	$m_{1,p}/M$	$m_{2,p}/M$	y_1/M	y_2/M	S_1^z/M^2	S_2^z/M^2
EQ_{00}	0.5000	0.5000	4.0	-4.0	0.0	0.0
NE_{00}	0.4909	0.7478	4.8348	-3.2232	0.0	0.0
EQ_{+0}	0.3444	0.5000	4.0	-4.0	0.2	0.0
EQ_{+-}	0.3444	0.3444	4.0	-4.0	0.2	-0.2
NE_{+-}	0.3436	0.7140	4.8	-3.2	0.2	-0.2
NEa_{+-}	0.3436	0.7140	7.2	-4.8	0.2	-0.2
NEb_{+-}	0.3436	0.5496	4.8	-3.2	0.2	-0.4486

Table 3.1: Directly specified parameters of the numerical simulations scaled by the fiducial mass M. S_A^z is the spin (Bowen-York angular momentum) of hole A. (All spins lie on the z-axis.) NEa_{+-} (increased separation) and NEb_{+-} (equal dimensionless spin parameter) are variants of NE_{+-}. See the text for further discussion of these simulations.

Run	m_1	m_2	q	η	a_1	a_2	a_1/m_1	a_2/m_2	l	M_{ADM}
EQ_{00}	0.514	0.514	1.0	0.250	0.0	0.0	0.0	0.0	12.24	1.0
NE_{00}	0.514	0.771	0.667	0.240	0.0	0.0	0.0	0.0	12.24	1.24
EQ_{+0}	0.516	0.514	1.004	0.250	0.389	0.0	0.758	0.0	12.24	1.0
EQ_{+-}	0.514	0.514	1.0	0.250	0.389	0.389	0.758	0.758	12.4 ± 0.2	1.0
NE_{+-}	0.516	0.773	0.668	0.240	0.388	0.259	0.752	0.335	12.6 ± 0.2	1.24
NEa_{+-}	0.513	0.764	0.672	0.240	0.390	0.262	0.759	0.342	17.0 ± 0.2	1.25
NEb_{+-}	0.516	0.784	0.658	0.239	0.388	0.572	0.752	0.730	13.0 ± 0.2	1.26

Table 3.2: Derived initial quantities scaled by the fiducial mass M. m_1 is the horizon mass of the lighter hole. q is the mass ratio defined by horizon masses. $a_A = |S_A|/m_A$ is the approximate Kerr spin parameter of hole A. l is the "proper separation" of the holes. (See the text for a discussion of this quantity.) M_{ADM} is calculated at a finite coordinate distance from the origin ($60M$ for all but run EQ_{+0}), using Eq. (12) of [86]. Errors in horizon masses m_1 & m_2 are $\sim 1.5\%$, which propagate into derived quantities q, η, a_1, a_2, a_1/m_1, and a_2/m_2.

3.3.1 Calculating Thrusts and Kicks

Because the recoil results from the emission of gravitational radiation, we calculate the radiation momentum flux, or *thrust*, which is given by Eq. (1.64). Then we integrate the thrust to calculate the total momentum radiated by the system up to time t,

$$\Delta P^i = \int_{t_0}^{t} dt' \frac{dP^i}{dt'} \,.$$

(3.2)

Note that the time integration should begin at $-\infty$ to include all previous thrust contributions, but we must start at a finite time t_0. The thrust calculation also includes a time integration which should begin at $-\infty$, but which we must start at t_0.

3.3.2 Bowen-York Radiation Pulses

Bowen-York black hole initial data includes extraneous radiation, the "Bowen-York pulse". This radiation is extraneous because it is not a result of the intended astrophysical scenario, but also it is not error from solving the Hamiltonian constraint. It results from the nonphysical approximations, such as conformal flatness, made in generating the initial data (*cf.* [52]). The pulse is concentrated near the apparent horizons, and much of the extraneous radiation falls into the black holes, producing an increase in the apparent-horizon masses of $\sim 1\%$. The radiation from the pulse that does escape the black holes mixes with the physical radiation, and so the pulse contributes to the error in the recoil kicks calculated from the radiation. The pulse occurs immediately in the simulations, and it has a short duration, so it

should be distinct from the astrophysical radiation.

Fig. 3.1 illustrates the Bowen-York pulse in the dominant $(l = 2, m = 2)$ mode of ψ_4 (extracted at $R_{\text{ext}} = 30M$), for the equal-mass configurations: zero-spin (EQ_{00}), single-spin (EQ_{+0}) and double-spin (EQ_{+-}). The three cases agree well in the later "astrophysical" part of the waveform. The greatest differences occur early, during the Bowen-York pulse. The pulse does not appear to influence the later waveform.

By starting the time integrations in the thrust and kick calculations from a time t_0 after the Bowen-York pulse, we can minimize the error introduced by the pulse. We estimate the duration of the pulse as $30M$ for the equal-mass (EQ) simulations and $45M$ for the unequal-mass (NE) data. These estimates result from treating each hole as a perturbed Kerr hole of approximately the derived mass and spin from the initial data, and assuming that each relaxes to a Kerr black hole by emission of quasinormal modes (QNMs). The most slowly damped mode will have an "e-folding" time $\tau_e \approx 12m$ for a hole of mass m [21], and it will have decayed by two orders of magnitude after $\sim 55m$. This serves as an upper limit on the duration of the pulse.

3.4 Results

We study the effect of unequal masses and spin separately and then the combined effect. Nonequal masses should produce a longitudinal kick, while adding
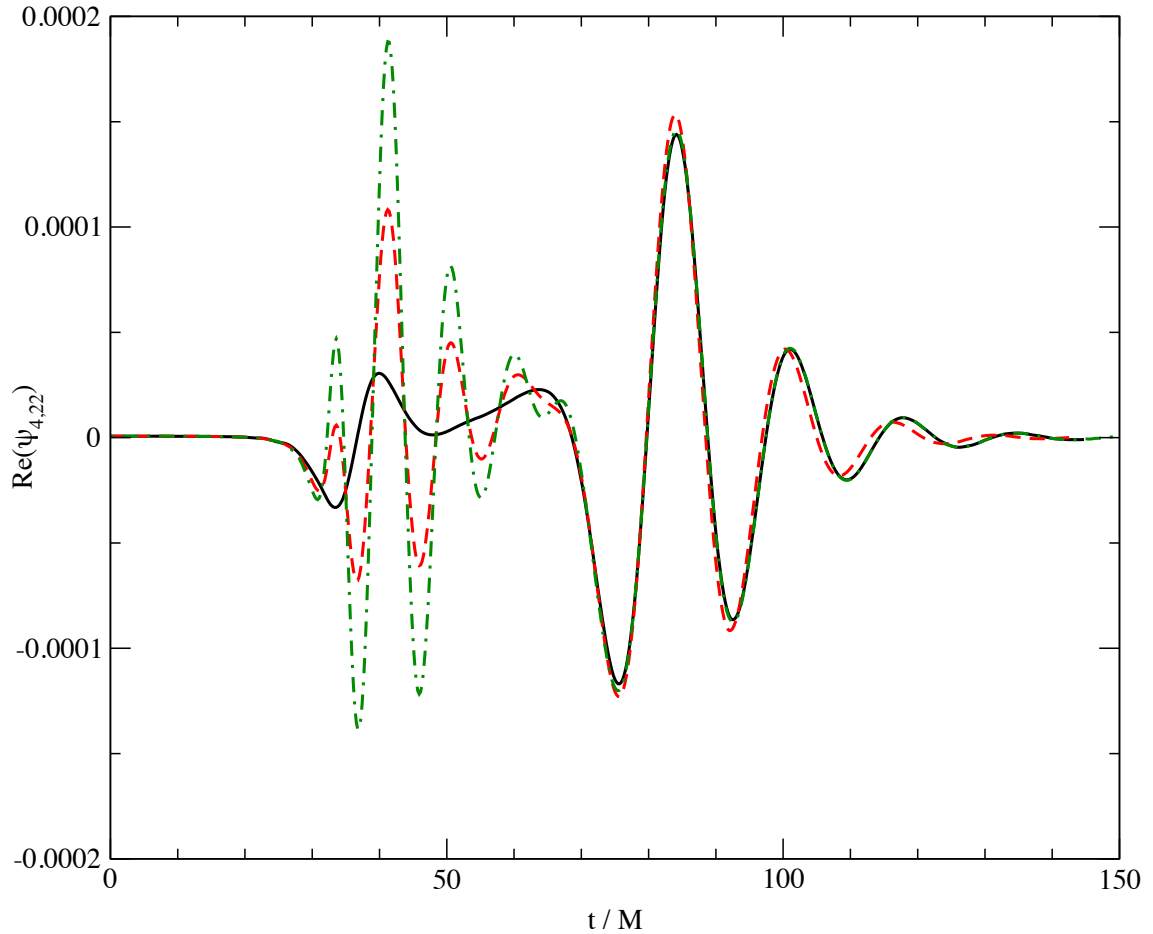
Figure 3.1: The dominant mode $(l = 2, m = 2)$ ψ_4 extracted at coordinate distance $R_{\text{ext}} = 30\,M$ for equal-mass configurations: EQ_{00} (solid/black), EQ_{+0} (dashed/red), and EQ_{+-} (dot-dashed/green).

spin to equal-mass black holes should produce a transverse kick. (To produce equal masses in the case where only one black hole has spin (EQ_{+0}), we increased the puncture mass of the second black hole.) The kicks for each simulation are listed in Table 3.3. Because the punctures are located on the y-axis initially, ΔP^y is the longitudinal kick, and ΔP^x is the transverse kick. The table includes the integration start time t_0 in each case. We estimate the kick errors by varying t_0 by $\pm 5M$.

3.4.1   Kicks Due to Unequal Masses

For unequal-mass binaries without spin, the kicks are longitudinal and modest, producing velocities of ~ 2.7 km/s in the case of a mass ratio $q = 1.5$. Fig. 3.2 shows both the thrust dP^y/dt (top panel) and the kick ΔP^y for this case. While the thrust varies in sign with the variation in radiation asymmetry, the resulting kick of 1.14 ± 0.08 in the y direction shows the measurable net momentum flux from the asymmetry. Because of the symmetries of the system, there are no thrusts in the x or z directions. In dimensionful units, the kick velocity is ~ 2.7 km/s. Fig. 3.2 includes the thrusts and kicks for four different extraction radii, and the final kicks across radii match closely. The difference for the $30M$ and $40M$ extraction radii is $\sim 0.5\%$. The agreement between the $50M$ and $60M$ extraction radii is lower than between the $40M$ and $50M$ radii. This appears to be because the $60M$ extraction radius lies in a lower resolution portion of the grid. The final kicks across extraction radii for the nonspinning mergers are listed as NE_{00} in Table 3.3.
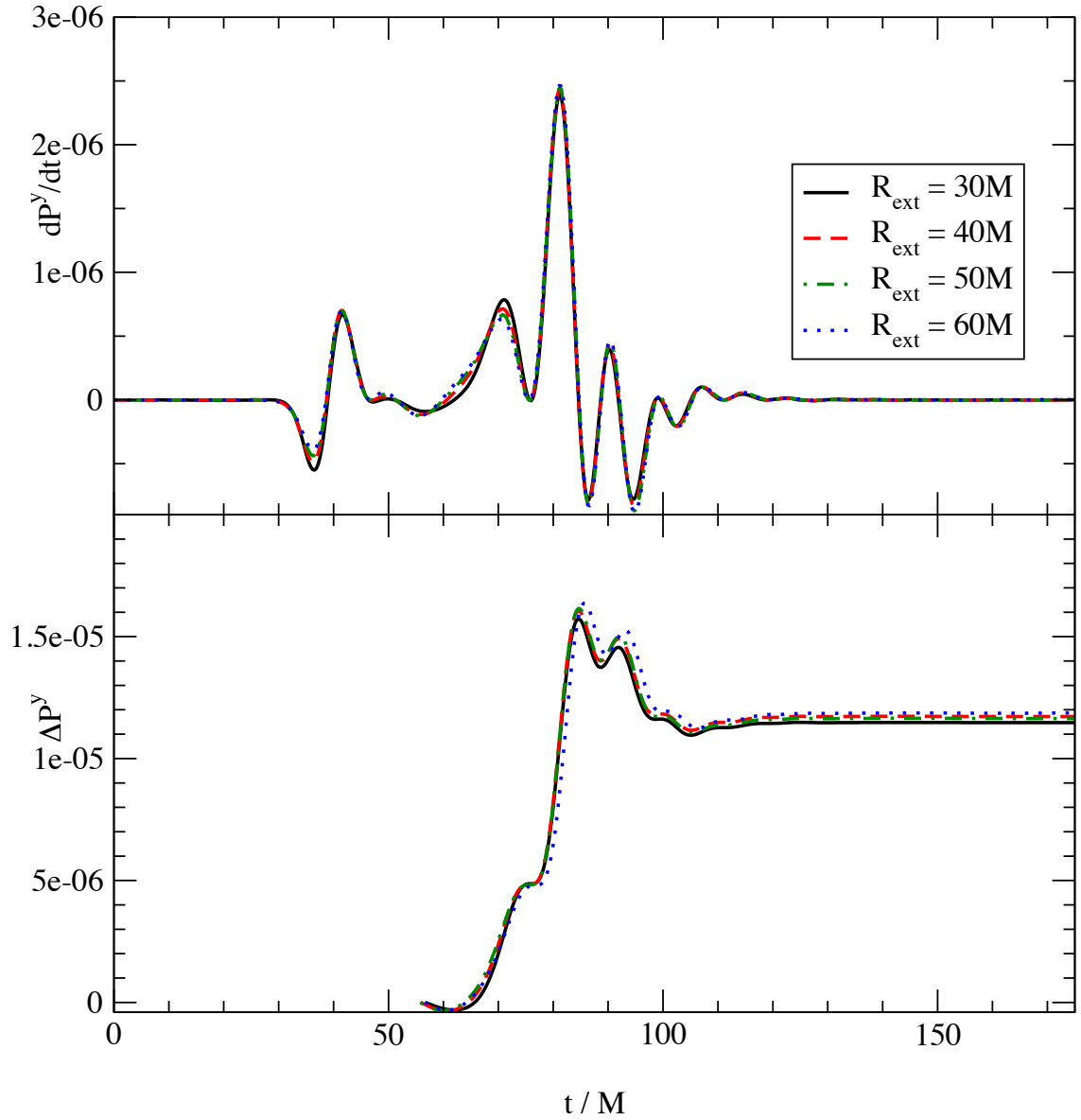
Figure 3.2: Longitudinal thrust dP^y/dt (top) and kick ΔP^y (bottom) for NE_{00} extracted at four radii R_{ext}: $30M$, $40M(10.6M)$, $50M(21.0M)$ and $60M(31.4M)$. The data are aligned so that the highest thrust peaks coincide, consistent with time-shift formula Eq. (14) in [43]. The necessary time shifts are listed in parentheses.

3.4.2 Kicks Due to Spin with Equal Masses

For the equal mass cases where one or both black holes possess spin, with the spins perpendicular to the initial infall direction, the kicks are transverse to both the initial infall direction and the spins. With one spin (EQ_{+0}), the velocities ~ 15 km/s, while with two anti-aligned spins (EQ_{+0}), the velocities ~ 30 km/s. Fig. 3.3 includes the thrusts and resulting kicks for the two cases. Comparing them, the EQ_{+-} kick is roughly twice EQ_{+0}, just as $a_1 + a_2$ is twice as large for EQ_{+-} as for EQ_{+0}. For EQ_{+0} the only interaction is "spin-orbit" since there is no other spin. While "spin-spin" effects are possible in the case of two spins, the results suggest that these effects are much smaller than the "spin-orbit" effects since the kicks scale with the sum, rather than the product, of a_1 and a_2. This is consistent with PN predictions of the interactions in which the "spin-spin" effects are higher PN order than the "spin-orbit" effects [62]. Again, the final kicks for these cases are listed in Table 3.3.

3.4.3 Kicks Due to Unequal Masses and Spin

When the black holes have unequal masses and spins, the kick magnitudes are $\sim 20 - 30$ km/s, with the longitudinal kick magnitude 0.15 the magnitude of the transverse kick. Again, the spins are perpendicular to the infall/kick plain, anti-aligned, and magnitude $\sim 0.2 - 0.4M$. In this case, the Bowen-York pulse and the physical signal overlap in the transverse thrust as seen in Fig. 3.4, which includes
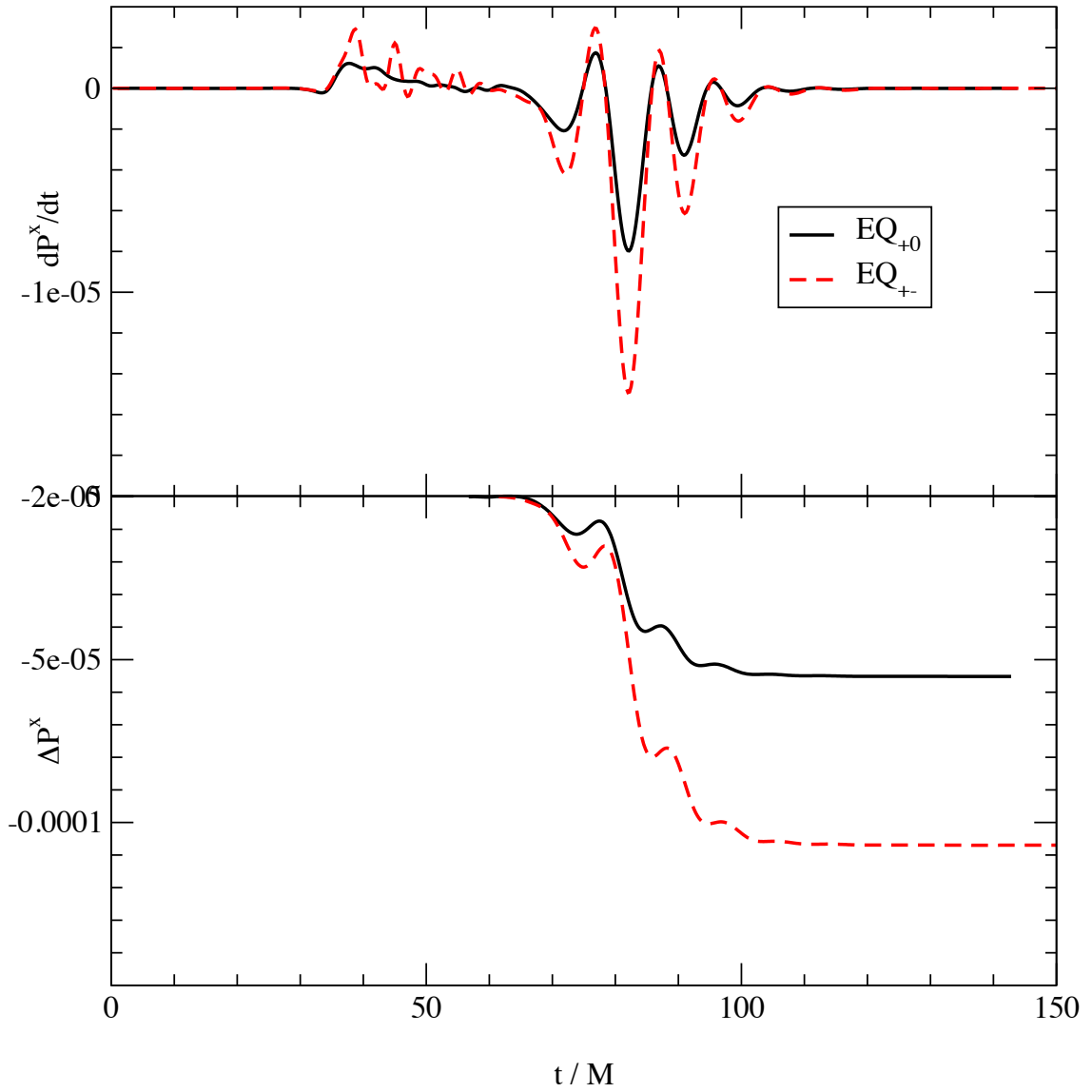
Figure 3.3: Transverse thrust dP^x/dt and kick ΔP^x for the spinning, equal-mass cases EQ$_{+0}$ (black/solid) and EQ$_{+-}$ (red/dashed). The extraction radius R_{ext} is $30M$ for both.

one case with the same initial separation d used in previous cases ($NE_{+-}, d = 8$) and one case with a larger initial separation ($EQ_{+0}, d = 12$). With the increased separation, the thrust due to the Bowen-York pulse is more distinct, while the physical thrust is very similar to the NE_{+-} physical thrust, delayed by $22M$. For the more separated case, the physical thrust begins at approximately $t = 114M$ at extraction radius $60M$. From this we infer that the NE_{+-} physical thrust begins at $t = 92M$ which we choose as the integration start time for the NE_{+-} kick, $t_0 = 92M$. In Figs. 3.5 and 3.6 we present the momentum kicks in the transverse (x) and longitudinal (y) directions for the NE_{+-} data set, comparing them with the kicks seen from the NE_{00} and EQ_{+-} cases before. Note that while the spin angular momentum S on each black hole is the same for EQ_{+-} and NE_{+-}, the transverse kick is actually larger for EQ_{+-}, so the transverse kick is not a simple function of $|S_1| + |S_2|$. For NEb_{+-}, the spin magnitude of the larger black hole is approximately twice the spin magnitude of the smaller black hole, and the NEb_{+-} transverse kick is larger than the EQ_{+-} and NE_{+-} transverse kicks. Comparing NE_{+-} and NEb_{+-} for which the total mass and the mass ratio are the same, we find that the transverse kicks roughly scale with $a_1 + a_2$, as before. NE_{+-} and NEb_{+-} have sizable errors because the Bowen-York pulse occurs so close to the burst in the physical radiation which leads to uncertainty in t_0. Within the kick uncertainties, the unequal-mass cases are consistent. As before, Table 3.3 lists the kicks, both longitudinal and transverse, for these combined cases.
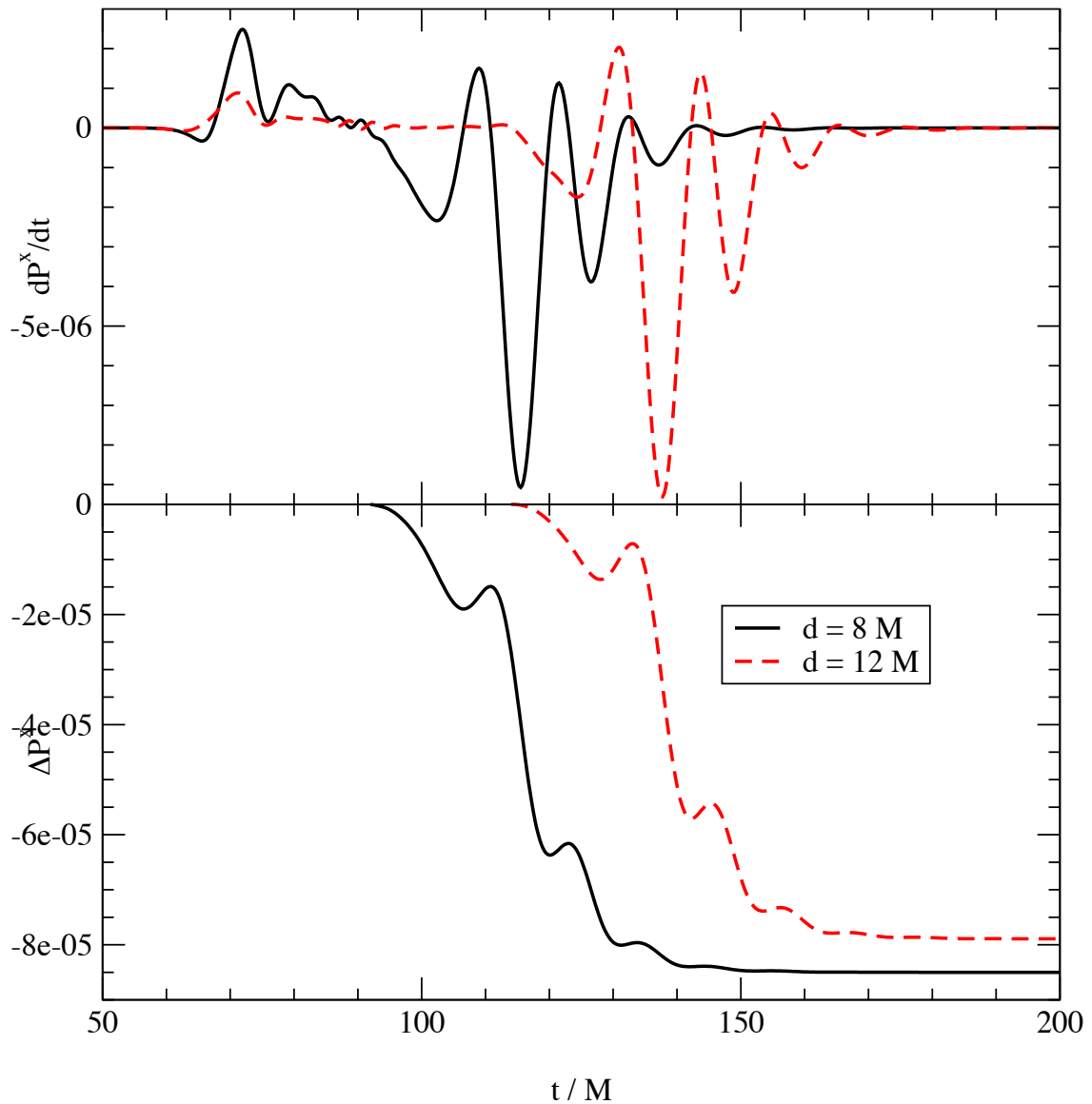
Figure 3.4: Transverse thrust dP^x/dt (top) and kick ΔP^x (bottom) for unequal masses and

spin, for coordinate separations $8M$ (NE$_{+-}$; black/solid) and $12M$ (NEa$_{+-}$; red/dashed).

