Binary Black Hole Coalescence

Frans Pretorius

Department of Physics, Princeton University, Princeton, NJ 08544
(Dated: August 12 2007)

The two-body problem in general relativity is reviewed, focusing on the final stages of the coalescence of the black holes as uncovered by recent successes in numerical solution of the field equations.

Contents

I. Introduction 2

II. The two-body problem in general relativity 3
   A. Stages of a merger
      1. Astrophysical binaries
      2. The black hole scattering problem

III. Contemporary successful numerical solution methods 7
   A. Historical background 8
   B. Historical background continued—ingredients to assemble a successful numerical 2-body code
      1. Fix the character of the differential equations
      2. Find a formulation admitting a numerically well-posed initial boundary value problem
      3. Curb truncation-error-induced growth of constraints
      4. Provide well behaved dynamical coordinates conditions
      5. Specify good outer boundary conditions
      6. Deal with black hole singularities
      7. Provide consistent and relevant initial data
   C. Generalized harmonic evolution
      1. Source function evolution
   D. BSSN with ‘moving punctures’
      1. Moving punctures
   E. Comparison of the two techniques
   F. Numerical algorithms

IV. Results from black hole merger simulations 19
   A. Equal mass, minimal eccentricity and spin
      1. Gravitational wave structure
   B. Unequal mass, minimal eccentricity and spin
   C. Equal mass, non-negligible spin, minimal eccentricity
      1. Qualitative features of a merger of spinning black holes
      2. Recoil velocities
   D. Equal mass, large eccentricity, minimal spin
   E. Generic

V. Implications, prospects and questions 28
   A. Black holes in our universe
      1. Consequences of radiated energy
      2. Consequences of radiated momentum
      3. Implication of waveform structure for detection efforts
   B. The black hole scattering problem
   C. High energy particle experiments and black hole collisions

VI. Conclusion 32

References 33
I. INTRODUCTION

A black hole is one of the most fascinating and enigmatic predictions of Einstein’s theory of general relativity. Its interior can have rich structure and is intrinsically dynamical, where space and time itself are inexorably led to a singular state. The exterior of an isolated black hole is, on the other hand, remarkably simple, described uniquely by the stationary Kerr solution. The dynamics of black holes are governed by laws analogous to the laws of thermodynamics, and indeed when quantum processes are included, emit Hawking radiation with a characteristic thermal spectrum. Most remarkable however, is that black holes, “discovered” purely through thought and the mathematical exploration of a theory far removed from every day experience, appear to be ubiquitous objects in our universe.

The evidence that black holes exist, though circumstantial, is quite strong [1]. The high luminosity of quasars and other active galactic nuclei (AGN) can be explain by gravitational binding energy released through gas accretion onto supermassive ($10^6 - 10^9 M_\odot$) black holes at the centers of the galaxies [2, 3], several dozen X-ray binary systems discovered to date have compact members too massive to be neutron stars and exhibit phenomena consistent with matter interactions originating in the strong gravity regime of an inner accretion disk [4], and the dynamical motion of stars and gas about the centers of nearby galaxies and our Milky Way Galaxy infer the presence of very massive, compact objects there, the most plausible explanation being supermassive black holes [5, 6, 7].

To conclusively prove that black holes exist one needs to “see” them, or conversely see the compact objects masquerading as black holes. The only direct way of observing black holes is via the gravitational waves they emit when interacting with other matter/energy (an isolated black hole does not radiate). The quadrupole formula says that the typical magnitude $h$ of the gravitational waves emitted by a binary with reduced mass $\mu$ on a circular orbit measured a distance $r$ from the source is (for a review of gravitational wave theory see [8])

$$ h = \frac{16\mu v^2}{r}, $$

where $v$ is the average tangential speed of the two members in the binary (and geometric units are used—Newton’s constant $G = 1$ and the speed of light $c = 1$). This formula suggests that the strongest sources of gravitational waves are simply the most massive objects that move the fastest. To reach large velocities in orbit, the binary separation has to be small; black holes, being the most compact objects allowed in the theory, can reach the closest possible separations and hence largest orbital velocities. Therefore, modulo questions about source populations in the universe, a binary black hole interaction offers one of most promising venues of observing black holes through gravitational wave emission.