NEa$_{+-}$ is the only case that uses a larger coordinate separation. The extraction radius

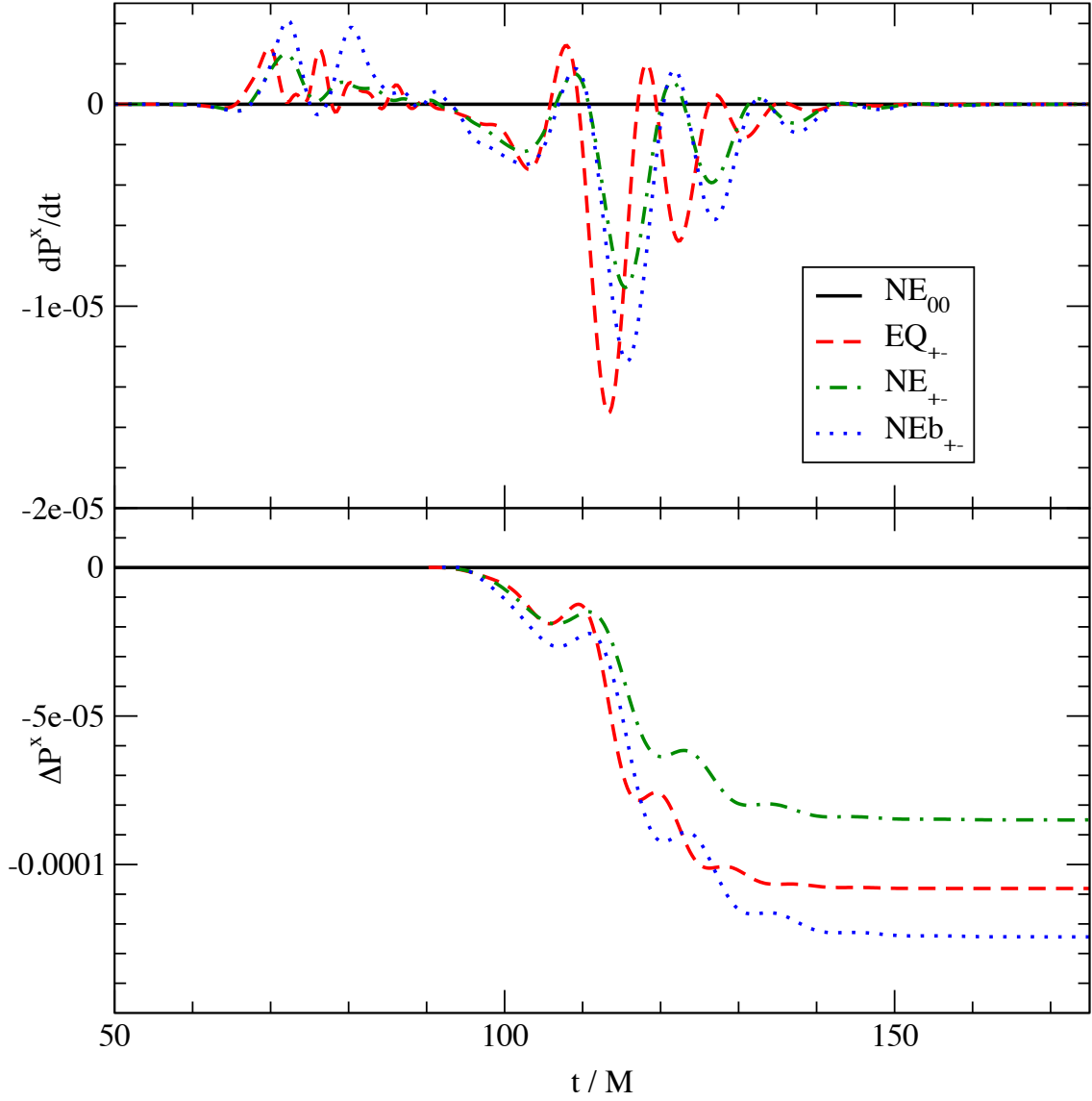R_{ext} is $60M$ in both cases.

Figure 3.5: Transverse thrust dP^x/dt (top) and kick ΔP^x (bottom) for NE_{00} (black), EQ_{+-} (red), and NE_{+-} (green), and NEb_{+-} (blue). In all cases the extraction radius R_{ext} is $60\,M$.

3.4.4 Analysis of results

To further analyze the results in Table 3.3, we develop expectations of the kicks due to mass ratio and black holes spins using post-Newtonian analysis. To estimate the momentum flux and the resulting recoil kick, we adapt (3.31a-b) of [62] to head-on mergers, yielding these expressions for the thrust

$$\dot{\vec{P}}_N = \frac{16}{105}\frac{\delta m}{m_T}\eta^2 \left(\frac{m_T}{r}\right)^4 \dot{r}\left(-\dot{r}^2 + \frac{2m_T}{r}\right)\hat{e}_y, \qquad (3.3)$$

and

$$\dot{\vec{P}}_{SO} = -\frac{16}{15}\frac{\eta^2\dot{r}^2}{r}\left(\frac{m_T}{r}\right)^4 (a_1 + a_2)\hat{e}_x, \qquad (3.4)$$

where $m_T = m_1 + m_2$, $\delta m = m_1 - m_2$, and r is the separation of the two black holes. $\dot{r} < 0$, so both expressions predict momentum loss due to radiation.

The relationship between black hole spins and the kicks resulting from them in black hole mergers was uncertain before such mergers were simulated. Sums of the magnitudes of the spin angular momenta (S), the Kerr parameters (S/m), or the dimensionless Kerr parameters (a/m) are possible scaling factors for the kicks. Eqs. (3.3) and (3.4) suggest that the kicks scale with $a_1 + a_2$, the sum of the Kerr parameter magnitudes. The results of our simulations support this prediction, and the initial parameters were chosen to test the prediction and other possible scaling factors. The accuracy of the prediction is surprising since the radiation responsible
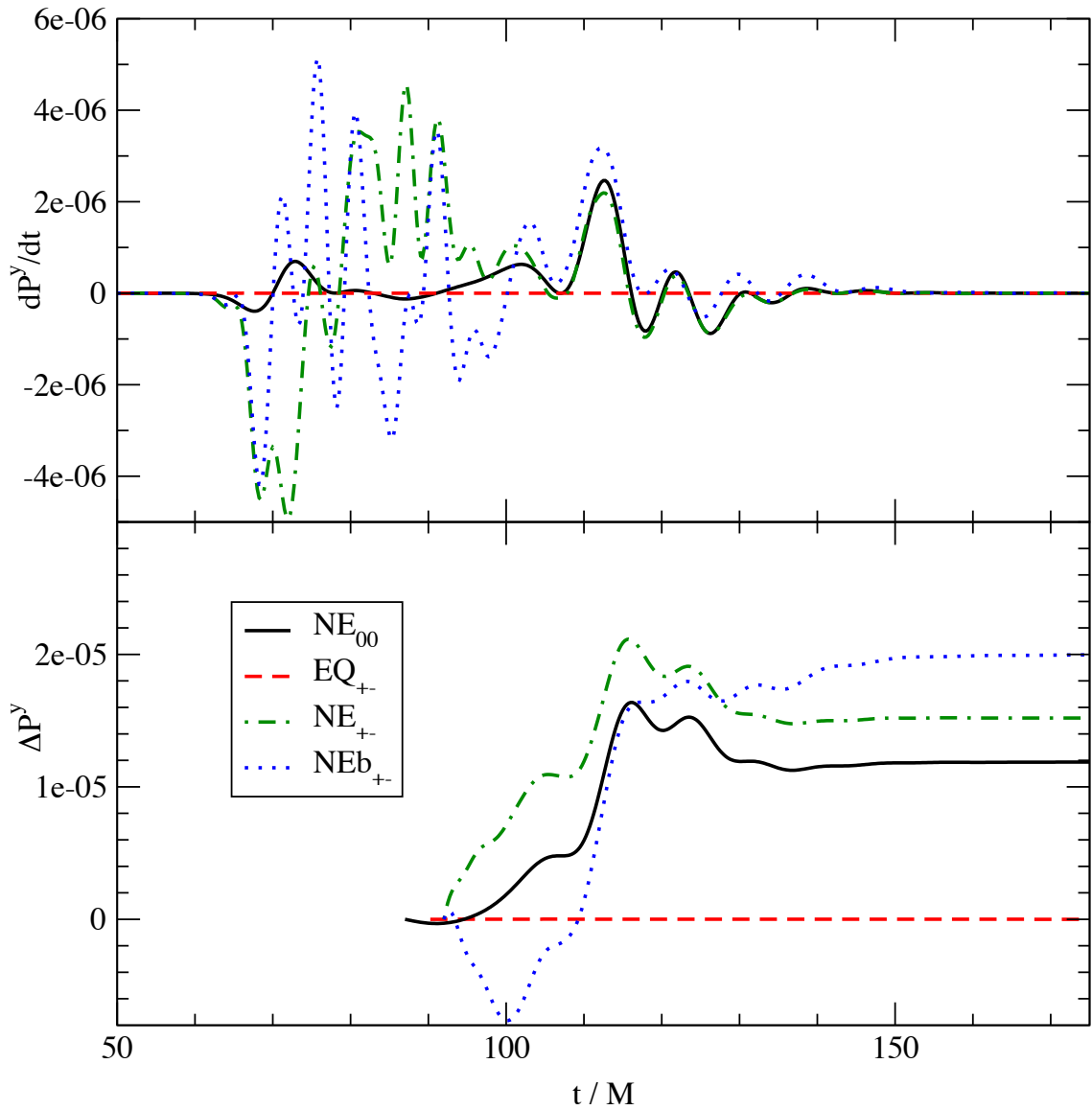
Figure 3.6: Longitudinal thrust dP^y/dt (top) and kick ΔP^y (bottom) for NE_{00} (black), EQ_{+-} (red), and NE_{+-} (green), and NEb_{+-} (blue). In all cases the extraction radius R_{ext} is $60\,M$

R_{ext}	t_0 (M)	ΔP^x $\times 10^{-5}(M)$	ΔP^y	v_x	v_y (km/s)	v
				NE_{00}		
$30M$	51.0	0.0	1.12 ± 0.02	0.0	2.71 ± 0.05	2.71 ± 0.05
$40M$	61.7	0.0	1.14 ± 0.04	0.0	2.76 ± 0.10	2.76 ± 0.10
$60M$	82.9	0.0	1.16 ± 0.04	0.0	2.81 ± 0.10	2.81 ± 0.10
				EQ_{+0}		
$20M$	46.1	-5.13 ± 0.04	0.0	-15.38 ± 0.12	0.0	15.38 ± 0.12
$30M$	56.6	-5.51 ± 0.01	0.0	-16.52 ± 0.03	0.0	16.52 ± 0.03
				EQ_{+-}		
$30M$	59.6	-10.70 ± 0.08	0.0	-32.08 ± 0.24	0.0	32.08 ± 0.24
$40M$	70.0	-10.53 ± 0.09	0.0	-31.57 ± 0.27	0.0	31.57 ± 0.27
$60M$	90.5	-10.81 ± 0.15	0.0	-32.41 ± 0.45	0.0	32.41 ± 0.45
				NE_{+-}		
$40M$	72.0	-8.47 ± 0.31	1.36 ± 1.10	-20.41 ± 0.75	3.28 ± 2.65	20.67 ± 0.85
$60M$	92.0	-8.50 ± 0.27	1.52 ± 1.36	-20.48 ± 0.67	3.66 ± 3.28	20.80 ± 0.88
				NEa_{+-}		
$40M$	94.0	-7.89 ± 0.26	1.04 ± 0.08	-18.92 ± 0.63	2.49 ± 0.20	19.08 ± 0.63
$60M$	114.0	-7.89 ± 0.21	1.00 ± 0.07	-18.92 ± 0.51	2.40 ± 0.17	19.07 ± 0.51
				NEb_{+-}		
$40M$	72.0	-12.33 ± 0.56	2.03 ± 0.64	-29.41 ± 1.34	4.84 ± 1.53	29.81 ± 1.35
$60M$	92.0	-12.44 ± 0.39	2.00 ± 0.55	-29.67 ± 0.93	4.77 ± 1.31	30.05 ± 0.94

Table 3.3: Final integrated momentum kicks and corresponding kick velocities. In each case, we have removed "BY pulse" effects through deferring integration until after the passage of the pulse. Quoted errors at each extraction radius are obtained by varying the integration starting time by $\pm 5M$.

for the kicks occurs late in the merger when the post-Newtonian approximations no longer hold. Fig. 3.7 compares transverse thrusts of the other spin runs to EQ_{+-}. Because the total mass m_T of the black holes provides the length scale of the system, the spin scaling factor is $(a_1 + a_2)/m_T$. The thrust amplitude for each run is scaled by its spin scaling factor relative to the EQ_{+-} spin scaling factor. In other words, the factor for the EQ_{+-} transverse thrust is set to 1, and the remaining transverse thrusts are scaled accordingly. Time is scaled by the ADM mass M_{ADM}, and the thrusts are aligned so the peaks coincide. With the PN-predicted scaling, the thrust amplitudes agree within 20% and thus validate the PN prediction outside its expected time regime.

3.5 Discussion

In this study of head-on black hole mergers, we investigated recoil kicks produced by spins and unequal masses. The spins were perpendicular to the initial separation vector of the puncture black holes, which is analogous to spins parallel or anti-parallel to the orbital angular momentum in inspiral mergers. A head-on collision approximates the plunge of an inspiral merger, but it has simpler dynamics. The infall direction is clear, which provides clear directionality for the recoil kicks and simplifies separation of the mass ratio and spin kicks. This allows us to interpret our results unambiguously. In our simulations the spin kicks are orthogonal to and significantly larger than the mass ratio kicks. The thrust and kick scale with the
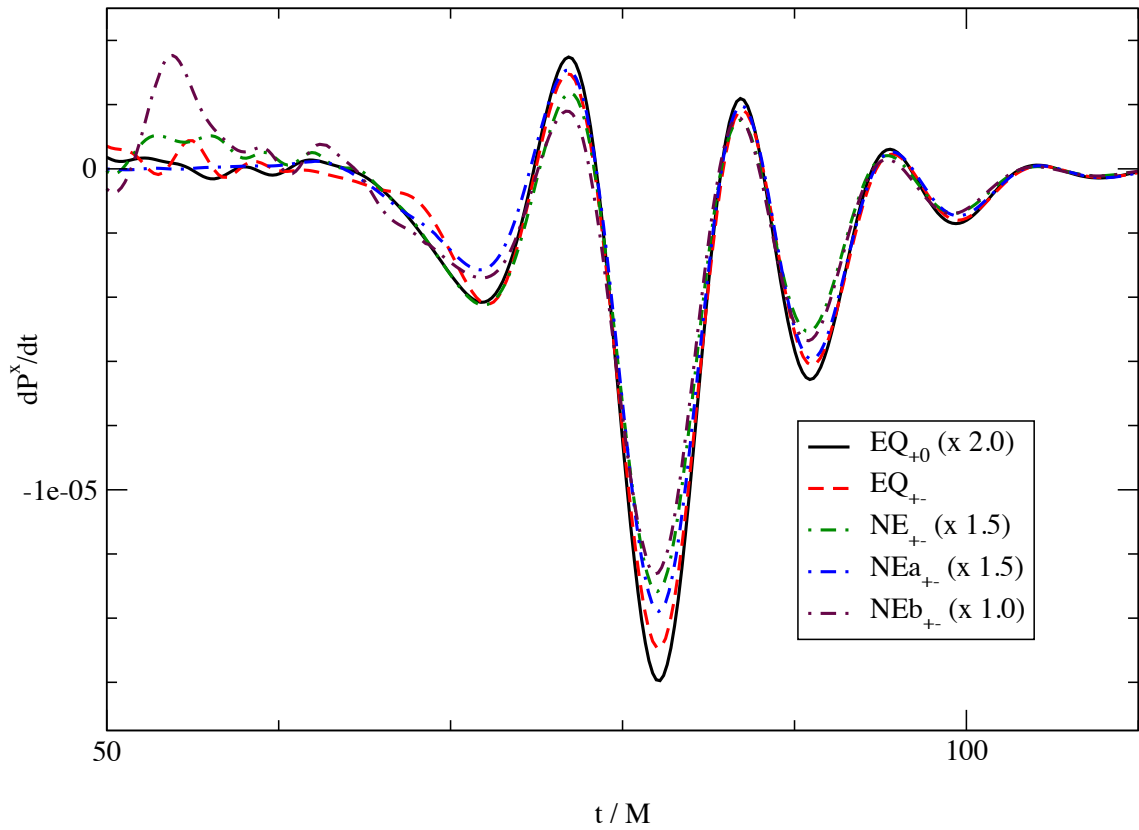
Figure 3.7: Transverse thrust dP^x/dt for all spinning data sets: EQ_{+0}, EQ_{+-}, NE_{+-}, NEa_{+-}, and NEb_{+-}. Time has been rescaled by the ADM mass for each data set, and translated relative to EQ_{+-} to line up peaks. Amplitudes have been scaled relative to EQ_{+-} according to PN predictions. These data were extracted at $R_{\text{ext}} = 30\,M$.

sum of the Kerr parameter magnitudes $a_1 + a_2$. These results are consistent with leading-order PN predictions (see [62]), which predict the orthogonality of the two types of kicks and the spin scaling. Since the PN predictions are not valid in the time regime when the bulk of the kick-producing radiation is emitted, this consistency is surprising. In the PN analysis the mass ratio kick results from coupling of the mass-quadrupole and mass-octupole moments, while the spin kick results from coupling of the mass-quadrupole and current-quadrupole moments. Spin does not contribute to the mass-octupole, and to leading-order, it does not contribute to the mass-quadrupole. This is validated by the mass-quadrupole-dominated ($l = 2, m = 2$) waveforms of the equal-mass cases (EQ_{00}, EQ_{+0}, and EQ_{+-}) in Fig.3.1. The three cases are indistinguishable after the initial Bowen-York pulses, despite the presence of spins in later cases.

The head-on kick study provides a transition from mass ratio kick investigations to more extensive investigations of combined kicks. The head-on collisions allow us to directly separate the spin kick from the mass ratio kick rather than determining the angle separating the kick contributions by fitting the data, which is necessary in the case of inspirals. Therefore, the head-on kick study provides clear validation of PN expressions (as in [62]) to predict the spin scaling of kicks in the orbital plane ("in-plane kicks").

Exploration of kicks has progressed since the present results were published, as numerical relativity simulations have encompassed larger mass ratios (with 100:1 results published recently [72]) and sampled the space of spin configurations more generally. Among the most important results is "superkicks", which are not present

95

in the head-on results. These are kicks oriented out of the plane, with magnitudes significantly larger ($\sim$ 10 times) than in-plane spin kicks, that originate from components of spin within the orbital plane. They were suggested in [38] and later investigated for equal-mass [45] and unequal-mass [10] cases.

With these developments, the empirical formulas used to predict recoil kicks have evolved as well [105, 73]. The study of these formulas has been an ongoing endeavor of the Goddard group. As a member of the group, I have contributed to this study, while I developed the moving patches technique discussed in Chapter 4. In the process, I have coauthored four papers on the subject of kick formulas.[39, 9, 10, 105] The revised formulas, motivated by PN analysis, maintain the same form and spin scaling confirmed by the head-on study discussed in this chapter (Eqs. (3.3) and (3.4)), with the addition of the out-of-plane superkick component.

Beyond the empirical description of kicks they provide, these formulas are crucial for calculations of astrophysical processes such as the retention of the remnant supermassive black holes in galactic mergers [77], the formation of supermassive black holes through hierarchical structure formation in the early universe [106], and growth of intermediate-mass black holes in globular clusters [58]. By predicting magnitudes, probabilities, and distributions for the recoil kicks essential to these astrophysical processes, the formulas are critical for evaluating different models of the processes. Therefore the accuracy of the kick formulas, to which the present results contributed, is crucial for the study of the astrophysical processes.

Chapter 4

Moving Patches

The material in this chapter describes the moving patches technique, a project I conducted independently, while I was a member of the NASA Goddard Numerical Relativity group. Though I was fortunate to receive assistance from the members of the group, the research discussed in this chapter is my own. It was my primary focus from 2009 to 2011.

4.1 Overview

A black hole binary merger simulation calculates a spacetime that represents two black holes, but we treat the black holes as separate entities that approach and merge, and this suggests a natural division of the dynamics

1. the motion of the black holes across the grid

2. the evolution of the individual black holes

where the first is a dynamical process while the second is a secular process. The division holds well until merger when the positions of the individual black holes become meaningless. Such a division is suggested by the heuristic IRS model discussed in Chap. 2 and in [11], which treats the binary as a rigid body rotating at the orbital frequency with a secular decrease in length. Considering the black holes as

separate bodies and dividing the dynamics as above suggests a technique to improve the accuracy of black hole binary simulations:

1. Track the positions of the black holes.

2. Evolve the fields near each black hole in a coordinate frame that moves with the black hole's position. (*comoving*)

This eliminates the dynamical component of the fields' evolution. What remains is the secular evolution. Because the secular evolution is much slower, we may be able to calculate it more accurately, as we might calculate a single non-moving black hole. The dynamical component of the simulation becomes the task of tracking the black hole positionon the simulation grid.

Consider a simulation of a single non-moving black hole representing the Kerr black hole. The Kerr black hole is stationary, so the metric components are constant in time. For a suitable choice of gauge, the simulated black hole should be stationary as well. In fact, this is a requirement of a suitable gauge. In actual simulations of a single non-moving black hole, the fields are very close to constant in time after an initial period of rapid gauge adjustment. After the gauge "settles in", the fields' evolution equations become

$$\partial_t u_a \approx 0 \qquad\qquad (4.1)$$

for all the fields u_a, as we would expect for the Kerr black hole. Since the fields are not changing, we can calculate them more accurately. For Eq. (4.1) to be true requires that the right-hand-sides of the actual evolution equations approach 0. This involves the precise cancellation of terms involving different fields and their

derivatives so that the sum remains 0. Once the fields' time variation ceases, the evolution equation maintains this quiescent state. The fields are not stationary in a black hole binary simulation, but the variation is dominated by advection of the fields, so it is like two non-moving black holes with advection added. On the right-hand-sides, the non-advection terms almost cancel each other with the residual producing the secular evolution. Errors from the large advection terms disrupt the near-cancellation and interfere with the secular evolution. Evolving each black hole in a comoving frame eliminates the advection terms and therefore increases the accuracy of the evolution near the black hole.