Joseph Weber pioneered the science of gravitational wave detection with the construction of resonant bar detectors. Weber claimed to have detected gravitational waves [9], though no similar detectors constructed following his claims were able to observe the putative (or any other) source, and the general consensus is that given the sensitivity of Weber’s detector and expected strengths of sources it is very unlikely that it was a true detection [10]. Note that the existence of gravitational waves is not in doubt—the observed spin down rate of the Hulse-Taylor binary pulsar [11] and several others discovered since, is in complete accord with the general relativistic prediction of spin down via gravitational wave emission. Today a new generation of gravitational wave detectors are operational, including laser interferometers (LIGO [12], VIRGO [13], GEO600 [14], TAMA [15]) and resonant bar detectors (NAUTILUS [16], EXPLORER [17], AURIGA [18], ALLEGRO [19], NIOBE [20]). A future space-based observatory is planned (LISA [21]), and pulsar timing and cosmic microwave background polarization measurements also offer the promise of acting as gravitational wave “detectors” (for reviews see [22, 23]). The ultimate success of gravitational wave detectors, in particular with regards to using them as more that simply detectors, but tools to observe and understand the universe, relies on source modeling to predict the structure of the waves emitted during some event. Even if an event is detected with a high signal-to-noise ratio (SNR), there simply is not enough information contained in such a one dimensional time series to “invert” it to reconstruct the event; rather template banks of theoretical waveforms from plausible sources need to be built and used to decode the signal. In rare cases an electromagnetic counterpart may be detected, for example during a binary neutron star merger if this is a source of short gamma ray bursts, which could identify the event without the need for a template. Though even in such a case, to extract information about the event, its environment, etc. requires source modeling.

Gravitational wave detectors have therefore provided much of the impetus for trying to understand the nature of binary black hole collisions, and the gravitational waves emitted during the process. However, from a theoretical perspective black hole collisions are fantastic probes of the dynamical, strong-field regime of general...
relativity. What is already know about this regime—the inevitability of spacetime singularities in gravitational collapse via the singularity theorems of Penrose and Hawking \[24, 25\]; the spacelike, chaotic "mixmaster" nature of these singularities conjectured by Belinsky, Khalatnikov and Lifshitz (BKL) \[26\]; the null, mass-inflation singularity discovered by Poisson and Israel \[27\] that, together with regions of BKL singularities could generically describe the interiors of black hole; the rather surprising discovery of critical phenomena in gravitational collapse by Choptuik \[28, 29\]; etc—together with the sparsity of solutions (exact, numerical or perturbative), suggests there is potentially a vast landscape of undiscovered phenomena. Of particular interest, and potential application to high energy particle collision experiments, are ultra-relativistic black hole collisions. It is beyond the scope of this article to delve much into these aspects of black hole coalescence, though a brief overview of this will be given in Sec. IV.

The two body problem in general relativity, introduced in more detail in Sec. II, is a very rich and complicated problem, with no known closed-form solution. Perturbative analytic techniques have been developed to deal with certain stages of the problem, in particular the inspiral prior to merger and ringdown after merger. Numerical solution of the full field equations are required during the merger, and this aspect of the problem is the main focus of this article. Much effort has been expended by the community over the past 15-20 years to numerically solve for merger spacetimes, and within the last two years an understanding of this phase of the two body problem is finally being attained. Sec. III summaries the difficulties in discretizing the field equations, and describes the methods known at present that work for black hole collisions, namely generalized harmonic coordinates and BSSN (Baumgarte-Shapiro-Shibata-Nakamura) with moving punctures. Preliminary results are discussed in Sec. IV, though given the rapid pace at which the field is developing much of this will probably be dated in short order. Sec. V concludes with a discussion of some astrophysical and other implications of the results.

II. THE TWO-BODY PROBLEM IN GENERAL RELATIVITY

Consider the classical two body problem of finding the motion of two masses interacting only via the Newtonian gravitational force, and given the initial positions and velocities of the objects. The solution is well known—in the center of mass frame each body travels within the same plane along a conic section with a focus at the center of mass, and the type of conic (ellipse, hyperbola, parabola) depends on the net energy of the system (bound, unbound, marginally unbound). In Newtonian gravity this setup is an idealization to the dynamics of two "real" objects in that one treats them as point sources without any internal structure. If one were to extend the problem to include the structure of the bodies there would be an infinite class of two body problems depending on the nature of the material composition of the objects.