4.1.1 Motivation

The most direct impact of the technique suggested above would be increased accuracy in the evolved fields. Normally, greater accuracy is gained through higher resolution in time and space. For three-dimensional finite difference simulations, increasing the spatial resolution by 20% typically requires 72% more memory and twice the computational speed. Though the available computational power continues increasing, numerical relativity simulations demand ever-increasing resolution as they explore more ambitious initial configurations, such as the recent results of a 100:1 mass ratio merger by Lousto and Zlochower [72] and the merger of black holes with near-extremal spins ($\hat{a} = 0.95$) by Lovelace *et al.*. The effective resolution of a simulation is determined by the mass of the smaller black hole. So a simulation with a nominal maximum resolution of $M_{ADM}/80$ in terms of the total ADM mass and

a smaller black hole of mass $m_< = M_{ADM}/10$ has an effective resolution of $m_</8$, the resolution of a $M_{ADM} = 1$ single black hole simulation of maximum resolution $M_{ADM}/8$. Extremal spin also reduces the effective resolution of a simulation since the spin produces a more compact black hole. This can be seen in the case of a Kerr black hole, where a maximal spin reduces the innermost stable circular orbit from a radius of $6M$ for a Schwarzschild black hole to $1M$. The moving frame technique could produce the increased accuracy required by more ambitious initial configurations without increasing the spatial resolution. Also it could increase the speed of simulations, since a given accuracy could be obtained from a lower spatial resolution. This would allow simulations of longer inspirals, and therefore longer waveforms. Finally, since the technique eliminates much of the dynamics near the black holes, it could allow larger time steps when combined with an implicit time stepping scheme, which would also increase the speed of simulations.

4.1.2   Changing Coordinate Systems

General relativity is a covariant theory, so the tensor equations do not depend on the particular coordinate system, and we can choose the coordinate system used to express the equations. The coordinates specify points in spacetime, and also they provide a natural basis for the tensor components in calculations. If we were using a coordinate basis, and we changed the coordinate system, normally we would change to the new coordinate basis, but we could use the previous coordinate basis, just as we could use other non-coordinate bases. For the technique suggested above, there is

no need to change bases, so all quantities are expressed in the fixed grid coordinate system where the black holes move. This technique of changing the coordinate system without changing the basis is a modification of the dual coordinate technique [94] used by the Caltech-Cornell collaboration for numerical relativity simulations.

In a simulation of a black hole binary, suppose that x^i is the *fixed* coordinate system in which the black holes move, and u_a is one of the fields, a function of time and space:

$$u_a = u_a(x^i, t) \,. \tag{4.2}$$

Now let $x^{j'}$ be a coordinate system that moves with one of the black holes. There exist transformations between the two coordinate systems, $x^{j'} = x^{j'}(x^i, t)$ and its inverse $x^i = x^i(x^{j'}, t)$. If we substitute the inverse transformation into Eq. (4.2), we obtain a new function of $x^{j'}$ and t,

$$\widetilde{u}_a(x^{j'}, t) = u_a(x^i(x^{j'}, t), t) \,. \tag{4.3}$$

Making such a coordinate transformation modifies the system of evolution equations by changing the functional form (Eq. (4.3)) and the derivatives. For spatial derivatives, the coordinate transformation adds a factor to the partial derivative,

$$\partial_{j'} \widetilde{u}_a = \partial_i u_a \frac{\partial x^i}{\partial x^{j'}} \,. \tag{4.4}$$

where $\partial_{j'}$ and ∂_i are the spatial derivatives in the moving and fixed coordinate systems, respectively. Since this approach changes only the spatial coordinates, time derivatives are taken with respect to the same time coordinate, but the coordinate transformation does add an additional term to the time derivative. From Eq. (4.3)

101

we have,

$$\partial_t \widetilde{u}_a = \partial_t u_a + \dot{x}^i \partial_i u_a \,. \tag{4.5}$$

In a black hole binary the inspiral motion of each black hole consists of a slow radial infall and a tangential velocity, so one approach to comoving coordinates is global linear transformations. The dual coordinates method [94] uses a rotation and a rescaling of the coordinate system. Rotating the coordinate system with angular frequency Ω, matched to the orbital motion of the system, counteracts the tangential velocities, while rescaling the coordinates by a factor $a(t)$ counteracts the radial infalls. The approach presented below is much simpler since it requires no rotations or rescalings.

4.1.3 Moving Patches Technique

We create a local comoving coordinate system around each black hole as suggested above, where the coordinate transformation to the fixed grid is a simple translation rather a global transformation. For black hole A, the transformation is

$$x_A^i(t) = \delta_{j'}^i x_A^{j'} - q_A^i(t) \,, \tag{4.6}$$

where $q_A^i(t)$ is a translation that minimizes the motion of black hole A in its local coordinates $x_A^{j'}$. The label A is necessary since each black hole has a separate transformation to the fixed grid coordinate system. With this transformation, Eqs. (4.4) and (4.5) become,

$$\partial_{j'} \widetilde{u}_a = \delta_{j'}^i \partial_i u_a \,, \tag{4.7}$$

102

and

$$\partial_t \widetilde{u_a} = \partial_t u_a - \partial_t q^i \partial_i u_a \,. \tag{4.8}$$

These smaller grids, or patches, "move" over the larger fixed grid. The movement is manifested through the coordinate transformations between the patches and the fixed grid. Each patch boundary introduces a boundary for the fixed grid, and the patch and the fixed grid must communicate field data at the boundary. The coordinates of the boundary points are transformed for these communications, and the change of the coordinates reflects the motion of the grid.

The moving patches technique is the use of a local patch coordinate system moving with the black hole to calculate the fields' evolution near each black hole. In the patch coordinate system the evolution equations acquire an additional advection term that corresponds to the black hole's motion on the larger grid. The advection velocity appearing in the additional terms is the patch velocity, the time derivative of the coordinate transformation to the fixed grid from the patch grid. The technique improves the accuracy of the evolved fields and the data derived from them. By eliminating the black holes' motion on the grid, the technique also reveals the impact of black holes' motion on simulation errors. In what follows, I present my implementation of the technique in HAHNDOL as well as comparisons between simulations using the technique ("patch" runs) and simulations using a single grid ("non-patch" runs) to evaluate the technique and investigate the error generated from black hole motion.

4.2 Implementation

The moving patches technique requires

- creation of a patch grid for each black hole

- specification of the coordinate transformation between each patch and the fixed grid

- modification of the evolution equations on the moving patches

- synchronization of field data between each patch and the fixed grid at their interface

4.2.1 Grid Generation and Initial Fill

Since PARAMESH is block-based, creating a patch grid for a black hole in HAHNDOL means creating a collection of blocks that cover a region including and surrounding the black hole. The coordinates of these blocks are in the local coordinate system associated with the black hole, and the blocks are assigned a common number for the patch number field (which is a PARAMESH user-defined block field). This implicitly defines the patch. The patch is also explicitly defined by a data structure in HAHNDOL, `patch`, which stores metadata about the patch, such as the black hole with which the patch is associated and the coordinate range of the patch.

The standard initial grid creation procedure in HAHNDOL reads rules for fixed-mesh refinement (FMR) from a parameter file. These FMR rules define refinement zones with reference to the fixed coordinate system origin or a puncture location.

For moving patches HAHNDOL creates one patch grid for each puncture following the FMR rules associated with that puncture. The patch grid only includes the highest refinement levels with the minimum patch level specified by the parameter file. HAHNDOL maintains an association between the patch grid and the puncture in `patch` and updates the patch's coordinate transformation based on the motion of the puncture. PARAMESH assumes a single coordinate system for all blocks. To avoid assigning overlapping coordinates to patch and fixed grid blocks, HAHNDOL sets the initial translation $q_A^i(0)$ to an integer multiple of the maximum dimension of the fixed grid, $q_A^i(0) = \delta_1^i n * \max \left(x_{max}^k - x_{min}^k \right), (k = 1, 2, 3)$. Though this initial translation is necessary to place the patch in a separate computational domain in PARAMESH, it has no effect on the evolution.

The moving patches technique offers no benefit in solving the constraints for the initial data, so HAHNDOL solves for the initial data on the fixed grid. Then HAHNDOL copies the field data to each patch from the corresponding portion of the fixed grid.

4.2.2   Evolution

Once the evolution begins, HAHNDOL updates $\dot{q}_A^i$ at the beginning of each iteration to minimize the puncture motion on the patch. Recall that the puncture tracker performs a first-order Euler integration of $-\beta^i(x_p^j)$ to calculate the track of each puncture. With moving patches, HAHNDOL uses $\beta^i(x_p^j)$ for puncture A to update $\dot{q}_A^i$, the velocity in the advection term Eq. (4.8) that is added to the field

evolution equations. Currently, $\dot{q}_A^i = -\beta^i(x_A^j)$, so in principle $q_A^i(t)$ is the same as puncture track A, In the future, $\dot{q}_A^i$ could be a function of $\dot{\beta}^i(x_p^j)$ or a higher time derivative of $\beta^i(x_p^j)$ to provide a smoother function. The current choice for $\dot{q}_A^i$ eliminates all puncture motion on the local patch. By comparison, the dual coordinates approach [94] uses a feedback control system to drive the apparent horizon centers back to their initial values by adjusting $\partial^3 a/\partial t^3$ and $\partial^3 \Omega/\partial t^3$, rather than adjusting the scaling factor a and the rotational frequency Ω directly. If $\dot{q}_A^i$ were driven to 0 by a similar feedback control system, there would be slight puncture motion on the patch but the smoother coordinate transformations might further reduce the errors in simulations using the moving patches technique.

While $\dot{q}_A^i$ is updated at each iteration, the coordinate transformation $x_A^i(x_A^{j'}, t)$ must be updated within the RK4 iteration. The patches are in motion, so when a field u_a is evaluated at the midpoint of an iteration, the patch locations are different than their locations at the beginning of the iteration, and so the coordinate transformation from each patch to the fixed grid is different as well. Specifically, the transformation for patch A, at the midpoint and end of iteration t_n is

$$x_A^i(x_A^{j'}, t_n + \tfrac{1}{2}\Delta t) = x_A^i(x_A^{j'}, t_n) + \tfrac{1}{2}\Delta t\, \dot{q}_A^i, \tag{4.9}$$

and

$$x_A^i(x_A^{j'}, t_{n+1}) = x_A^i(x_A^{j'}, t_n + \tfrac{1}{2}\Delta t) + \tfrac{1}{2}\Delta t\dot{q}_A^i \tag{4.10}$$

where Δt is the integration timestep. If the coordinate transformations are not updated at the midpoint, the two-way synchronization between each patch and the fixed grid uses incorrect grid points, and the simulation becomes unstable. Though

106

x_A^i is updated within the RK4 iteration, the result is not an RK4 integration of x_A^i, since $\dot{x}_A^i$ is constant throughout the iteration. HAHNDOL performs the two-way synchronization between each patch and the fixed grid after the PARAMESH guard cell filling. In the first step of the synchronization, HAHNDOL interpolates from the fixed grid to fill the guard cells on the outer boundary of each patch, and in the second step, it interpolates from each patch back to the portion of the fixed grid overlapped by that patch.

The interpolation in the synchronization could introduce error that would counteract the accuracy gain from the coordinate transformation. Since the finite differencing in HAHNDOL is sixth-order, I increased the accuracy of the HAHNDOL interpolator to sixth-order by increasing the stencil to six points. I confirmed the accuracy of the interpolator with polynomial test data. For a sixth-order polynomial function of the coordinates, the interpolation error was sixth-order convergent, while for lower-order polynomial functions, the error was zero to machine precision. Since then the sixth-order interpolator has been used routinely for simulations by the Goddard group.

Initially the fixed grid includes the same refinement levels as the patches. Since the patches overwrite the portions of the fixed grid that they overlap, and the patches include the higher refinement levels around the punctures, there is no need for the fixed grid to include those higher refinement levels. Once evolution begins, the AMR procedure derefines the fixed grid until the maximum refinement level of the fixed grid is one less than the patches' minimum refinement level. For the initial implementation of moving patches in HAHNDOL, AMR does not run on

the patches because refinement to and derefinement from the patch minimum level is more complicated than the existing AMR in HAHNDOL. Therefore, the patches maintain their initial coordinate sizes and cube shapes, but since the patches move with the black holes, grid adaptation is not crucial.

4.3 Evaluating the Moving Patches Technique

The goal of the moving patches technique is to reduce the error in black hole simulations due to the motion of the black holes. Each evolution equation Eqs. (1.12)-(1.16) includes an advection term proportional to $\beta^k \partial_k u_a$. The shift can be written as

$$\beta^i(x^j) = \beta^i_{str}(x^j) + \beta^i_{adv} \,, \tag{4.11}$$

where β^i_{str} is the portion of the shift that counteracts grid stretching, and β^i_{adv} is a spatial constant equal to the shift at the puncture position. It is responsible for the puncture's motion on the grid. On the fixed grid, where the evolution equations are unaltered, this component of the shift produces the active field evolution responsible for the bulk motion of the black hole.

On the patch, the black hole motion is manifested only in the coordinate transformation to the fixed grid. Because of the changes to the evolution equations Eq. (4.8), the puncture does not move on the patch, and so $\beta^i_{adv} = 0$ on the patch, which eliminates part of the advection term and any error associated with it. HAH-NDOL uses fifth-order one-sided differencing near the puncture, so the fifth-order error should be reduced on the patch with the reduction proportional to $\beta^k_{adv} \partial_k u_a$.

Therefore the error reduction should be greater in the strong field region and greater for large puncture momentum, since the advection on the grid corresponds to the puncture momentum. Moving away from the puncture, where spatial derivatives of the fields are smaller and the advection terms are higher-order, the effect of the moving patches technique should be less dramatic. The actual error will depend on $\partial_i u_a$. Also, other larger error sources may mask the error reduction.

Constraint violations and puncture coordinate positions provide estimates of the error reduction. Because the constraints combine the evolution variables, constraint violations indicate the error of the system rather than the error of a particular field. Norms of constraint violations measure error for refinement levels. In contrast, puncture coordinate errors are local in space, and they are direct results of errors in the shift, though they represent errors in the system because of the coupling of the evolution equations. Though the puncture positions depend on the gauge choice, we can use them to assess the errors in a system's dynamics. Specifically we can measure errors in the phase, frequency, and a quantity we define which is equal to the local unit Newtonian angular momentum ℓ_{PT}, as calculated from the two puncture tracks. As we will see below, these errors indicate errors in the gravitational radiation, and accurate gravitational waveforms are among the most important results from numerical relativity (NR) simulations.

The moving patches technique, as implemented in HAHNDOL with PARAMESH, slowed down the simulations considerably, so evolutions using the technique did not evolve long enough to make direct comparisons of waveforms. The patch synchronization process, which updates the patches and the fixed grid at the patch

boundary, could not be integrated into the regular PARAMESH guard cell filling procedures that update the guard cells at block boundaries as well as the outer boundary of the grid. While PARAMESH uses efficient MPI communications that minimize interprocessor communications and allow data exchange between multiple processor pair simultaneously, the patch synchronization uses the HAHNDOL interpolator, which was designed for much smaller data exchanges and therefore uses the relatively inefficient MPI broadcast communications. The performance limitations are discussed further in the Performance subsection below.

4.3.1 Head-On Runs

Head-on runs provide a simple test of the moving patches technique with a non-trivial evolution. Since they use octant symmetry, the head-on runs are faster, which facilitates testing the code. For the head-on runs, I used Brill-Lindquist initial data (no linear momentum or spin) with equal-mass punctures located at $x = \pm 8M$ on a grid that extended $128M$ from the origin. I performed the simulation at three resolutions: $M/16$, $M/20$, and $M/24$, both with and without moving patches.

Since constraint violations show deviations from Einstein's equations Eq. (1.1) and therefore error, I compared the Hamiltonian constraint convergence with and without moving patches to make sure that the moving patches technique did not decrease the order of convergence. In particular, I was concerned that the synchronization between the fixed grid and the patches would introduce error that would counteract any gain in accuracy on the patches themselves. Though I had tested

the interpolator with test data, I wanted to make sure it would work in an actual evolution.

At $t = 30M$, I compared one-dimensional cuts along the x axis at the boundary between a patch and the fixed grid. The boundary is located in refinement level 7 of the fixed grid, the finest fixed grid level in the moving patches runs. Initially, the patch boundary crosses the x axis at $x = 4$ on the inner side, closest to the origin, and $x = 12$ on the outer side. As the simulation runs, both crossing points drift toward the origin as the black hole falls inwards and the patch moves with it.

While these one-dimensional cuts show the convergence at specific points in space, the presence of high frequency noise complicates their interpretation. Nonetheless, they are useful for evaluating the convergence behavior at the edge of the patch. Because the punctures follow straight-line trajectories in head-on simulations (for nonspinning punctures), we can make meaningful comparisons between the cases with and without patches. For more general simulations, such as inspirals, one-dimensional cuts are less informative since the punctures move across the grid, and they generally do not lie on the cuts. The L1 norm over an entire refinement level is more useful in these cases because it provides a broader evaluation of the constraint convergence, and the puncture positions affect the norm less.

We can expect convergence up to sixth-order for C_H. The finite differencing error increases to sixth-order moving away from the puncture as the upwinding for the advection terms decreases. The time integration error is much smaller than other error sources, and so it does not determine the overall convergence. At refinement boundaries, we expect the convergence order of C_H to decrease by two, since

it includes second. When the finite difference stencils include points from different resolutions, the cancellations necessary for sixth-order convergence are incomplete, and the finite differencing decreases by one order for first derivatives and two orders for second derivatives [61]. The constraint convergence is even lower near the puncture due to the large field gradients.

The one-dimensional cuts show similar fifth-order convergence for C_H for the patch and non-patch runs. I only compared the portion of level 7 of the fixed grid that was not covered by the patch, since the covered portion is overwritten during every iteration. Fig. 4.1 shows C_H on level 7 of the fixed grid in the inner portion, while Fig. 4.2 shows C_H in the outer portion. Because the patch runs show convergence at the same order as the non-patch runs, the interpolation shows no sign of decreasing the constraint convergence.

4.3.2   Inspirals

Inspirals should provide the most robust and realistic test of the moving patches technique. Head-on mergers provide an initial test of the technique, but they are a special case. Inspirals address the question of advection-generated error in a more general case and better demonstrate the technique's usefulness. The moving patches technique would be especially useful for inspiral cases such as large mass ratio inspirals and inspirals of highly spinning black holes. Large mass ratio simulations require higher resolution to adequately resolve the smaller black hole. Highly spinning black holes require higher resolution because their characteristic scales are
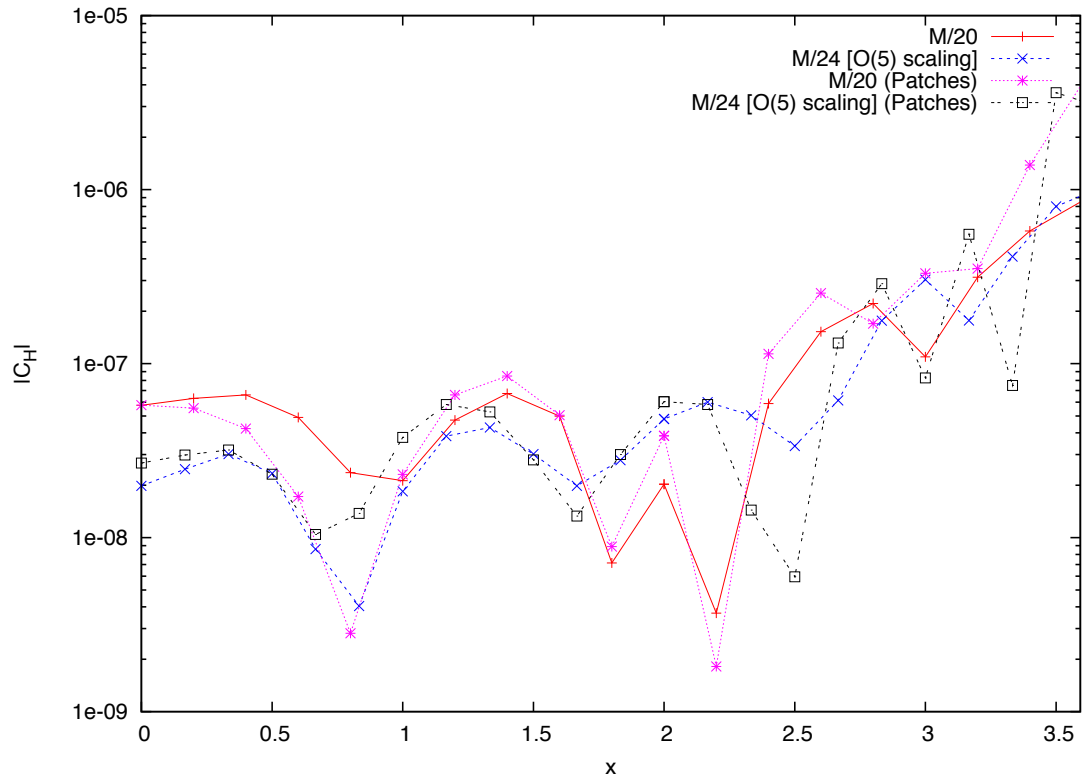
Figure 4.1: C_H to the left of the patch (level 7 of the fixed grid) for the head-on tests.

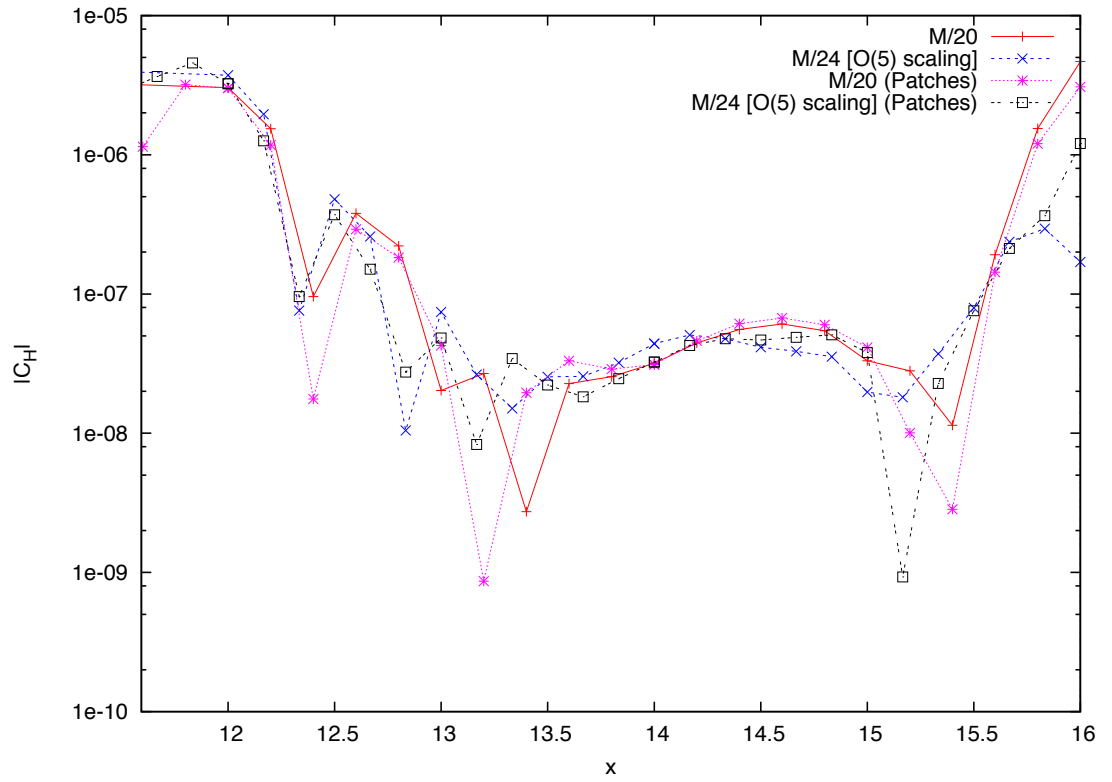The convergence is fifth-order, both with and without moving patches.

Figure 4.2: C_H to the right of the patch (level 7 of the fixed grid) for the head-on tests.

The convergence is fifth-order, both with and without moving patches.

smaller relative to their masses. By improving the accuracy of simulations, moving patches could lower the resolution required for both types of simulations.

For the inspiral tests, I used the moving patches technique to rerun a simulation previously run by the Goddard group and compared the patch results with the existing results. The previous simulation was a merger of two spinning equal-mass black holes with the spin axes of the first and second black hole aligned and anti-aligned, respectively, with the orbital angular momentum vector. For both black holes, the magnitude of the dimensionless spin parameter $\hat{a}$ was 0.8.

The initial data were calculated with the `TwoPunctures` elliptic solver [7] using initial data parameters derived from Husa *et al.*[60], which sought to reduce eccentricity in black hole binary evolutions with a post-Newtonian evolution up to the desired coordinate separation of $11M$. Husa *et al.*[60], used the momenta at separation $11M$ in the post-Newtonian evolution as the starting momenta for a puncture evolution of equal-mass, nonspinning black holes initially separated by $11M$. The momenta and coordinate separation from Husa *et al.*[60] were used for the Goddard simulations (both with and without patches). Since the Goddard simulations added spin to the black holes, the initial puncture masses were decreased to keep the apparent horizon masses at $0.5M$. The parameters for the initial data are listed in Table 4.1.

The outer boundaries were located $1536M$ from the origin. The previous simulation was performed at the following resolutions: $3M/128$, $3M/160$, $M/64$, and $3M/224$. To emphasize the relative resolutions of the simulations, I will refer

Puncture	$m_p(M)$	$y(M)$	$P^x(M)$	$P^y(M)$	S^z/M^2
1	0.301805	-5.5	0.0900993	0.000709412	0.8
2	0.301805	5.5	-0.0900993	-0.000709412	-0.8

Table 4.1: Initial data parameters for inspirals. $x = 0$, $z = 0$, $P^z = 0$, $S^x = S^y = 0$ for both punctures.

to these simulations as I_{128}, I_{160}, I_{192}, and I_{224}, respectively.[1] The simulation with patches had a resolution of $3M/128$, and I will refer to it as P_{128}. It used patches for the top three refinement levels, levels 12-14.

As currently implemented in PARAMESH, simulations using moving patches are inefficient on the large numbers of processors required for these runs, and the moving patches technique is not ready for production applications. Nonetheless, we can apply the currently limited code to study the impact of moving patches on accuracy of simulations. Because of the performance impact of the moving patches technique, P_{128} ended at $t = 173M$.

4.3.2.1 Constraint Violations

As discussed previously, by reducing error in the advection terms of the fields' evolution equations, the moving patches technique should decrease the error in the fields, which in turn should reduce the constraint violations. Fig. 4.3 shows the L1

---

[1] The label for the $M/64$ simulation results from the identity $\frac{1}{64} = \frac{3}{192}$

norm of the Hamiltonian constraint ($|C_H|_{L1}$) for level 14, the highest refinement level, for all the simulations. P_{128} lies between I_{160} and I_{192}, and it is smoother than all of the non-patch runs indicating a lower level of numerical noise. Fig. 4.4 shows the L1 norm of the x momentum constraint ($|C_{M^x}|_{L1}$) for level 14. For $|C_{M^x}|_{L1}$ the P_{128} results are similar to the I_{160} results. The relative improvement in accuracy is less for $|C_{M^x}|_{L1}$ than for $|C_H|_{L1}$. $|C_{M^x}|_{L1}$ is roughly 20 times smaller than $|C_H|$, and so it may already be less affected by error from the advection terms. Also note that C_H includes second spatial derivatives of g_{ij} in the spatial scalar curvature R while C_{M^i} only has first spatial derivatives of K_{ij} and K. The second derivatives are more sensitive to noise, and so the advantage of eliminating the fields' advection across the grid using the moving patches technique is more likely to be observed in C_H. The results for $|C_{M^y}|_{L1}$ and $|C_{M^z}|_{L1}$ are similar, though $|C_{M^z}|_{L1}$ is much smaller.

Figs. 4.5 and 4.6 show $|C_H|_{L1}$ and $|C_{M^x}|_{L1}$ for the second finest level, level 13. For $|C_H|_{L1}$ P_{128} is similar to I_{160}, after the first 30M of evolution. For $|C_{M^x}|_{L1}$ P_{128} is very close to I_{128}. For level 13, $|C_{M^y}|_{L1}$ and $|C_{M^z}|_{L1}$ are similar to $|C_{M^x}|_{L1}$, as they were for level 14. Fig. 4.7 shows $|C_H|_{L1}$ for level 12, the coarsest patch level. In this case, the error is larger for P_{128} than for I_{160}, though it is still significantly less than that of I_{128}. $|C_{M^x}|_{L1}$, $|C_{M^y}|_{L1}$, and $|C_{M^z}|_{L1}$ are not shown for level 12, but they are similar to the results for level 13. So the moving patches technique shows the greatest error reduction, as measured by the constraint L1 norms, at the highest refinement level, and the relative error reduction decreases moving into more coarse patch levels farther from the punctures.

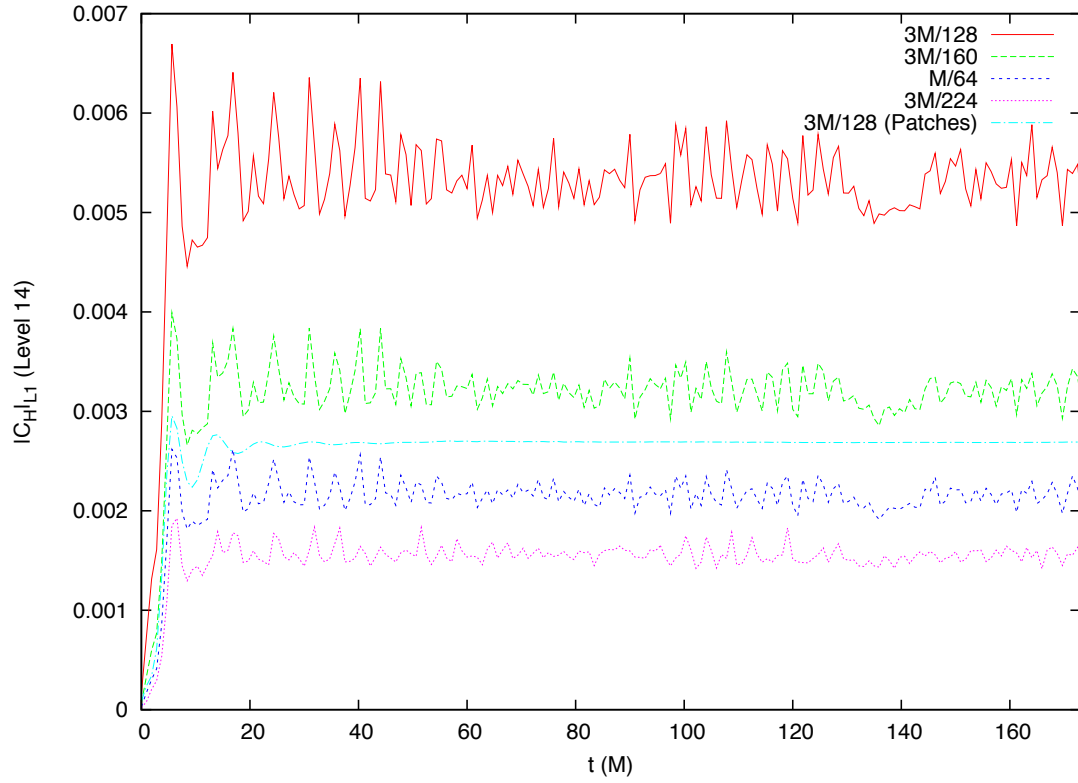Fig. 4.8 and 4.9 show $|C_H|_{L1}$ and $|C_{M^x}|_{L1}$ for level 9, which is on the fixed

Figure 4.3: Hamiltonian constraint (L1 norm) for level 14 (finest) in the inspiral test. The Hamiltonian constraint for the patches run shows a significant reduction.
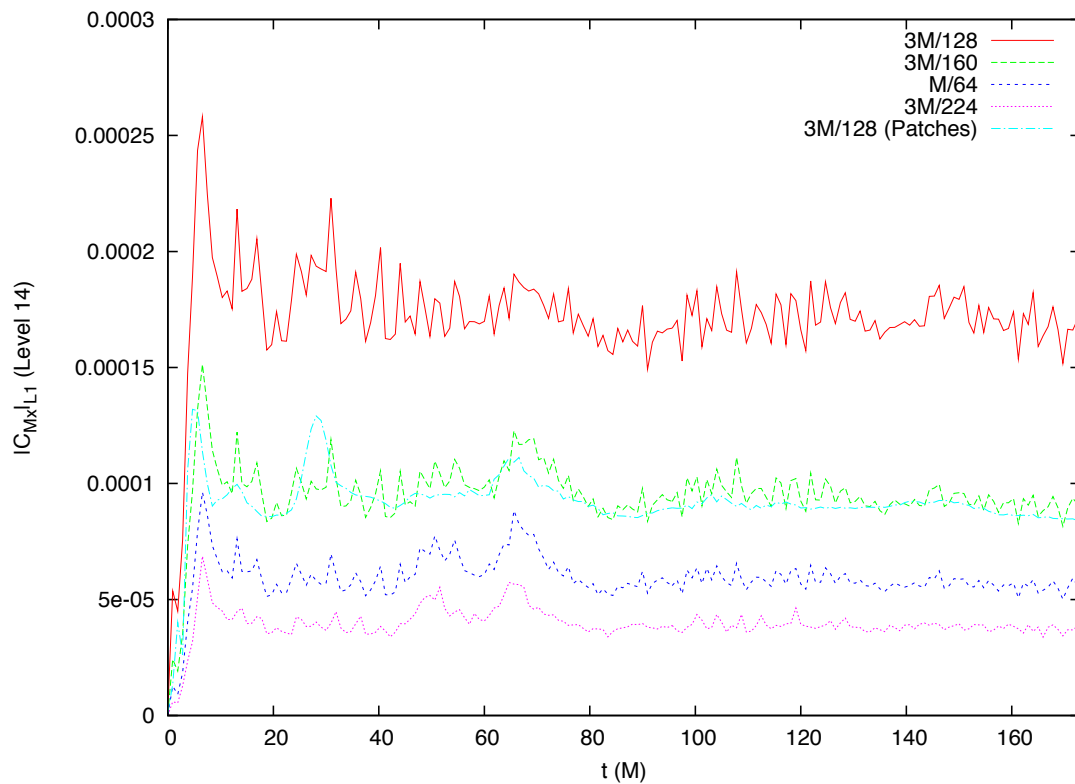
Figure 4.4: X momentum constraint (L1 norm) for level 14 (finest) in the inspiral test. The X momentum constraint for the patches run shows a reduction, though it is is not as large as the reduction for the Hamiltonian constraint in Fig. 4.3.
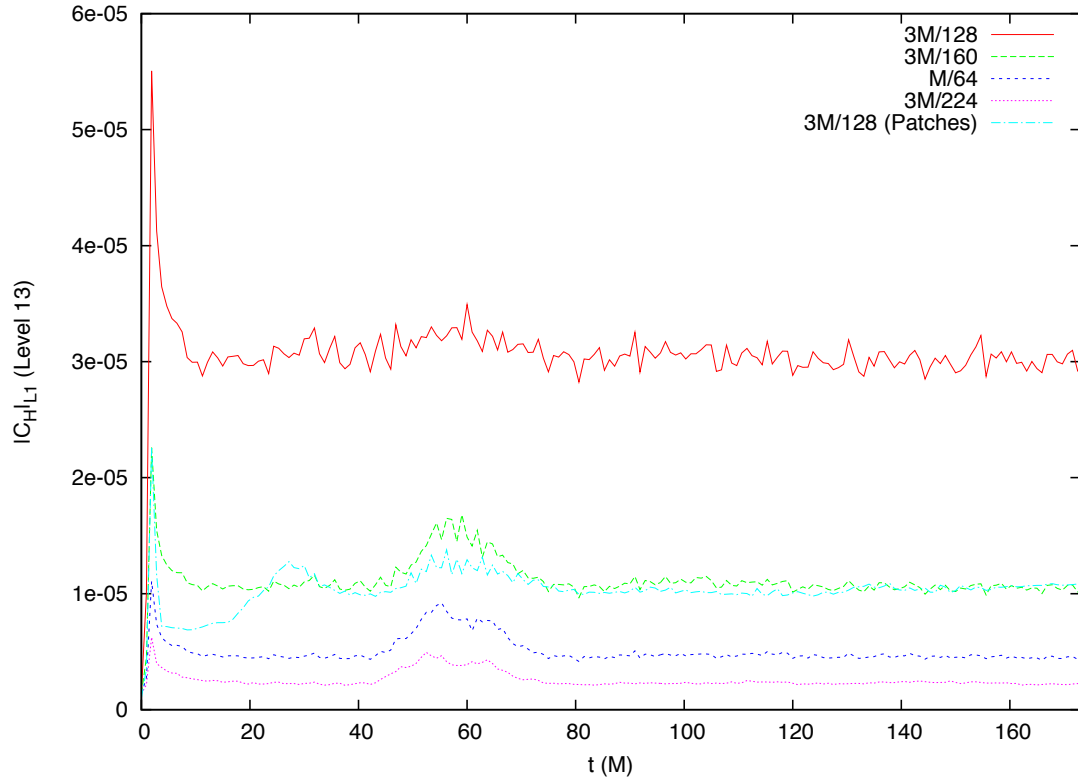
Figure 4.5: Hamiltonian constraint (L1 norm) for level 13 (second finest) in the inspiral test. The reduction is not as large as the reduction for the finest level in Fig. 4.3.
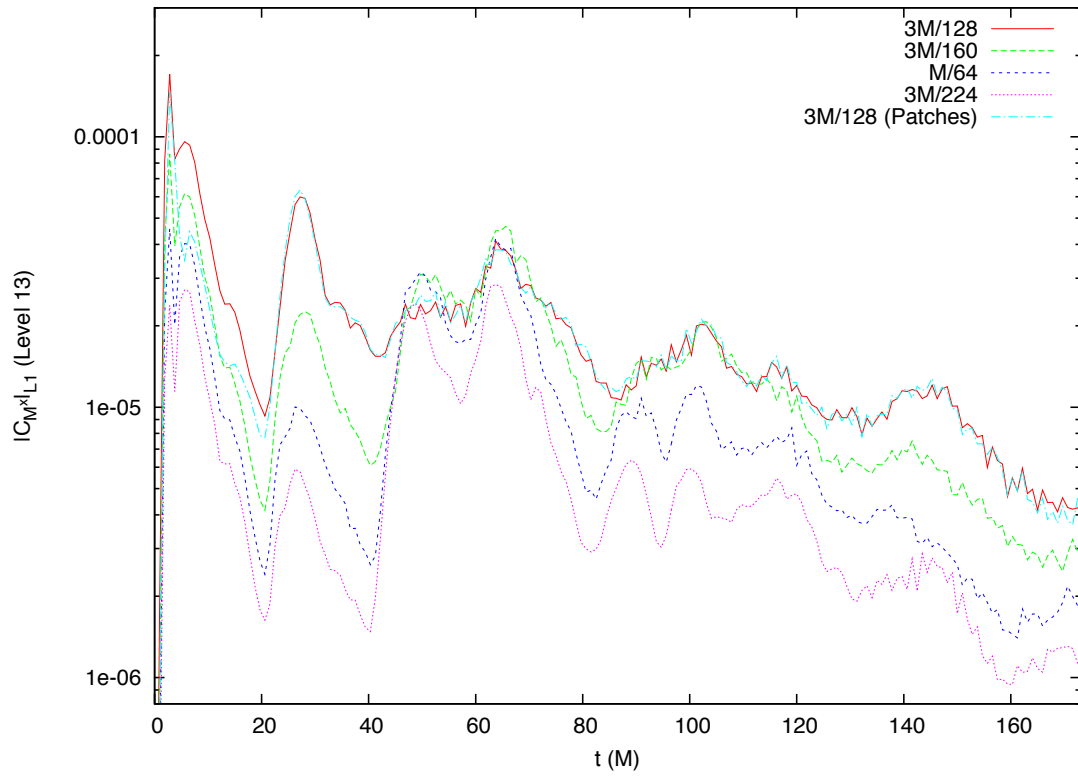
Figure 4.6: X momentum constraint (L1 norm) for level 13 (second finest) in the inspiral test. No reduction is evident.
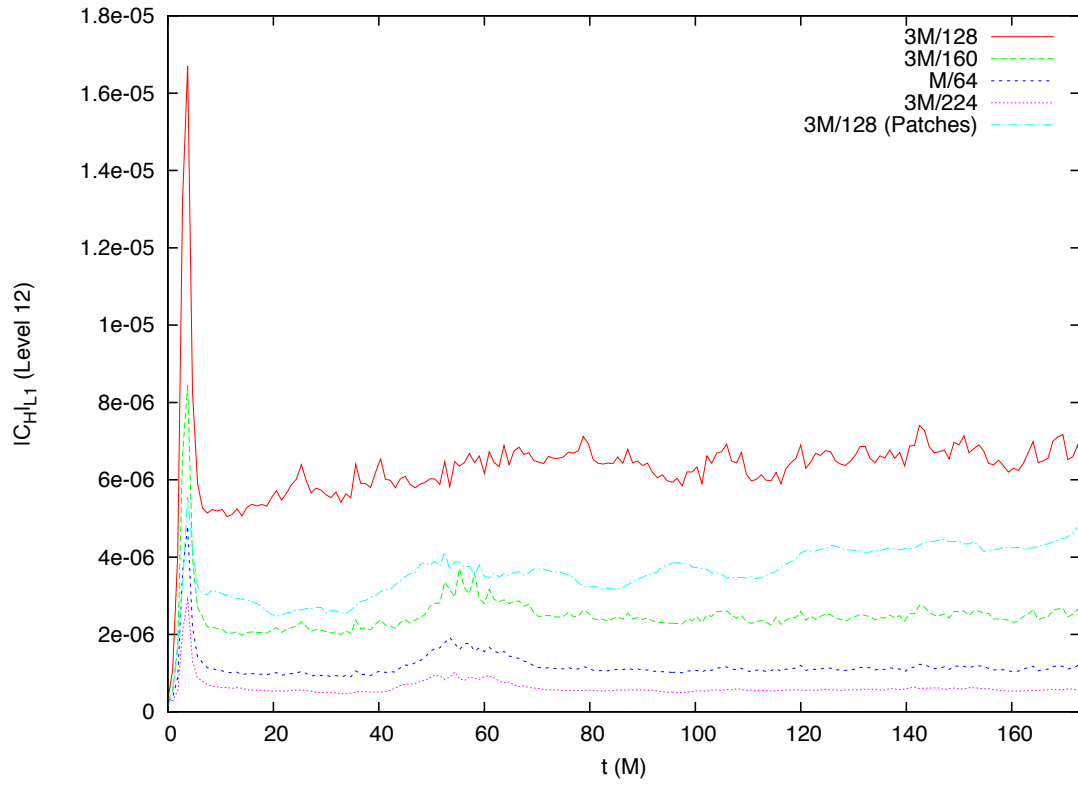
Figure 4.7: Hamiltonian constraint (L1 norm) for level 12 (coarsest patch level) in the inspiral test. Some reduction is still observed.

grid and which includes the wave zone. For both constraints, P_{128} is very similar to I_{128}. ($|C_{M^y}|_{L1}$ and $|C_{M^z}|_{L1}$ are similar to $|C_{M^x}|_{L1}$ again.) The patches do not overlap any part of level 9, so they affect this level indirectly. Therefore the similar constraint results for patch and non-patch runs are expected. Since radiation is extracted in the wave zone, these results suggest that the moving patches technique has not adversely affected the region in which the quantities of primary interest are computed. Combined with the constraint results for the higher levels, this indicates that with regard to the constraints, the moving patches technique offers increased accuracy approaching the punctures for a given resolution. This is consistent with the earlier heuristic decomposition of the field dynamics into the black holes' orbital motions and the gradual changes as angular momentum is lost and the black holes' approach. Moving away from the punctures, the orbital motions' have less impact on the fields, and therefore the moving patches technique offers less advantage.

As explained in Chapter 1, we conventionally describe the resolutions of simulations simply by referencing the finest resolution present in each, but when comparing simulations that differ in the sizes of the various refinement regions, the highest resolution is an insufficient description. Because P_{128} does not use AMR on the patches, while the non-patch inspiral runs do, it has a different grid structure. P_{128} maintains the FMR grid specified in the HAHNDOL parameter file, while the non-patch runs refine their grids to adjust to the fields. The highest refinement zones (levels 12-14) for the non-patch runs are smaller than those zones in P_{128}. Table 4.2 lists distance s from a puncture for regions of different maximum grid spacing maxΔx, for P_{128}, I_{128}, and I_{160}. s is not a radial distance, but rather the
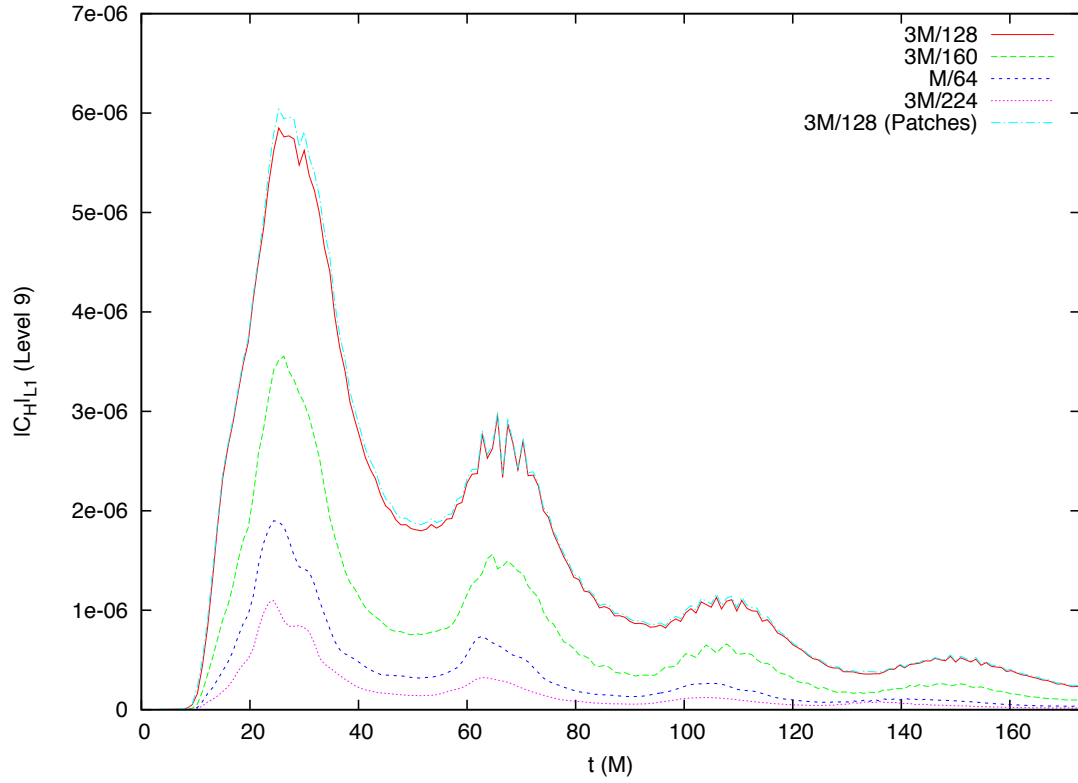
Figure 4.8: Hamiltonian constraint (L1 norm) for level 9 (wave zone) in the inspiral test. There is no significant impact from the interpolation on the patch boundary.
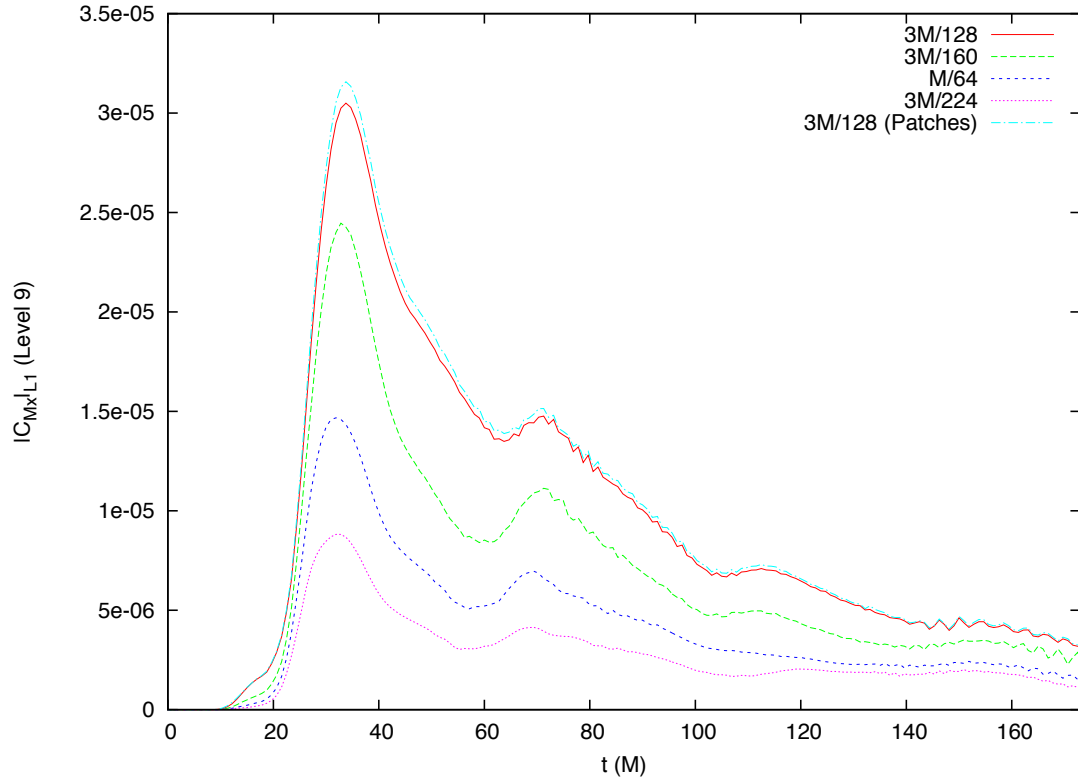
Figure 4.9: X momentum constraint (L1 norm) for level 9 (wave zone) in the inspiral test. There is no significant impact from the interpolation on the patch boundary.

Run	$\frac{3M}{32}$	$\frac{3M}{64}$	$\frac{3M}{128}$
P_{128}	3.23165	1.61583	0.80791
I_{128}	2.78132	1.55281	0.65175
	$\frac{3M}{40}$	$\frac{3M}{80}$	$\frac{3M}{160}$
I_{160}	2.78132	1.55281	0.65175

Table 4.2: Size of regions for maximum grid spacing. The first column is the run identifiers. The headings of the remaining columns are maximum grid spacings. The table lists the linear size in M units of the portion of the grid for which the grid spacing does not exceed the grid spacing listed in the column heading, for three inspiral runs. Larger sizes correspond to higher resolution grids.

half-side length of a cube of the same volume as the region, so s is the distance to the region's edge along a coordinate direction. P_{128} includes more of the grid than I_{128} for each $\max\Delta x$, so it has a higher effective resolution. Because I_{160} is higher resolution, each region has $\max\Delta x$ smaller by a factor of 0.8, while s is the same as for I_{128}. Fig. 4.10 makes the comparison between P_{128} and I_{160} more clear. It compares s as a function of $\max\Delta x$ for the three cases, and we see that P_{128} lies between I_{160} and I_{128}

The different grid structure of P_{128} and the resulting higher effective resolution that it produces complicates comparisons of the constraint violations norms with the non-patch runs. Some of the improvement may be due to the increased effective resolution of P_{128}. The constraint violations are normalized by volume, so differ-
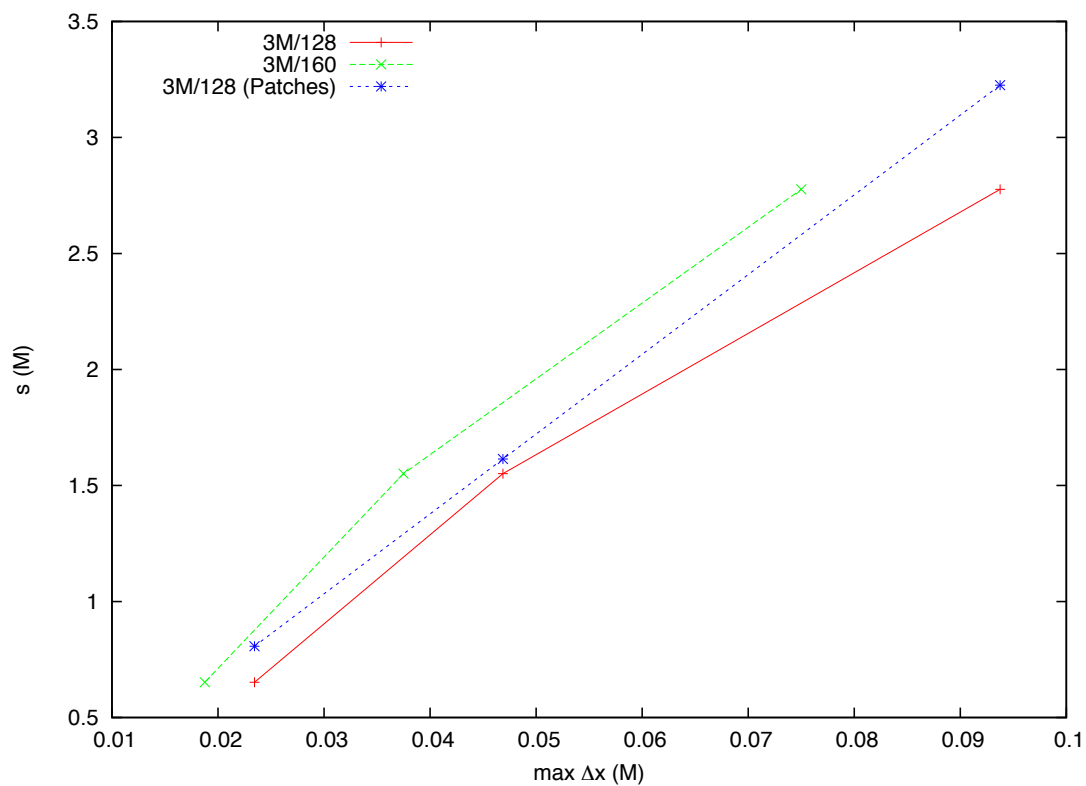
Figure 4.10: Maximum grid spacing Δx for distance s from punctures. s is a characteristic length of the volume for which the grid spacing does not exceed Δx. The larger size of the high-resolution portion of the patch run increases the effective resolution.

ences in size do not impact the norms directly, but farther from the punctures, the field gradients decrease, and so the finite differencing becomes more accurate. So including portions of the grid farther from the puncture should decrease the norms of any refinement level. Fig. 4.10 suggests that the effect of differing grid structures may make P_{128} comparable to a non-patch run with resolution between the I_{128} and I_{160} resolutions. The constraint accuracy of P_{128} surpasses that of I_{160} in the finest grid regions, an effect that cannot be ascribed to the grid structure.

4.3.2.2 Error in Phase and Frequency

To detect gravitational waves and to characterize the sources of the waves, accurate phase is crucial. Because of the low signal-to-noise ratio expected for gravitational waves, detector signals are matched to a template bank of calculated gravitational waves from different possible sources. Such matched filtering correlates the data to templates in the bank. A high correlation to a template indicates the presence of that waveform in the data. Therefore minimizing waveform phase error is crucial in black hole binary merger simulations, motivating special attention to the impact of patches on this type of error.

In a NR simulation, the phase, frequency, and amplitude of a given ℓm mode of the gravitational radiation vary by resolution. The different resolutions start with the same initial data, representing the same initial physical configuration, and the same initial gauge and gauge conditions. Variations in the physical and gauge fields result from solving for the initial data and evolving the fields, which are both

resolution-dependent. Quantities derived from the fields, such as the phase, frequency, and amplitude of the radiation, also vary by resolution. Comparing the results for different resolutions allows us to evaluate the quality of simulations and estimate the error. Though phase is important for gravitational wave detection, frequency provides a better indication of the error in a simulation at any point in time. The phase at a point in time is the result of the entire evolution up to that point. If resolution-dependent error sources instantly disappeared, the accumulated phase difference across resolutions would remain. Phase error accumulates. So when we compare two different resolutions, we cannot separate new phase error from existing phase error except by zeroing out the relative phase between two different resolutions and comparing the phases after the zero point. As the time derivative of phase, frequency better indicates the instantaneous state of the system. So comparing the frequency in time across resolutions better indicates the development of error in the system.

4.3.2.3   Frequency as a Measure of Error

Ultimately, we are interested in the phase and frequency of the gravitational radiation produced by the merger. To evaluate the moving patches technique, we want to compare the frequency error when patches are used to the error when patches are not used. Because of the performance issues in the implementation of the moving patches technique I was not able to run a patch simulation long enough to measure radiation. So as an alternative, I examined the orbital frequency calculated from

the puncture tracks, which is suggested by Fig. 2.1.

First, we must consider the relationship between this puncture frequency and the orbital frequency calculated from the radiation. Fig 4.11 shows the relative error in frequency as calculated from $\dot{h}_{22}$ and from the puncture tracks for I_{128}, I_{160}, and I_{192}. The relative error is defined as

$$\delta\omega_{src,res} = \left| 1 - \frac{\omega_{src,res}}{\omega_{src,224}} \right| . \tag{4.12}$$

$src = rad$ and PT for the radiation and puncture track frequencies, respectively, and $res = 128, 160$, and 192, indicates the resolution. I_{224} is the reference resolution. While this reference resolution does not represent the exact answer, it provides a simple error estimate. For a more precise error measure, I could perform a Richardson extrapolation, but as the goal is to demonstrate the similar behavior of the radiation and puncture frequencies, I have used the simpler error estimate.

In each resolution case, the relative errors of the radiation and puncture track frequencies track closely until the approach of merger and the formation of a common apparent horizon. At this point, we can no longer treat the system as though it were two separate bodies, and the puncture tracks no longer represent the system's dynamics well. The relative errors of the radiation frequencies show high frequency noise. I have used a logarithmic scale to compare the different resolutions better, but it does distort the noise. Fig. 4.12 shows a similar relative error measure earlier in the simulation. Here I define the relative error slightly differently,

$$\delta\omega_{src,res} = 1 - \frac{\omega_{src,res}}{\omega_{src,224}} . \tag{4.13}$$

I use this definition for Fig. 4.12 because early in the simulation, the noise in ω_{rad}

overshadows the trend, and the previous definition makes the trend unclear. This modulated high-frequency noise is routinely observed in HAHNDOL simulations. It appears to result from propagating initial data error that reflects off refinement boundaries in the simulation grid, and therefore it is separate from the error trend we are investigating. Again we see that the puncture frequency provides a good indication of the variation in the radiation frequency across resolutions until the approach of merger. The variation with resolution is established early. In particular, in both the puncture and the radiation frequencies, we can see I_{128} separating from I_{160} and I_{192} as early as $t = 150M$, a trend that continues up through merger.

Fig. 4.13 shows relative error in puncture frequency for I_{128}, I_{160}, I_{192}, and P_{128}, using Eq. (4.12) to define relative error and I_{224} as the reference frequency. P_{128} has a smaller relative error than both I_{128} and I_{160}. Though P_{128} ends at t=173M because of the performance limitations mentioned earlier, the improvement provided by the moving patches technique is already evident. Since the late frequency variation is established early, this improvement would continue through later times as well. The comparison of P_{128} with the non-patch runs is more straightforward for frequency than for the constraint violation norms because ω is determined by the shift values at the puncture locations only. Moderate variations ($< 20\%$) in the refinement boundaries far from the punctures (in terms of grid spacing) would have little influence on the punctures. In calculating puncture frequency, moving patches increases the effective resolution of P_{128} to at least $3M/160$.
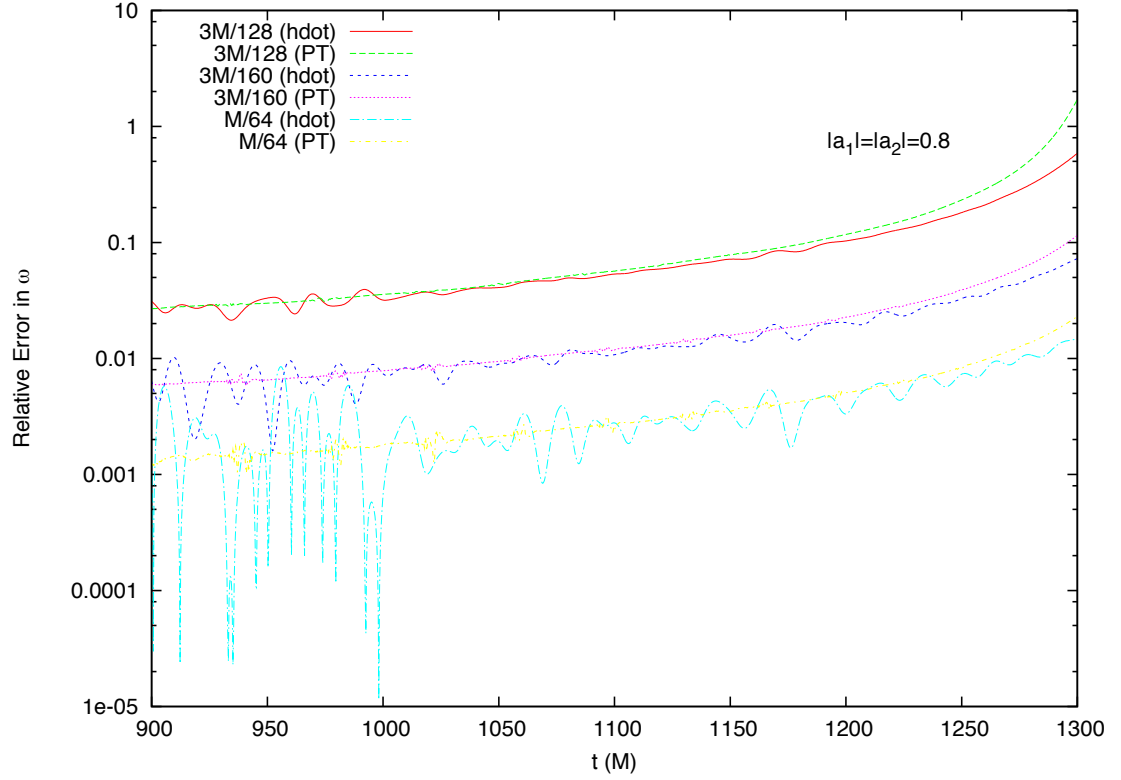
Figure 4.11: Relative frequency errors based on $\dot{h}_{22}$ and puncture tracks (non-patch). At each resolution, the relative errors for frequency, calculated with Eq. (4.12), from puncture tracks and $\dot{h}_{22}$ agree until approximately $100\,M$ before merger. I_{224} is the reference for the error calculations.
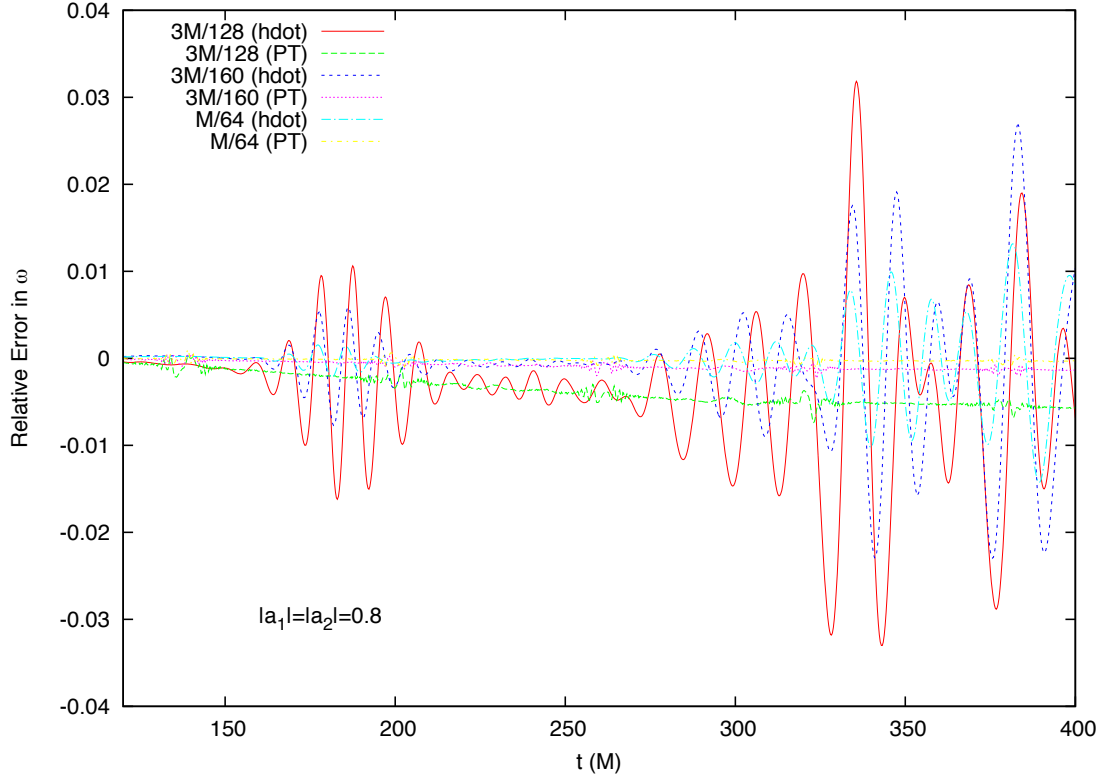
Figure 4.12: Relative frequency errors based on $\dot{h}_{22}$ and puncture tracks (non-patch, early). The relative frequency errors from puncture tracks and $\dot{h}_{22}$ agree at each resolution early in the simulations. Though qualitatively similar to Fig. 4.11, these early error comparisons use Eq. (4.13) to better show the trends for significantly smaller error magnitudes.
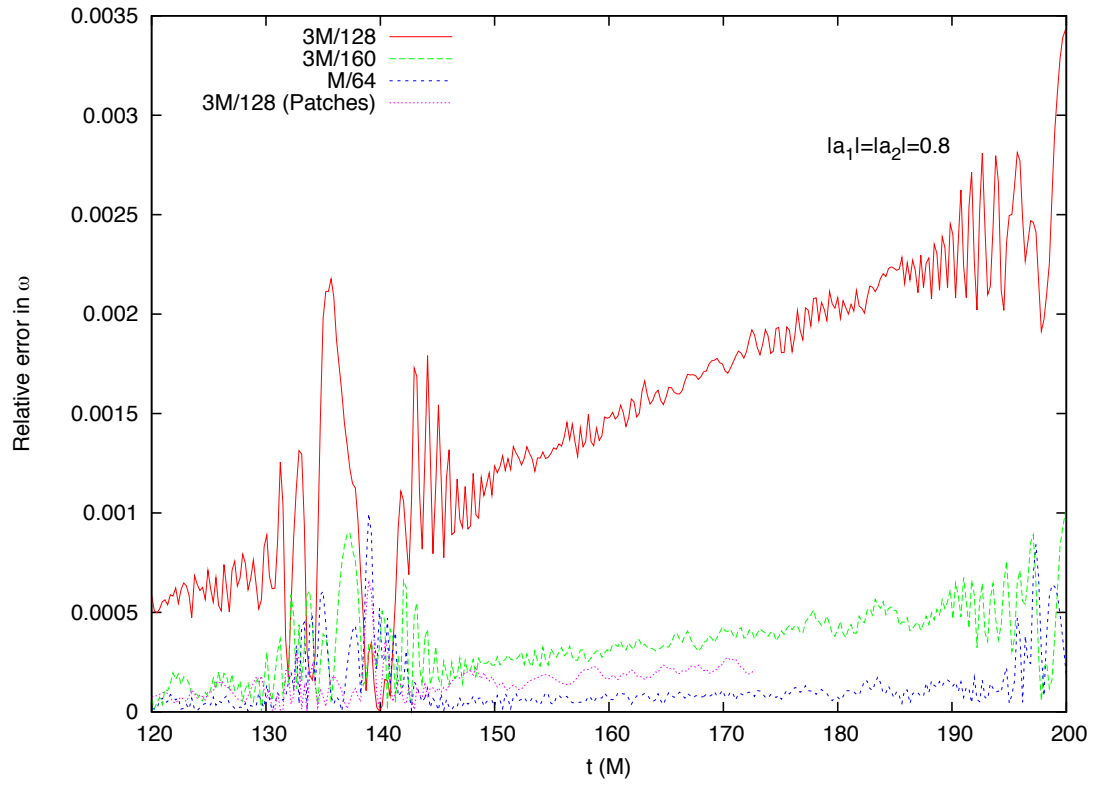
Figure 4.13: Relative frequency errors based on puncture tracks. The patch run errors are consistent with higher resolution non-patch runs.

4.3.2.4 A Comparison of Puncture Track Angular Momenta

The last section considered orbital frequency as a proxy for the secularly evolving state of the system, which was discussed previously and in more detail in [16, 17]. However, as we have seen, the orbital frequency is subject to noise and unintended eccentricity. Therefore, we return to the puncture track angular momentum L_{PT} that I developed in Chapter 1, since it may allow a more clear evaluation of the accuracy of the moving punctures technique.

Starting from Eq. (1.76) and adapting to the inspiral tests we are considering, we find that

$$\mathbf{L_{PT}} = \mathbf{L_N} + \frac{6.75}{32\mathbf{L_N}} . \tag{4.14}$$

I assume a total mass of $1M$, equal masses ($\eta = 0.25$), circular motion ($v = r^{-\frac{1}{2}}$), and a total spin of 0. With these assumptions, $L_N = 0.25\ell_{PT}$. I will use J for the actual angular momentum of the system to distinguish it from L_{PT}. Because the total spin of the system is 0, J is the same as the orbital angular momentum L.

Initially L_{PT} does not reflect the system's angular momentum. As an example, Fig. 4.14 shows that ℓ_{PT} for I_{224} starts at 0 and rapidly increases. Figs. 4.15 and 4.16 show the r^2 and ω factors of ℓ_{PT} separately, for the same run. r^2 has an initial modest adjustment, while ω has a more dramatic adjustment. r^2 is a function of the coordinates of the punctures while ω is a function of the puncture velocities, $\omega = r \left| \dot{x}_{p1}^i - \dot{x}_{p2}^i \right|$. The puncture velocities are 0 initially because $\dot{x}_p^i = -\beta^i(x_p^j)$, and initially the shift at the puncture locations does not correspond to the momenta of the punctures, but it evolves rapidly until the puncture velocities do reflect the
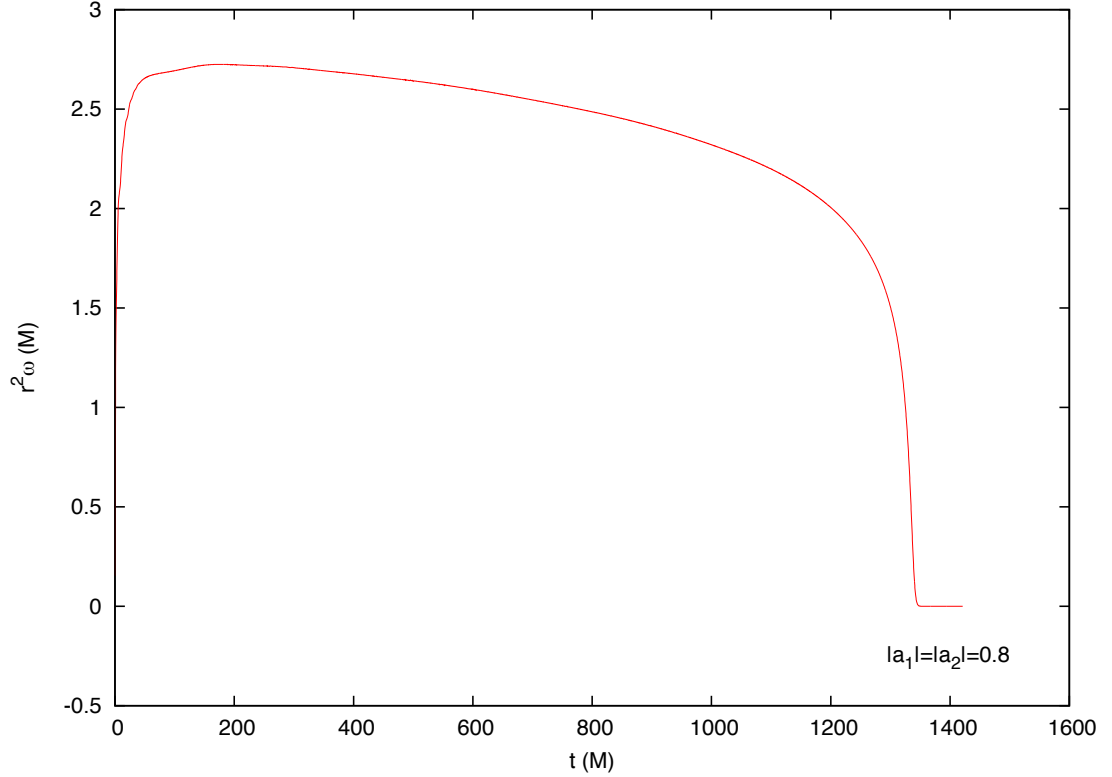
Figure 4.14: $r^2\omega$ for ultra-fine run (non-patch). After a brief initial adjustment period, $r^2\omega$ demonstrates physical behavior.

punctures' momenta. In the case of I_{224}, this occurs at approximately $t = 60M$, as shown by Fig. 4.17. At this point, ℓ_{PT} and therefore L_{PT} become relevant for examining the angular momentum in the simulation.

Having defined L_{PT} and discussed its early behavior, I now establish its usefulness as a diagnostic for angular momentum evolution. Fig. 4.18 compares J to L_{PT} for the two highest resolutions in the non-patch case. J is calculated by subtracting the radiated angular momentum from the initial angular momentum as calculated

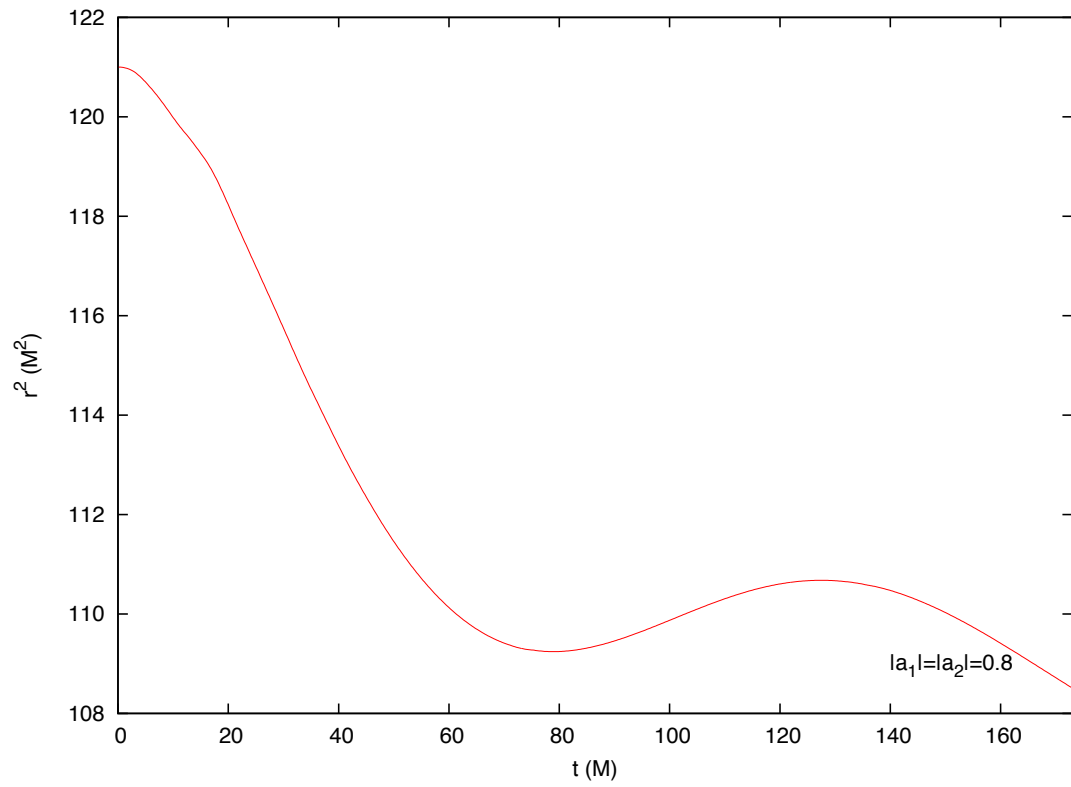Figure 4.15: r^2 for ultra-fine run (non-patch). r^2 has a slight initial adjustment.

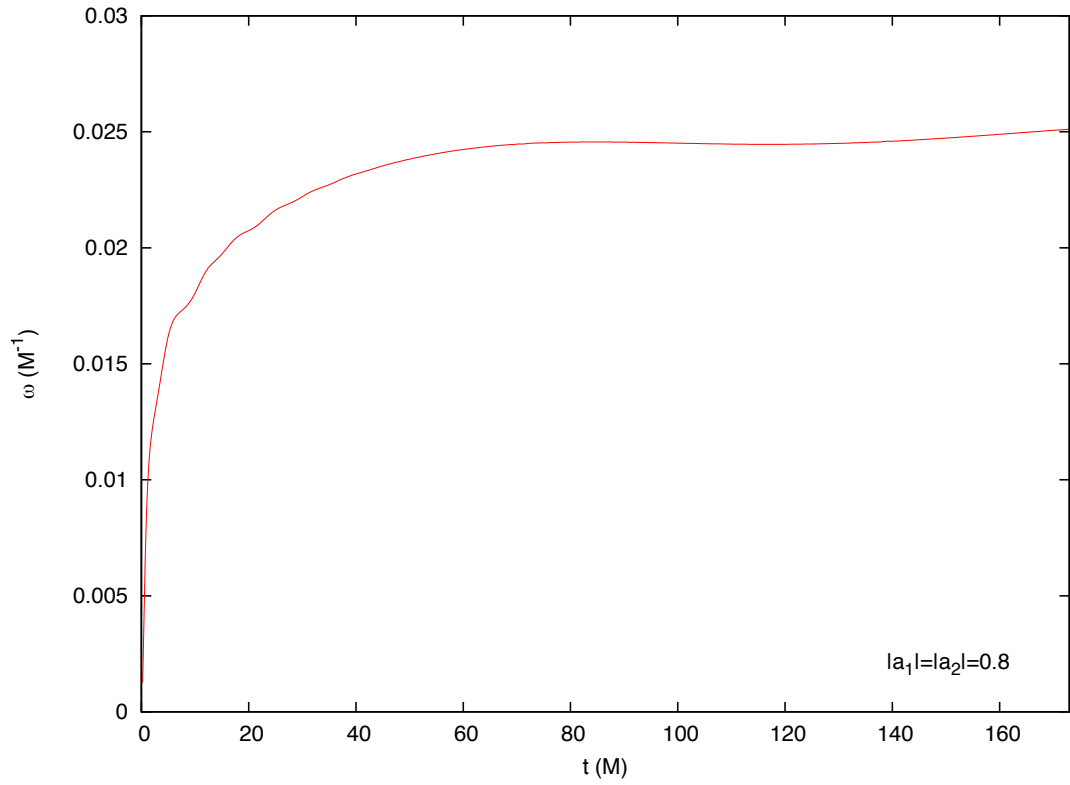Figure 4.16: ω for ultra-fine run (non-patch). Because ω is calculated from the puncture tracks, it undergoes a dramatic adjustment initially.
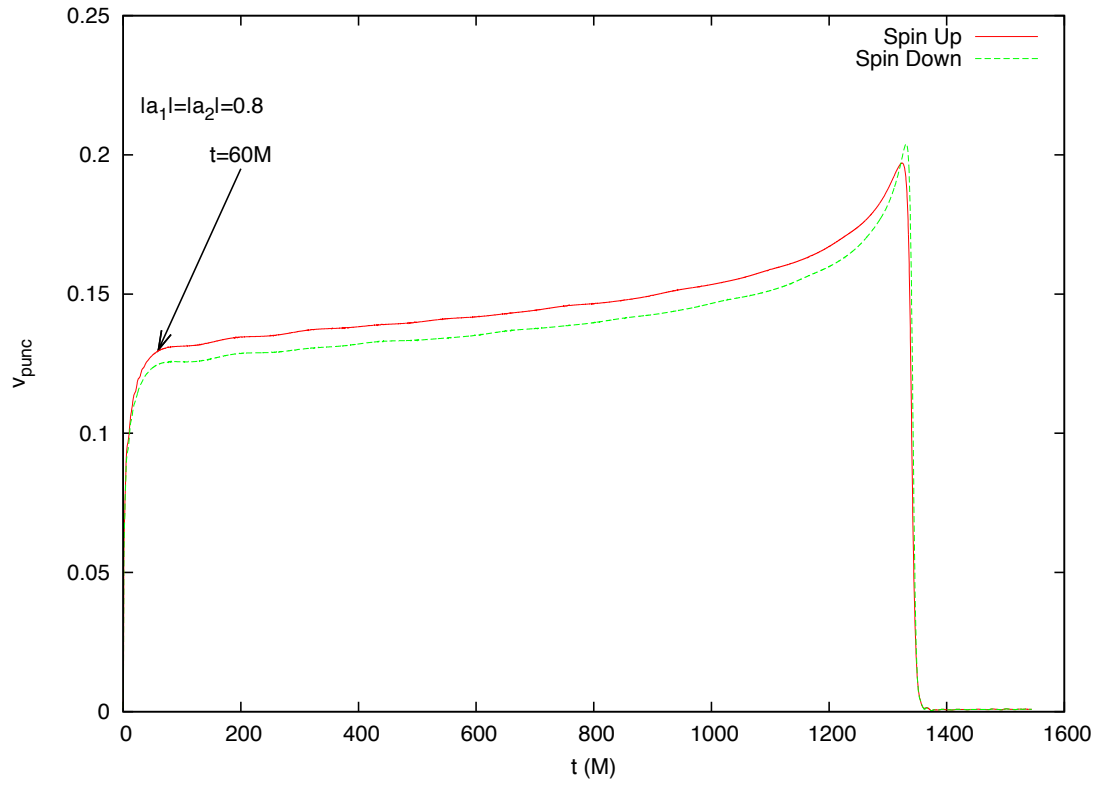
Figure 4.17: Puncture velocities for ultra-fine run (non-patch). In the first $60\,M$ the puncture velocities quickly adjust to be consistent with the Bowen-York momenta.

from the initial data. For both resolutions, L_{PT} follows J closely and therefore provides a good early indication of the system's angular momentum evolution. L_{PT} deviates from J significantly close to merger, but the variation due to resolution is apparent by this time. Fig. 4.19 shows the relative error in L_{PT} for I_{128}, I_{160}, and I_{192}, using I_{224} as the fiducial value. The periodic bursts of noise in L_{PT} coincide with one of the punctures crossing a coordinate axis. Similar bursts of noise in the puncture frequency as either puncture crosses an axis are routinely observed in HAHNDOL simulations. There is no reason crossing an axis would generate noise directly. However when a puncture crosses a coordinate axis, it is moving parallel to the other coordinate axis. (Recall that the simulations were performed in bitant symmetry, so the punctures remain in the $z = 0$ plane.) The motion parallel to a coordinate axis is a more likely source of the noise, but I have not investigated it further. I mention it only to eliminate any confusion it might cause.

With L_{PT} as a diagnostic, we can compare the patch run to non-patch runs. Fig. 4.20 compares L_{PT} for P_{128} and all the non-patch runs. P_{128} matches I_{224} better than I_{128} and I_{160} and as well as I_{192}. Fig. 4.21 shows this more clearly. It compares the relative error of P_{128}, I_{128}, I_{160}, and I_{192}, using I_{224} as the fiducial value. Based on L_{PT}, the patches technique at resolution $M/128$ performs at an accuracy comparable to I_{192}, supporting similar conclusions based on puncture-track frequency in Sec. 4.3.2.3. As in the case of puncture frequency, L_{PT} is measured far from the refinement boundaries, and so the comparison is straightforward.
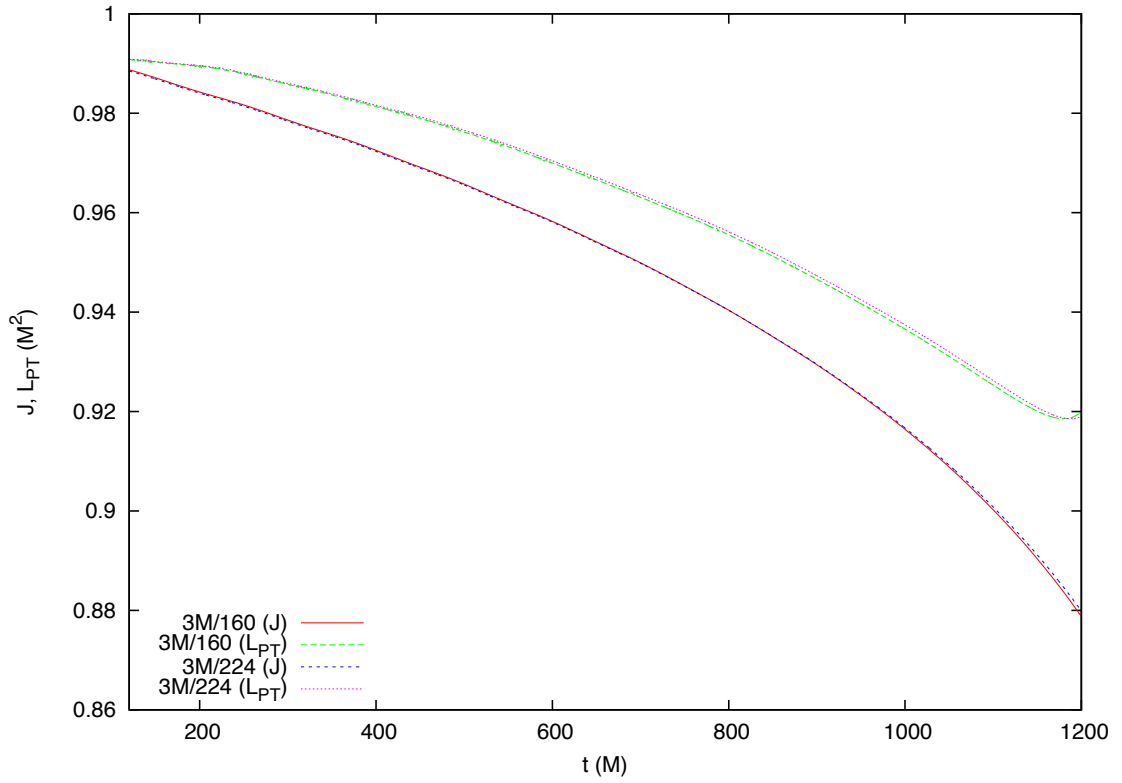
Figure 4.18: Angular momentum for runs (non-patch). L_{PT} and $J = J_{initial} - J_{radiated}$ agree from early in the simulations until near merger.
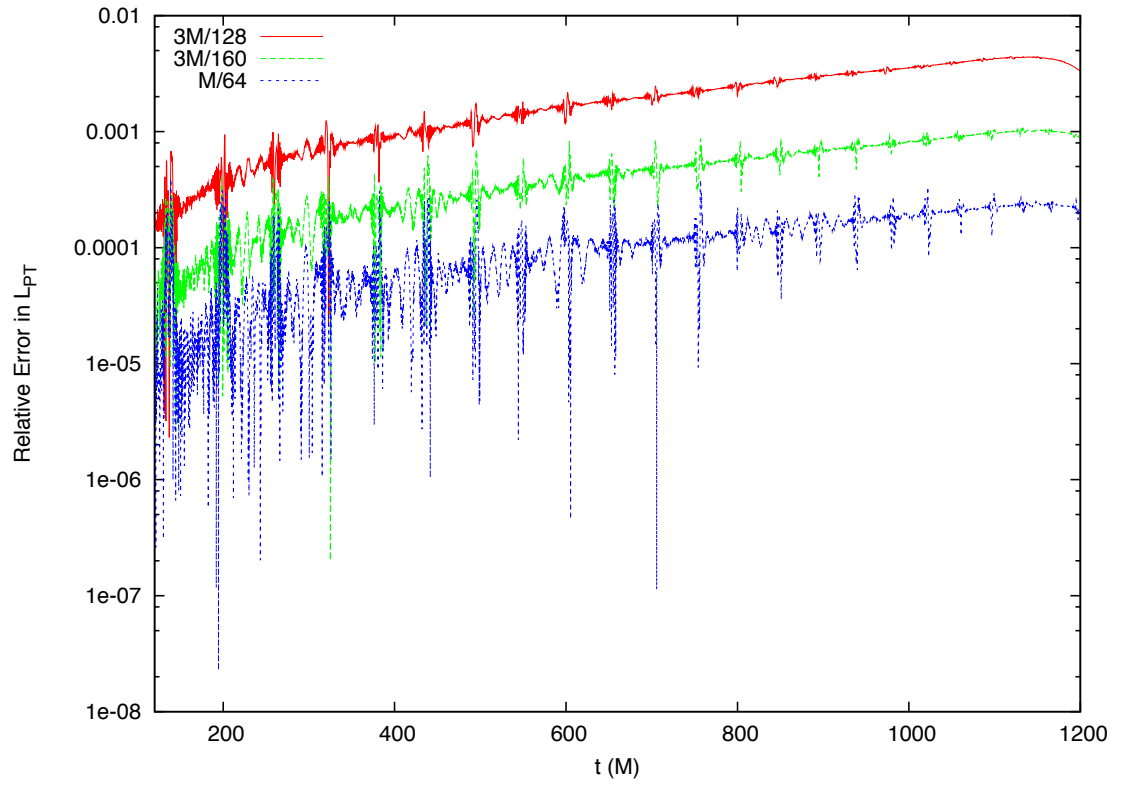
Figure 4.19: Relative error in L_{PT} (non-patch). Relative errors in I_{128}, I_{160}, and I_{192} I_{224} are compared to I_{224}. The periodic bursts of noise are associated with crossing coordinate axes.
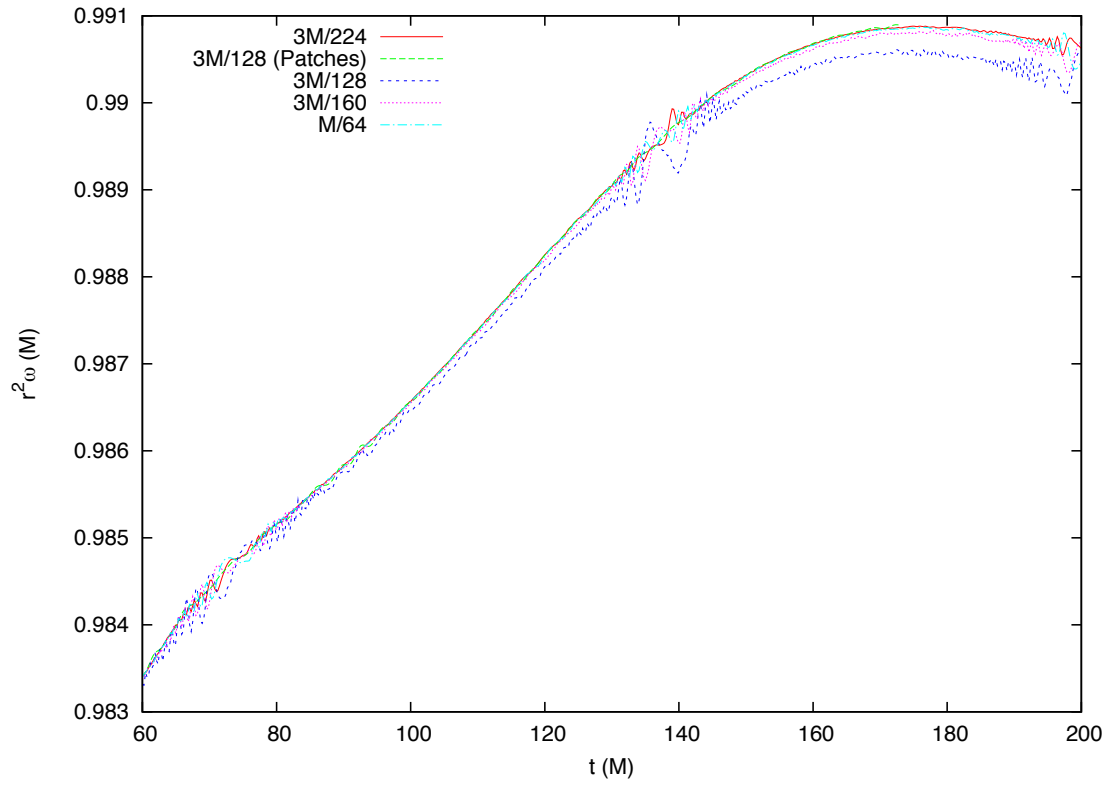
Figure 4.20: $r^2\omega$ for non-patch and patch runs. P_{128} is similar to I_{192} $(M/64)$ in its agreement with I_{224}.

Figure 4.21: Relative error in $r^2\omega$ for non-patch and patch runs. As in Fig. 4.20, P_{128} agrees with I_{224} at the same level as I_{192} ($M/64$).

4.3.3 Performance

The moving patches technique as implemented in HAHNDOL is very time-consuming. To assess the performance, I performed three runs with moving patches using the same parameters as P_{128} and varying the number of processors for each run. In addition I performed one run without moving patches with the same parameters on 128 processors. Each run lasted approximately two hours. The runs are summarize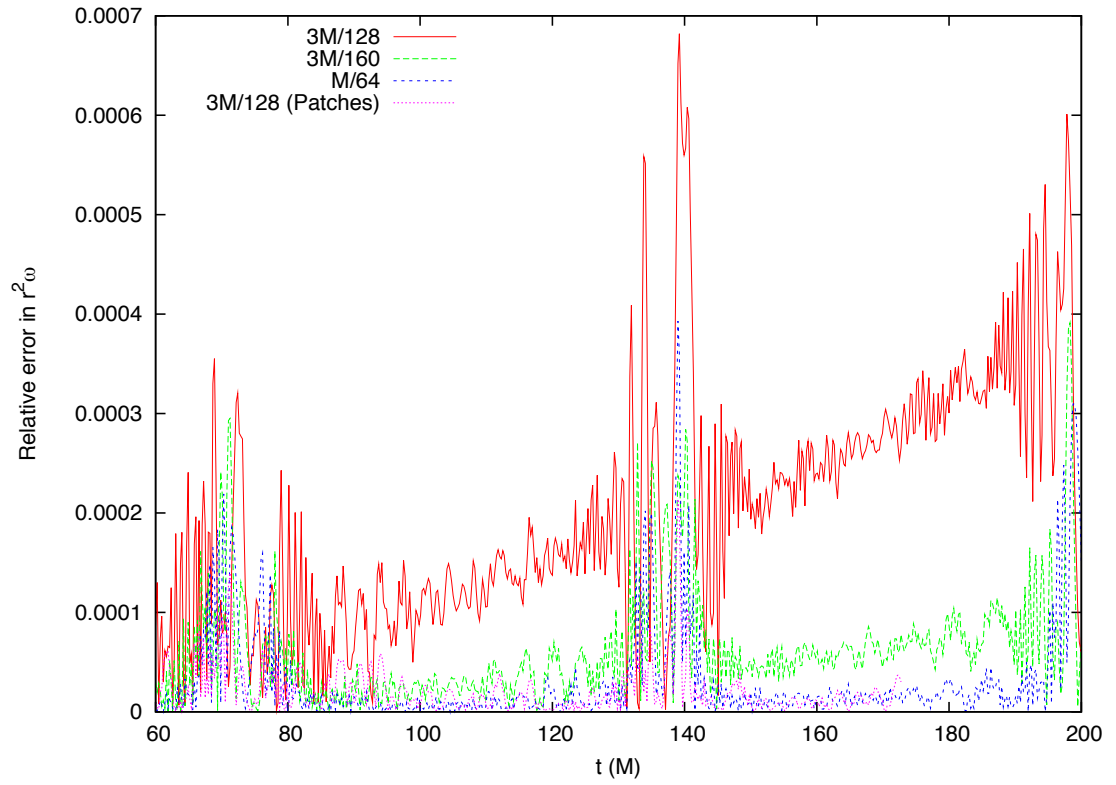d in Table 4.3.3. The impact of the moving patches is clear. Comparing the runs with 128 processors, the non-patch run evolves to $t = 11.7M$ in 117 minutes, while the patch run evolves to $t = 1.50M$ in 121 minutes, for a performance ratio of 7.8. The patch run evolves 7.8 times more slowly. Also, note the scaling behavior of the three patch runs. When the processor count is increased to 256 processors, the simulation reaches $t = 2.07$, a 33% increase for a doubling in processing power. A further doubling in processor count to 512 produces no increase in the time the simulation reaches.

The significant performance reduction results from the synchronization between the patches and fixed grid, in particular the interpolation from the patches back to the fixed grid. In the inspiral runs, the patches cover approximately 70 blocks of the fixed grid. In the P_{128} case, each block contains 3375 grid points including guard cells, so HAHNDOL interpolates approximately 240,000 points, four times per iteration, for 24 fields, which results in approximately 23 million interpolations per iteration. Interpolation from the fixed grid to the guard cells on the patch outer boundaries is quick in comparison. The patches contain 560 blocks,

Processor Count	Iterations	Final t (M)	Simulation Duration (min)
128	128	1.50	121
256	177	2.07	121
512	177	2.07	120
128 (NP)	1000	11.7	117

Table 4.3: Performance of moving patches. "NP" indicates the run without moving patches. The table shows the results of performance tests of moving patches with the results of a run that did not use moving patches. Moving patches has a significant performance cost. The columns are: the number of processors used for the test, the number of simulation iterations, the final coordinate time in the test, and the time required to run the test.

with approximately 43 block faces on each of the two outer boundaries and 675 guard cells on each face, which results in approximately 5.5 million interpolations per iteration.

Though these estimates are specific to P_{128}, they demonstrate the considerable computational cost of the synchronization. This is due to the HAHNDOL interpolator, which relies on interprocessor communications. These communications are slow compared to operations local to a processor, and all of the processors call the interpolator for the same set of points simultaneously. Each processor searches the local grid for the points and broadcasts the interpolated data to all processors, if it finds the points. Because this synchronous interpolation uses broadcast communications, it requires little coordination between processors, but it is inefficient, since it is a serial process. For the small number of points for which the interpolator was designed originally, the inefficiency is insignificant. For the patch synchronization, which requires on the order of a million points interpolated per iteration, the inefficient process produces a significant delay.[2]

Despite this limitation, I made the patch synchronization more efficient via several improvements to HAHNDOL. First, I modified the interpolator so that multiple points are interpolated by a single function call, which reduces the interprocessor

---

[2]Using communications between pairs of processors would be much faster since the interpolated points are distributed across the processors, and multiple synchronizations could occur simultaneously. However, such a parallel process requires careful coordination of the pairwise communications to prevent deadlocks where processors wait indefinitely for responses. Therefore the HAHNDOL interpolator does not use this approach.

communications. Next, I reordered the calculations, so that multiple fields are interpolated at each point in the same call, and eliminated redundant calculations. Finally, I rewrote the patch synchronization procedure to use the HAHNDOL interpolator more efficiently. Previously, the processors in turn requested points for interpolation. During each call of the interpolator, most processors were neither sending nor receiving useful data. In the modified procedure, each processor creates a local list of points for which it requires data. Where possible, it fulfills the request locally. Next, all of the processors broadcast their local point lists. As a result, each processor has the list of points needed by all of the processors. Then, the processor creates a "chunk" of points, a list comprising points from multiple processors, and calls the interpolator with the chunk. Many more processors send and receive useful data for each interpolator call. The chunk and interpolate sequence repeats until the global point list is exhausted. For each iteration in the synchronization process, the chunk comprises equal numbers of points from each processor. The efficiency gain is limited by the chunk size, which in turn is determined by memory limits on the simulation host. These changes improve the efficiency of the synchronization substantially, but the synchronization process remains a serial process with limited efficiency. Therefore, this part of the technique still severely limits the duration of simulations using moving patches in HAHNDOL.

The other components of the moving patches technique do not contribute significantly to the time required to run a simulation. The creation of the patch grids and copy of initial data to the patch grids only occur once. The modifications to the evolution equations require additional calculations at each point on the patches, but

148

they do not require additional interprocessor communications, and the number of additional calculations are not significant compared to the number of calculations required by the modified evolution equations. Updating the coordinate transformations has an insignificant impact on performance since it uses data local to the processors and involves simple calculations.

4.4 Conclusions

4.4.1 Results

The moving patches technique does reduce errors in simulations as shown by the constraint violations and errors in orbital frequency and L_{PT}. Synchronization between the patches and the fixed grid introduces interpolation error, but the head-on runs show that it does not lower the convergence of the constraint violations at the patch boundary. In the inspiral runs, the reduction in the constraint violation norms is larger for refinement zones closer to the punctures. This is consistent with expectations since the field gradients are larger closer to the punctures. The patch run has larger refinement zones and therefore higher effective resolution. This is partially responsible for the error reduction, and it is unclear how much the moving patches technique reduces the constraint violations.

The patch run has a finest resolution of $3M/128$, but it performs at higher effective resolutions for calculating frequency and L_{PT} from the puncture tracks. Error reduction for frequency and angular momentum is more clearly due to the moving patches technique because the puncture tracks are not strongly affected by

the specifics of the grid refinement.

In the case of the frequency, the patch run has a lower relative error than the non-patch run that has a finest resolution of $3M/160$, and therefore the patch run has an effective resolution of at least $3M/160$ by this measure. The behavior of the frequency derived from the puncture tracks is very similar to the frequency of the gravitational radiation, especially in comparisons of different resolutions. So the error reduction in the puncture frequency in the patch run indicates a similar error reduction would occur for the radiation frequency in a longer patch run.

For L_{PT}, which is an approximation of J for the system, the patch run performs at an effective resolution of at least $M/64$ since it has a lower relative error than the non-patch run at $M/64$. As in the case of frequency, L_{PT} behaves very similarly to the radiation-derived J for the system, and this indicates that a longer patch run would demonstrate a more accurate angular momentum for the system.

L_{PT} could be used as a diagnostic quantity in other simulations giving an early indication of error related to angular momentum. It is less affected by eccentricity facilitating comparisons of different runs. Also, since it is derived from the puncture tracks, it gives a much earlier indication of error than radiation-derived quantities such as J that use information from the wave zone which are delayed by the propagation time from the strong field region to the wave zone and by the time it takes the radiation amplitudes to increase sufficiently to extract accurate information. L_{PT} using patches is also less noisy since it is not subject to the noise associated with punctures crossing coordinate axes.

4.4.2 Future Development

The most important development for the moving patches technique is increasing the efficiency of it. The current implementation of moving patches in HAHNDOL is very inefficient, and it scales poorly. So it cannot evolve interesting simulations.

The poor performance is due to the synchronization between the fixed grid and the patches. The synchronization uses the HAHNDOL interpolator which uses broadcast communications between processors. Such broadcast communications are slow compared to operations local to a processor. The regular PARAMESH guard cell filling routines fill many more points in a shorter time using point-to-point communications which allow data exchanges between multiple processor pairs simultaneously. Using such point-to-point communications for the interpolations in the synchronization is a promising route for improving the efficiency of moving patches.

Another route for improving the efficiency is to only synchronize a buffer region of the fixed grid at the patch boundary. This would decrease the number of fixed grid points interpolated during the synchronization. The remaining fixed grid region covered by the patches would not need to be evolved.

A less promising route is to integrate the patches with PARAMESH. PARAMESH aggregates contiguous blocks of data on the same processor, which decreases the need for interprocessor communications. The patches are implemented in HAHNDOL above PARAMESH, so PARAMESH distributes the blocks from the patches, according to the local coordinates, on a small number of processors. If PARAMESH utilized the patch coordinate transformations, the patch blocks could be placed on the same

processor as corresponding fixed grid blocks, which would allow much faster synchronization. Because this would require significant modifications of PARAMESH, it is not a promising route, but it should be taken into account for future implementations of the moving patches technique with another grid system, such as CARPET in the CACTUS framework.

In addition to efficiency gains, the moving patches technique can be fully implemented by

- Allowing AMR on the patches

- Integrating patches with other HAHNDOL tools such as the Apparent Horizon finder

- Adding a mechanism to "peel off" patch levels as the patches approach each other

- Tuning the coordinate transformation to make it smoother

The first two items are related to the infrastructure of HAHNDOL and PARAMESH, the mechanism for grid refinement and how coordinates are treated. PARAMESH handles grid refinement at the block level. When it refines a block, PARAMESH changes it into a *parent* block and creates eight child (*leaf*) blocks that cover the same portion of the grid as the parent block at twice the resolution. The parent block continues to be updated with data from the child blocks. When PARAMESH derefines blocks, it deactivates the eight child blocks of a common parent block and changes the parent block back into a leaf block. If the derefined blocks are on the minimal refinement

level of a patch, the parent block no longer has valid data since PARAMESH cannot use the coordinate transformation information that exists at the HAHNDOL level. The minimum patch level can be treated as the minimum refinement level for AMR on the patch grid, but this constrains AMR and leads to poor grid refinement. Also, currently the patch boundaries are specified as static coordinate ranges because of the complexity of describing a more generic boundary consisting of the outer faces of the blocks in the minimum patch level.

If AMR ran on the patches, the patch boundaries no longer could be described as simple coordinate ranges, and a new boundary specification would be required. To integrate moving patches with other HAHNDOL tools, all of HAHNDOL would need to be capable of using the patch coordinate transformations. Currently parts of HAHNDOL use the PARAMESH block coordinate information explicitly, such as to calculate the distance of a grid point to a puncture. All such procedures would need to make the appropriate transformations to the coordinate information provided by PARAMESH. Though straightforward, this process must be done for all such procedures.

The second two items are directly related to the moving patches technique. As the punctures inspiral in a black hole binary simulation using moving patches, the moving patches will eventually overlap. While in principle this is acceptable since they would interact through their synchronizations with the fixed grid, it would introduce additional error because the exchange between patches would result from interpolations to and from the fixed grid. Also, the synchronizations would need to be carefully sequenced.

A better alternative is to eliminate the minimum refinement level on each patch and add that level to the fixed grid. This "peeling off" of the patch level would consist of adding the level to the fixed grid using AMR, interpolating to the new maximum fixed grid level from the patches, and eliminating the level from the patches. The patch refinement levels would be peeled off in sequence as the separation of the patch outer boundaries subsequently approached some minimum, such as $\frac{\sqrt{3}}{2}$ times the block size. A similar procedure would work for patch implementations on other platforms such as CACTUS/CARPET.

The coordinate transformation in the current implementation of moving patches produced smooth results in the tests. The patch velocities are adjusted at each iteration, which prevents large jumps in the velocities or the coordinate transformations derived from them. Nonetheless, smoother velocities would likely benefit both the evolution on the patches, which uses the patch velocities in the additional advection terms, and the synchronization of data between the patches and the fixed grid, which uses the derived coordinate transformations. As suggested previously, a feedback control system similar that used for dual coordinates [94] seems promising. In the case of moving patches, the control parameter for such as system would be the patch velocity $\dot{q}_A^i$ itself, and the condition would be $\dot{q}_A^i \to 0$. This would be achieved by adjusting $\partial^2 \dot{q}_A^i / \partial_t^2$.

4.4.3 Future Applications

In addition to decreasing error associated with black hole advection in simulations, the moving patches technique would allow larger time steps when used with an implicit time integration scheme. Currently HAHNDOL and other NR codes use explicit time integration schemes,[3] such as RK4. In such schemes, the Courant–Friedrichs–Lewy (CFL) condition,

$$\Delta t \leq \Delta t_C = C\Delta x \, , \tag{4.15}$$

limits the time step Δt to a multiple of the grid resolution Δx. The multiplier C is a constant determined by the system of equations and the integration scheme. For BSSNOK simulations using HAHNDOL and RK4, $0.5 < C < 1.0$. If $\Delta t > \Delta t_C$, the system solution will not converge in time. The CFL condition reflects the requirement that the domain of dependence of a given grid point (the portion of the simulation domain that affects the point) from the finite difference stencils include the analytical domain of dependence from the differential equations. The analytical domain of dependence is determined by the fastest characteristic modes of the system. The CFL condition requires that the time step be small enough that modes outside the finite difference domain of dependence of a grid point cannot reach the point. For BSSNOK using 1+log slicing, the fastest modes can be non-physical gauge waves [5], in which case Δt_C is determined by non-physical dynamics.

---

[3]Lau *et al.*[65] demonstrated a implicit-explicit (IMEX) integrator for simulations of a scalar field on curved backgrounds, but at this time implicit time integration is not used for simulations of black hole binary mergers in full general relativity.

The fields' time variation also bounds the time step by some Δt_D that characterizes the fastest *physical* dynamics of the system. If $\Delta t > \Delta t_D$, the simulation will not accurately resolve these dynamics. To get a rough estimate of this bound, consider a moving punctures inspiral simulation, like the one used to test moving patches. For a field u_a, the relative change in one time step is approximately

$$\left|\frac{\Delta u_a}{u_a}\right| = \Delta t \left|\frac{v_D^i \, \partial_i u_a}{u_a}\right| \tag{4.16}$$

$$= \Delta t \left|v_D^i \, \partial_i \log u_a\right| . \tag{4.17}$$

For accuracy, we can impose the condition that $|\Delta u_a / u_a| < \epsilon$. Then solving for Δt, we have that

$$\Delta t < \Delta t_D \equiv \frac{\epsilon}{|v_D^i \, \partial_i \log u_a|} . \tag{4.18}$$

We can estimate $u_a \sim r_p^n$ where r_p is the distance to the nearest puncture, and $|n| \leq 3$ based on [54] where the strongest radial dependence is r^{-3} for $\tilde{A}_{ij}$. Set $n \equiv 3$ for definiteness, assuming the strongest radial dependence and discarding the negative sign because of the absolute value function present in Eq. (4.18). Also, express r_p as a multiple k of the grid spacing Δx. We then find

$$\Delta t_D = \frac{\epsilon \, k \, \Delta x}{3 v_D} . \tag{4.19}$$

Fig. 4.17 shows that the puncture velocities are less than 0.15 for the first two-thirds of I_{224}. So take $v_D = 0.15$ to be a representative velocity for the dynamics of the inspiral. Setting $\epsilon = 0.01$ for a conservative condition that the field changes no more than 1% in a timestep and substituting for v_D, we have

$$\Delta t_D = \frac{k \, \Delta x}{45} . \tag{4.20}$$

156

For I_{224}, the grid spacing is approximately 75 points/M. The apparent horizon is approximately $0.5M$ from the puncture in binary simulations, which corresponds to $k \sim 38$ above. So at the apparent horizon $\Delta t_D \sim \Delta x \sim \Delta t_C$. Because the apparent horizon lies inside the event horizon, and physical influences cannot escape the event horizon, this gives us a good estimate of Δt_D.

Implicit time integration schemes are not subject to the CFL condition, but they are still limited by the dynamics of the simulation. They are also computationally expensive, so even if Δt_D were slightly larger than Δt_C, an implicit scheme would not increase the speed of the simulation since the slightly larger time steps would take considerably longer. If we could slow down the dynamics considerably, we could use an implicit scheme with much larger time steps to speed up our simulations. Moving patches offers much slower dynamics, so we could use much larger time steps on the patches. We can see from Eq. (4.19), that decreasing v_D leads to an arbitrarily large Δt_D. Taking $\epsilon = 0.01$ as before and setting $v_D = 0.001$ to estimate the very small puncture velocity on a moving patch, we see that $\Delta t_D \sim 3k\Delta x$. For points well within the apparent horizon ($k \sim 33$), $\Delta t_D \sim 100\Delta x$, which would result in much faster simulations, even with the computational expense of implicit time integration. Such a combined strategy could be used on the moving patches while explicit time integration was used on the fixed grid. Since the fixed grid would have considerably larger grid spacing than the finest portion of the moving patches, it could use proportionally larger time steps.

The estimates above show that the moving patches technique could allow much faster simulations if combined with implicit time integration. Though they are

rough, they demonstrate the dependence of the time steps on the *physical* dynamics, and the enormous efficiency gain possible when black hole motion is minimized. Combined with the accuracy gains in frequency and angular momentum demonstrated in the inspiral tests, they show the great potential this technique has for improving black hole simulations.

Appendix: Mathematical Details

A.1   Spatial Analogs

The Einstein tensor G_{ab} appearing in Einstein's equations is defined as

$$G_{ab} \equiv R_{ab} - \tfrac{1}{2}R\,g_{ab}\,, \tag{21}$$

where R_{ab} is the Ricci tensor and R is the Ricci scalar.

In the ""3+1" decomposition of spacetime, these and other four-dimensional tensors have spatial analogs on the three-dimensional spatial slices. The spatial tensors have the same definitions and expressions as the spacetime tensors with g_{ab} replacing g_{ab}. Below I give explicit expressions for these quantities.

The spatial metric g_{ab} defines distances on a spatial slice. It has an inverse g^{ab}, defined by $g_{ab}g^{bc} = \delta_a{}^c$. g_{ab} lowers contravariant indices on spatial tensors, and g^{ab} raises covariant indices on spatial tensors. Because the slice is three-dimensional, the trace of g_{ab} is 3.

Given g_{ab}, there is a unique derivative D_a that satisfies the condition $D_a g_{bc} = 0$. This is the covariant derivative for the spatial metric. The spatial Christoffel symbol (usually called the spatial connection)

$$\Gamma^a{}_{bc} = \tfrac{1}{2}g^{ad}\left(g_{db,c} + g_{dc,b} - g_{bc,d}\right) \tag{22}$$

provides the connection between D_a and the ordinary derivative ∂_a.

Using g_{ab} and $\Gamma^a{}_{bc}$, the spatial versions of the Riemann tensor, Ricci tensor,

and Ricci scalar are

$$R^d{}_{abc} = \Gamma^d{}_{ac,b} - \Gamma^d{}_{bc,a} + \Gamma^e{}_{ac}\,\Gamma^d{}_{eb} - \Gamma^e{}_{bc}\,\Gamma^d{}_{ea}\,, \tag{23}$$

$$R_{ab} = R^c{}_{acb}\,, \tag{24}$$

and

$$R = R^a{}_a\,. \tag{25}$$

A.2   Lie Derivatives

According to (C.2.14) of [108], the Lie derivative of an arbitrary tensor field is

$$\mathcal{L}_{\mathrm{v}}T^{a_1\ldots a_k}{}_{b_1\ldots b_l} = v^c\,\nabla_c T^{a_1\ldots a_k}{}_{b_1\ldots b_l} - \sum_{i=1}^{k} T^{a_1.c.a_k}{}_{b_1\ldots b_l}\nabla_c v^{a_i} +$$

$$\sum_{j=1}^{l} T^{a_1\ldots a_k}{}_{b_1.c.b_l}\nabla_{b_j} v^c\,, \tag{26}$$

where v^a is the vector field with respect to which the Lie derivative is defined, $T^{a_1\ldots a_k}{}_{b_1\ldots b_l}$ is the arbitrary tensor field with k contravariant and l covariant indices, and ∇_a is any derivative operator.[4] In the explicit sums, c is a dummy index that replaces one of the tensor field indices, while the missing index appears on v or ∇ as appropriate.

A.2.1   Lie Derivatives of γ_{ij} and K_{ij}

From Eq. (26) the Lie Derivatives of g_{ij} and K_{ij} are

$$\mathcal{L}_{\beta}\,g_{ij} = \beta^k \partial_k g_{ij} + g_{ik}\,\partial_j \beta^k + g_{kj}\,\partial_i \beta^k\,, \tag{27}$$

---

[4]In this section, for convenience and for consistency with the referenced equation, the dimension of tensors is *not* indicated by typeface. All tensors are of the same unspecified dimension.

and

$$\mathcal{L}_\beta K_{ij} = \beta^k \partial_k K_{ij} + K_{ik}\,\partial_j \beta^k + K_{kj}\,\partial_i \beta^k\,. \tag{28}$$

A.2.2 Additive Property of Lie Derivatives

If $v^a = u^a + w^a$, then Eq. (26) becomes

$$\mathcal{L}_{\mathrm{u+w}} T^{a_1\ldots a_k}{}_{b_1\ldots b_l} = (u^c + w^c)\nabla_c T^{a_1\ldots a_k}{}_{b_1\ldots b_l}$$

$$-\sum_{i=1}^{k} T^{a_1.c.a_k}{}_{b_1\ldots b_l}\nabla_c(u^{a_i} + w^{a_i}) + \sum_{j=1}^{l} T^{a_1\ldots a_k}{}_{b_1.c.b_l}\nabla_{b_j}(u^c + w^c)\,. \tag{29}$$

The addition of vectors and the derivative operator are linear. So, after distributing u^a and w^a and collecting terms by vector, Eq. (29) becomes

$$\mathcal{L}_{\mathrm{u+w}} T^{a_1\ldots a_k}{}_{b_1\ldots b_l} = \mathcal{L}_{\mathrm{u}} T^{a_1\ldots a_k}{}_{b_1\ldots b_l} + \mathcal{L}_{\mathrm{w}} T^{a_1\ldots a_k}{}_{b_1\ldots b_l} \tag{30}$$

$$= \left(\mathcal{L}_{\mathrm{u}} + \mathcal{L}_{\mathrm{w}}\right) T^{a_1\ldots a_k}{}_{b_1\ldots b_l}\,. \tag{31}$$

Therefore, the Lie derivative is additive in the vector field v^a, with respect to which it is defined.

A.2.3 Lie Derivatives of Tensor Densities

A tensor density is a tensor times $g^{w/2}$ where g is the determinant of the metric. w is the weight of the tensor density. If T is the tensor density, β^a is the vector along which the Lie derivative is computed, and ∂ is the derivative operator associated with the metric, then according to (22) of [5],

$$\mathcal{L}_\beta T = [\mathcal{L}_\beta T]_\partial^{w=0} + w\,T\partial_k \beta^k\,. \tag{32}$$

161

The first term is the Lie derivative of T treated as a tensor using the ∂ derivative operator, and the second term is the additional term due to the factor of $g^{w/2}$.

So from Eq. (32), the omitted $\mathcal{L}_\beta$ terms are

$$\mathcal{L}_\beta \phi = \beta^k \partial_k \phi + \tfrac{1}{6}\phi\, \partial_k \beta^k \,, \tag{33}$$

$$\mathcal{L}_\beta \tilde{g}_{ij} = \beta^k \partial_k \tilde{g}_{ij} + \tilde{g}_{ik}\, \partial_j \beta^k + \tilde{g}_{kj}\, \partial_i \beta^k - \tfrac{2}{3}\tilde{g}_{ij}\, \partial_k \beta^k \,, \tag{34}$$

$$\mathcal{L}_\beta \tilde{A}_{ij} = \beta^k \partial_k \tilde{A}_{ij} + \tilde{A}_{ik}\, \partial_j \beta^k + \tilde{A}_{kj}\, \partial_i \beta^k - \tfrac{2}{3}\tilde{A}_{ij}\, \partial_k \beta^k \,, \tag{35}$$

and

$$\mathcal{L}_\beta K = \beta^k \partial_k K \,. \tag{36}$$

K is a scalar, and so its Lie derivative is just the partial derivative along β^i.

Bibliography

[1] Paramesh: Parallel adaptive mesh refinement. `http://www.physics.drexel.edu/~olson/paramesh-doc/Users_manual/amr.html`.

[2] Miguel Alcubierre. *Introduction to 3+1 Numerical Relativity*. Oxford University Press, Oxford, U.K., 2008.

[3] Miguel Alcubierre, Gabrielle D. Allen, Bernd Brügmann, Edward Seidel, and Wai-Mo Suen. Towards an understanding of the stability properties of the 3+1 evolution equations in general relativity. *Phys. Rev. D*, 62:124011, 2000.

[4] Miguel Alcubierre and Bernd Brügmann. Simple excision of a black hole in 3+1 numerical relativity. *Phys. Rev. D*, 63:104006, 2001.

[5] Miguel Alcubierre, Bernd Brügmann, Peter Diener, Michael Koppitz, Denis Pollney, Edward Seidel, and Ryoji Takahashi. Gauge conditions for long-term numerical black hole evolutions without excision. *Phys. Rev. D*, 67:084023, 2003.

[6] Peter Anninos, Karen Camarda, Joan Massó, Edward Seidel, Wai-Mo Suen, and John Towns. Three-dimensional numerical relativity: The evolution of black holes. *Phys. Rev. D*, 52(4):2059–2082, 1995.

[7] Marcus Ansorg, Bernd Brügmann, and Wolfgang Tichy. A single-domain spectral method for black hole puncture data. *Phys. Rev. D*, 70:064011, 2004.

[8] R. Arnowitt, S. Deser, and Charles W. Misner. The dynamics of general relativity. In Louis Witten, editor, *Gravitation: An Introduction to Current Research*, pages 227–265. Wiley, New York, 1962.

[9] John G. Baker, William D. Boggs, Joan M. Centrella, Bernard J. Kelly, Sean T. McWilliams, M. Coleman Miller, and James R. van Meter. Modeling kicks from the merger of non-precessing black-hole binaries. *Astrophys. J.*, 668:1140–1144, 2007.

[10] John G. Baker, William D. Boggs, Joan M. Centrella, Bernard J. Kelly, Sean T. McWilliams, M. Coleman Miller, and James R. van Meter. Modeling kicks from the merger of generic black-hole binaries. *Astrophys. J.*, 682:L29–L32, 2008.

[11] John G. Baker, William D. Boggs, Joan M. Centrella, Bernard J. Kelly, Sean T. McWilliams, and James R. van Meter. Mergers of non-spinning black-hole binaries: Gravitational radiation characteristics. *Phys. Rev. D*, 78:044046, 2008.

[12] John G. Baker, Manuela Campanelli, Frans Pretorius, and Yosef Zlochower. Comparisons of binary black hole merger waveforms. *Class. Quantum Grav.*, 24:S25–S31, 2007.

[13] John G. Baker, Joan M. Centrella, Dae-Il Choi, Michael Koppitz, and James R. van Meter. Binary black hole merger dynamics and waveforms. *Phys. Rev. D*, 73:104002, 2006.

[14] John G. Baker, Joan M. Centrella, Dae-Il Choi, Michael Koppitz, and James R. van Meter. Gravitational wave extraction from an inspiraling configuration of merging black holes. *Phys. Rev. Lett.*, 96:111102, 2006.

[15] John G. Baker, Joan M. Centrella, Dae-Il Choi, Michael Koppitz, James R. van Meter, and M. Coleman Miller. Getting a kick out of numerical relativity. *Astrophys. J.*, 653:L93–L96, 2006.

[16] John G. Baker, Sean T. McWilliams, James R. van Meter, Joan M. Centrella, Dae-Il Choi, Bernard J. Kelly, and Michael Koppitz. Binary black hole late inspiral: Simulations for gravitational wave observations. *Phys. Rev. D*, 75:124024, 2007.

[17] John G. Baker, James R. van Meter, Sean T. McWilliams, Joan M. Centrella, and Bernard J. Kelly. Consistency of post-Newtonian waveforms with numerical relativity. *Phys. Rev. Lett.*, 99:181101, 2007.

[18] Thomas W. Baumgarte. Innermost stable circular orbit of binary black holes. *Phys. Rev. D*, 62:024018, 2000.

[19] Thomas W. Baumgarte and Stuart L. Shapiro. Numerical integration of Einstein's field equations. *Phys. Rev. D*, 59:024007, 1998.

[20] Christopher Beetle, Marco Bruni, Lior M. Burko, and Andrea Nerozzi. Towards wave extraction in numerical relativity: Foundations and initial-value formulation. *Phys. Rev. D*, 72:024013, 2005.

[21] Emanuele Berti, Vitor Cardoso, and Clifford M. Will. Gravitational-wave spectroscopy of massive black holes with the space interferometer LISA. *Phys. Rev. D*, 73:064030, 2006.

[22] Carles Bona, Joan Massó, Edward Seidel, and Joan Stela. New formalism for numerical relativity. *Phys. Rev. Lett.*, 75:600–603, 1995.

[23] J. M. Bowen and James W. York Jr. Time-asymmetric initial data for black holes and black-hole collisions. *Phys. Rev. D*, 21:2047–2056, 1980.

[24] Michael Boylan-Kolchin, Chung-Pei Ma, and Eliot Quataert. Core formation in galactic nuclei due to recoiling black holes. *Astrophys. J.*, 613:L37, 2004.

[25] Michael Boyle, David A. Brown, Lawrence E. Kidder, Abdul H. Mroué, Harald P. Pfeiffer, Mark A. Scheel, Gregory B. Cook, and Saul A. Teukolsky. High-accuracy comparison of numerical relativity simulations with post-Newtonian expansions. *Phys. Rev. D*, 76:124038, 2007.

[26] Steve R. Brandt and Bernd Brügmann. A simple construction of initial data for multiple black holes. *Phys. Rev. Lett.*, 78(19):3606–3609, 1997.

[27] Duncan A. Brown, Jeandrew Brink, Hua Fang, Jonathan R. Gair, Chao Li, Geoffrey Lovelace, Ilya Mandel, and Kip S. Thorne. Prospects for detection of gravitational waves from intermediate-mass-ratio inspirals. *Phys. Rev. Lett.*, 99:201102, 2007.

[28] J. David Brown and Lisa L. Lowe. Multigrid elliptic equation solver with adaptive mesh refinement. *J. Comp. Phys.*, 209:582–598, 2005.

[29] Bernd Brügmann, Wolfgang Tichy, and Nina Jansen. Numerical simulation of orbiting black holes. *Phys. Rev. Lett.*, 92:211101, 2004.

[30] Alessandra Buonanno, Gregory B. Cook, and Frans Pretorius. Inspiral, merger and ring-down of equal-mass black-hole binaries. *Phys. Rev. D*, 75:124018, 2007.

[31] Alessandra Buonanno and Thibault Damour. Effective one-body approach to general relativistic two-body dynamics. *Phys. Rev. D*, 59:084006, 1999.

[32] Alessandra Buonanno, Yi Pan, John G. Baker, Joan M. Centrella, Bernard J. Kelly, Sean T. McWilliams, and James R. van Meter. Approaching faithful templates for nonspinning binary black holes using the effective-one-body approach. *Phys. Rev. D*, 76:104049, 2007.

[33] Lior M. Burko, Thomas W. Baumgarte, and Christopher Beetle. Towards a novel wave–extraction method for numerical relativity: III. Analytical examples for the Beetle–Burko radiation scalar. *Phys. Rev. D*, 73:024002, 2006.

[34] Manuela Campanelli, Carlos O. Lousto, Pedro Marronetti, and Yosef Zlochower. Accurate evolutions of orbiting black-hole binaries without excision. *Phys. Rev. Lett.*, 96:111101, 2006.

[35] Manuela Campanelli, Carlos O. Lousto, and Yosef Zlochower. Spin-orbit interactions in black-hole binaries. *Phys. Rev. D*, 74:084023, 2006.

[36] Manuela Campanelli, Carlos O. Lousto, and Yosef Zlochower. Spinning-black-hole binaries: The orbital hang up. *Phys. Rev. D*, 74:041501(R), 2006.

[37] Manuela Campanelli, Carlos O. Lousto, Yosef Zlochower, Badri Krishnan, and David Merritt. Spin flips and precession in black-hole-binary mergers. *Phys. Rev. D*, 75:064030, 2007.

[38] Manuela Campanelli, Carlos O. Lousto, Yosef Zlochower, and David Merritt. Large merger recoils and spin flips from generic black-hole binaries. *Astrophys. J.*, 659:L5–L8, 2007.

[39] Dae-Il Choi, Bernard J. Kelly, William D. Boggs, John G. Baker, Joan M. Centrella, and James R. van Meter. Recoiling from a kick in the head-on collision of spinning black holes. *Phys. Rev. D*, 76:104026, 2007.

[40] D. Christodoulou. Reversible and irreversible transformations in black-hole physics. *Phys. Rev. Lett.*, 25(22):1596–1597, 1970.

[41] Gregory B. Cook. Three-dimensional initial data for the collision of two black holes. II. Quasi-circular orbits for equal-mass black holes. *Phys. Rev. D*, 50(8):5025–5032, 1994.

[42] Matthew D. Duez, Stuart L. Shapiro, and Hwei-Jang Yo. Relativistic hydrodynamic evolutions with black hole excision. *Phys. Rev. D*, 69:104016, 2004.

[43] David R. Fiske, John G. Baker, James R. van Meter, Dae-Il Choi, and Joan M. Centrella. Wave zone extraction of gravitational radiation in three-dimensional numerical relativity. *Phys. Rev. D*, 71:104036, 2005.

[44] Y. Fourès-Bruhat. Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires. *Acta Math.*, 88:141–225, 1952.

[45] José A. González, Mark D. Hannam, Ulrich Sperhake, Bernd Brügmann, and Sascha Husa. Supermassive recoil velocities for binary black-hole mergers with antialigned spins. *Phys. Rev. Lett.*, 98:231101, 2007.

[46] José A. González, Ulrich Sperhake, and Bernd Brügmann. Black-hole binary simulations: the mass ratio 10:1. *Phys. Rev. D*, 79:124006, 2009.

[47] José A. González, Ulrich Sperhake, Bernd Brügmann, Mark D. Hannam, and Sascha Husa. Maximum kick from nonspinning black-hole binary inspiral. *Phys. Rev. Lett.*, 98:091101, 2007.

[48] Philippe Grandclement, Eric Gourgoulhon, and Silvano Bonazzola. Binary black holes in circular orbits. II. Numerical methods and first results. *Phys. Rev. D*, 65:044021, 2002.

[49] Kayhan Gültekin, M. Coleman Miller, and Douglas P. Hamilton. Growth of intermediate-mass black holes in globular clusters. *Astrophys. J.*, 616:221–230, 2004.

[50] Kayhan Gültekin, M. Coleman Miller, and Douglas P. Hamilton. Three-body dynamics with gravitational wave emission. *Astrophys. J.*, 640:156–166, 2006.

[51] Zoltán Haiman. Constraints from gravitational recoil on the growth of supermassive black holes at high redshift. *Astrophys. J.*, 613:36–40, 2004.

[52] Mark D. Hannam, Sascha Husa, Bernd Brügmann, José A. González, and Ulrich Sperhake. Beyond the Bowen-York extrinsic curvature for spinning black holes. *Class. Quantum Grav.*, 24:S15–S24, 2007.

[53] Mark D. Hannam, Sascha Husa, Bernd Brügmann, José A. González, Ulrich Sperhake, and Niall Ó Murchadha. Where do moving punctures go? *J. Phys. Conf. Ser.*, 66:012047, 2007. In the proceedings of 29th Spanish Relativity Meeting (ERE 2006): Einstein's Legacy: From the Theoretical Paradise to Astrophysicsl Observation, Palma de Mallorca, Spain, 4-8 Sep 2006.

[54] Mark D. Hannam, Sascha Husa, and Niall Ó Murchadha. Bowen-York trumpet data and black-hole simulations. *Phys. Rev. D*, 80:124007, 2009.

[55] Mark D. Hannam, Sascha Husa, Frank Ohme, Bernd Brügmann, and Niall Ó Murchadha. Wormholes and trumpets: Schwarzschild spacetime for the moving-puncture generation. *Phys. Rev. D*, 78:064020, 2008.

[56] Frank Herrmann, Ian Hinder, Deirdre M. Shoemaker, Pablo Laguna, and Richard A. Matzner. Gravitational recoil from spinning binary black hole mergers. *Astrophys. J.*, 661:430–436, 2007.

[57] Frank Herrmann, Deirdre M. Shoemaker, and Pablo Laguna. Unequal mass binary black hole plunges and gravitational recoil. *Class. Quantum Grav.*, 24:S33–S42, 2007.

[58] Kelly Holley-Bockelmann, Kayhan Gültekin, Deirdre M. Shoemaker, and Nicolás Yunes. Gravitational wave recoil and the retention of intermediate mass black holes. *Astrophys. J.*, 686:829–837, 2008.

[59] Peter Hübner. A scheme to numerically evolve data for the conformal Einstein equation. *Class. Quantum Grav.*, 16:2823–2843, 1999.

[60] Sascha Husa, Mark D. Hannam, José A. González, Ulrich Sperhake, and Bernd Brügmann. Reducing eccentricity in black-hole binary evolutions with initial parameters from post-Newtonian inspiral. *Phys. Rev. D*, 77:044037, 2008.

[61] Breno Imbiriba, John G. Baker, Dae-Il Choi, Joan M. Centrella, David R. Fiske, J. David Brown, James R. van Meter, and Kevin M. Olson. Evolving a puncture black hole with fixed mesh refinement. *Phys. Rev. D*, 70:124025, 2004.

[62] Lawrence E. Kidder. Coalescing binary systems of compact objects to $(post)^{5/2}$-Newtonian order. V. Spin effects. *Phys. Rev. D*, 52:821–847, 1995.

[63] Michael Koppitz, Denis Pollney, Christian Reisswig, Luciano Rezzolla, Jonathan Thornburg, Peter Diener, and Erik Schnetter. Recoil velocities from equal-mass binary-black-hole mergers. *Phys. Rev. Lett.*, 99:041102, 2007.

[64] Heinz-Otto Kreiss and Joseph Oliger. Methods for the approximate solution of time dependent problems. *Global Atmospheric Research Programme publications series*, 10, 1973.

[65] Stephen R. Lau, Harald P. Pfeiffer, and Jan S. Hesthaven. IMEX evolution of scalar fields on curved backgrounds. *Commun. Comput. Phys.*, 6:1063–1094, 2009.

[66] Noam I. Libeskind, Shaun Cole, Carlos S. Frenk, and John C. Helly. The effect of gravitational recoil on black holes forming in a hierarchical universe. *Mon. Not. R. Astron. Soc.*, 368:1381–1391, 2006.

[67] A. Lichnerowicz. L'intégration des équations de la gravitation relativiste et la problème des n corps. *J. Math Pures et Appl.*, 23:37–63, 1944.

[68] Alan P. Lightman, William H. Press, Richard H. Price, and Saul A. Teukolsky. *Problem Book in Relativity and Gravitation*. Princeton University Press, Princeton, NJ, 1975.

[69] Lee Lindblom, Mark A. Scheel, Lawrence E. Kidder, Robert Owen, and Oliver Rinne. A new generalized harmonic evolution system. *Class. Quantum Grav.*, 23:S447–S462, 2006.

[70] Carlos O. Lousto, Manuela Campanelli, and Yosef Zlochower. Remnant masses, spins and recoils from the merger of generic black-hole binaries. *Class. Quantum Grav.*, 27:114006, 2010.

[71] Carlos O. Lousto and Yosef Zlochower. Modeling gravitational recoil from precessing highly-spinning unequal-mass black-hole binaries. *Phys. Rev. D*, 79:064018, 2009.

[72] Carlos O. Lousto and Yosef Zlochower. Extreme-mass-ratio-black-hole-binary evolutions with numerical relativity. *Phys. Rev. Lett.*, 106:041101, 2011.

[73] Carlos O. Lousto and Yosef Zlochower. Modeling maximum astrophysical gravitational recoil velocities. *Phys. Rev. D*, 83:024003, 2011.

[74] Peter MacNeice, Kevin M. Olson, Clark Mobarry, Rosalinda de Fainchtein, and Charles Packer. PARAMESH: A parallel adaptive mesh refinement community toolkit. *Computer Physics Communications*, 126:330–354, 2000.

[75] Piero Madau and Eliot Quataert. The effect of gravitational-wave recoil on the demography of massive black holes. *Astrophys. J.*, 606:L17–L20, 2004.

[76] Sean T. McWilliams, James I. Thorpe, John G. Baker, and Bernard J. Kelly. Impact of mergers on LISA parameter estimation for nonspinning black hole binaries. *Phys. Rev. D*, 81:064014, 2010.

[77] David Merritt, Miloš Milosavljević, Marc Favata, Scott A. Hughes, and Daniel E. Holz. Consequences of gravitational radiation recoil. *Astrophys. J.*, 607:L9–L12, 2004.

[78] Miroslav Micic, Tom Abel, and Steinn Sigurdsson. The role of primordial kicks on black hole merger rates. *Mon. Not. R. Astron. Soc.*, 372:1540–1548, 2006.

[79] M. Coleman Miller and Edward J. M. Colbert. Intermediate-mass black holes. *Int. J. Mod. Phys. D*, 13:1–64, 2004.

[80] M. Coleman Miller and Douglas P. Hamilton. Four-body effects in globular cluster black hole coalescence. *Astrophys. J.*, 576:894, 2002.

[81] M. Coleman Miller and Douglas P. Hamilton. Production of intermediate-mass black holes in globular clusters. *Mon. Not. R. Astron. Soc.*, 330:232, 2002.

[82] Charles W. Misner, Kip S. Thorne, and John A. Wheeler. *Gravitation.* W. H. Freeman, San Francisco, 1973.

[83] Hideaki Mouri and Yoshiaki Taniguchi. Mass segregation in star clusters: Analytic estimation of the timescale. *Astrophys. J.*, 580:844, 2002.

[84] Takashi Nakamura, Ken ichi Oohara, and Y. Kojima. General relativistic collapse to black holes and gravitational waves from black holes. *Prog. Theor. Phys. Suppl.*, 90:1–218, 1987.

[85] Ezra T. Newman and Roger Penrose. Note on the Bondi-Metzner-Sachs group. *J. Math. Phys.*, 7:863–870, 1966.

[86] Niall Ó Murchadha and James W. York Jr. Gravitational energy. *Phys. Rev. D*, 10(8):2345–2357, 1974.

[87] Ryan M. O'Leary, Frederic A. Rasio, John M. Fregeau, Natalia Ivanova, and Richard O'Shaughnessy. Binary mergers and growth of black holes in dense star clusters. *Astrophys. J.*, 637:937–951, 2006.

[88] Kevin M. Olson. PARAMESH: A parallel adaptive grid tool. In A. Deane, A. Ecer, G. Brenner, D. Emerson, J. McDonough, J. Periaux, N. Satofuka, and D. Tromeur-Dervout, editors, *Parallel Computational Fluid Dynamics 2005: Theory and Applications: Proceedings of the Parallel CFD Conference, College Park, MD, U.S.A.* Elsevier, 2006.

[89] Kevin M. Olson and Peter MacNeice. An overview of the PARAMESH AMR software and some of its applications. In T. Plewa, T. Linde, and G. Weirs, editors, *Adaptive Mesh Refinement-Theory and Applications, Proceedings of the Chicago Workshop on Adaptive Mesh Refinement Methods, Series: Lecture Notes in Computational Science and Engineering*, volume 41, Berlin, 2005. Springer.

[90] Frans Pretorius. Evolution of binary black-hole spacetimes. *Phys. Rev. Lett.*, 95:121101, 2005.

[91] Lewis Fry Richardson. The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam. *Phil. Trans. R. Soc. London Ser. A*, 210:307–357, 1910.

[92] Olivier Sarbach, Gioel Calabrese, Jorge A. Pullin, and Manuel Tiglio. Hyperbolicity of the BSSN system of Einstein evolution equations. *Phys. Rev. D*, 66:064002, 2002.

[93] Mark A. Scheel, Michael Boyle, Tony Chu, Lawrence E. Kidder, Keith D. Matthews, and Harald P. Pfeiffer. High-accuracy waveforms for binary black hole inspiral, merger, and ringdown. *Phys. Rev. D*, 79:024003, 2009.

[94] Mark A. Scheel, Harald P. Pfeiffer, Lee Lindblom, Lawrence E. Kidder, Oliver Rinne, and Saul A. Teukolsky. Solving Einstein's equations with dual coordinate frames. *Phys. Rev. D*, 74:104006, 2006.

[95] Jeremy D. Schnittman, Alessandra Buonanno, James R. van Meter, John G. Baker, William D. Boggs, Joan M. Centrella, Bernard J. Kelly, and Sean T. McWilliams. Anatomy of the binary black hole recoil: A multipolar analysis. *Phys. Rev. D*, 77:044031, 2008.

[96] Edward Seidel and Wai-Mo Suen. Towards a singularity-proof scheme in numerical relativity. *Phys. Rev. Lett.*, 69(13):1845–1848, 1992.

[97] Masaru Shibata and Takashi Nakamura. Evolution of three-dimensional gravitational waves: Harmonic slicing case. *Phys. Rev. D*, 52:5428–5444, 1995.

[98] Larry L. Smarr and James W. York Jr. Kinematical conditions in the construction of spacetime. *Phys. Rev. D*, 17:2529–2551, 1978.

[99] Larry L. Smarr and James W. York Jr. Radiation gauge in general relativity. *Phys. Rev. D*, 17:1945–1956, 1978.

[100] Kenneth L. Smith. *Dynamic Singularity Excision in Numerical Relativity*. PhD thesis, The Pennsylvania State University, University Park, Pennsylvania, 2004.

[101] Saul A. Teukolsky. Rotating black holes: Separable wave equations for gravitational and electromagnetic perturbations. *Phys. Rev. Lett.*, 29:1114–1118, 1972.

[102] Jonathan Thornburg. Coordinates and boundary conditions for the general relativistic initial data problem. *Class. Quantum Grav.*, 4:1119–1131, 1987.

[103] Jonathan Thornburg. A fast apparent-horizon finder for three-dimensional Cartesian grids in numerical relativity. *Class. Quantum Grav.*, 21:743–766, 2004.

[104] James R. van Meter, John G. Baker, Michael Koppitz, and Dae-Il Choi. How to move a black hole without excision: gauge conditions for the numerical evolution of a moving puncture. *Phys. Rev. D*, 73:124011, 2006.

[105] James R. van Meter, M. Coleman Miller, John G. Baker, William D. Boggs, and Bernard J. Kelly. Test of a General Formula for Black Hole Gravitational Wave Kicks. *Astrophys. J.*, 719:1427–1432, 2010.

[106] Marta Volonteri. Gravitational recoil: Signatures on the massive black hole population. *Astrophys. J.*, 663:L5–L8, 2007.

[107] Marta Volonteri and Rosalba Perna. Dynamical evolution of intermediate mass black holes and their observable signatures in the nearby universe. *Mon. Not. R. Astron. Soc.*, 358:913–922, 2005.

[108] Robert M. Wald. *General Relativity*. The University of Chicago Press, Chicago, 1984.

[109] Jeffrey Winicour. Characteristic evolution and matching. *Living Rev. Relativity*, 1(5), 1998. `http://www.livingreviews.org/lrr-1998-5`.

[110] Jaiyul Yoo and Jordi Miralda-Escudé. Formation of the black holes in the highest redshift quasars. *Astrophys. J.*, 614:L25–L28, 2004.

[111] James W. York Jr. Gravitational degrees of freedom and the initial-value problem. *Phys. Rev. Lett.*, 26:1656–1658, 1971.

[112] James W. York Jr. Role of conformal three-geometry in the dynamics of gravitation. *Phys. Rev. Lett.*, 28:1082–1085, 1972.

[113] James W. York Jr. Kinematics and dynamics of general relativity. In Larry L. Smarr, editor, *Sources of Gravitational Radiation*, pages 83–126. Cambridge University Press, Cambridge, England, 1979.