In general relativity (GR) the two body problem is on one hand a significantly more challenging problem than its Newtonian counterpart due to the complexity of solving the Einstein field equations, yet if attention is restricted to the vacuum case is much simpler in that one can formulate the exact problem without idealization: given an initial spacelike slice of a vacuum spacetime containing two black holes, what is the subsequent evolution of the spacetime exterior to the event horizon?\(^1\) If Penrose's cosmic censorship conjecture holds the solution will be unique and entirely independent of the interior structure of the black holes due to the causal structure of the spacetime. A wrinkle in this clean picture of the two body problem in GR is that now, rather than a simple set of mass, position and velocity parameters, there are infinitely many degrees of freedom required to describe the initial conditions. These include the initial properties of each black hole and the gravitational wave content of the spacetime. To constrain the possibilities one could restrict the class of initial conditions to black holes that were, at some time in the past, sufficiently separated to each be well described by a Kerr metric with given mass and spin vector, require the initial spacetime slice to have “minimal” gravitational wave content and possess an asymptotic structure such that the black hole positions and relative velocities can unambiguously be defined. This class of initial conditions will cover the vast majority of conceivable astrophysical black hole binary configurations, and the black hole scattering problem setup discussed later on.

So what makes the two body problem so interesting in GR? First of all, it almost goes without saying that

---

\(^1\) This is not a technically precise definition, as the global structure of the spacetime is being ignored, and to capture the spirit of the two body problem in a technically precise manner applicable to situations in our universe would probably require defining it using the concepts of isolated horizons \[33\], and furthermore restrict the solution to the future domain of dependence of an initial finite volume of the spacetime.
as gravity is one of the fundamental forces influencing our existence and shaping the structure of the cosmos, and since GR is the best theory of gravity at our disposal, we want to understand all the details of of the more basic interactions in GR. A less prosaic reason to study this problem is the rich and fascinating phenomenology of solutions: what in Newtonian gravity is entirely describable by the mathematics of conic sections is now a problem that is unlikely to have been a closed form solution in any but the most trivial scenarios, featuring regimes with complicated orbital dynamics, and is accompanied by the emission of gravitational radiation. It is this latter feature which has the most profound implication: the two body problem in GR for any bound system is unstable, and will eventually result in the decay of the orbit and collision of the two black holes. If cosmic censorship holds the collision will always result in a single black hole, and then, from the “no hair” theorems of Israel and Carter [30, 31, 32], one knows that the exterior structure of the remnant black hole will eventually settle down to the Kerr solution. For an idea of how unstable orbits are to gravitational radiation, the time to merger $t_m$, in units of the Hubble time $t_H$, for an equal mass binary with each black hole having a mass $M$, initially separated by $R_0$ times the Schwarzschild radius $R_s = 2GM/c^2$, is roughly

$$\left( \frac{t_m}{t_H} \right) \approx \left( \frac{M}{M_\odot} \right) \left( \frac{R_0}{10^6 R_s} \right)^4$$

(2)

For example, two solar mass black holes initially a million Schwarzschild radii, or $\approx 3 \times 10^6 km$ apart, will merge within a Hubble time; two $10^9 M_\odot$ supermassive black holes need to be within $\approx 6 \times 10^3$ of their Schwarzschild radii, or roughly 1 parsec, to merge within the age of the universe.

In the following section the qualitative features of the two body interaction in general relativity are described.

### A. Stages of a merger

This article is primarily concerned with numerical solution of the field equations as a tool to study the two body problem. However, it is only in the final stages of coalescence where full numerical solution is required to obtain an accurate depiction of the spacetime. This stage of a merger occurs on a very short time scale compared to other phases of the two body interaction, which is fortunate, for due to the computational complexity of solving the field equations it is not feasible to evolve the spacetime for times much longer than this. In the following two sections a more detailed discussion of the various stages of a merger is given, in particular to set the scope for the remainder of the article and to highlight how much interesting phenomenology in the two body problem is not addressed by full numerical simulation. We break the discussion up into two classes of merger, “astrophysical”, and the black hole scattering problem. A merger scenario is considered astrophysical if, to some approximation and non-negligible likelihood the initial conditions could be realized by a binary system in our universe. The latter classification deals with the gedanken experiment of colliding two black holes with ultra-relativistic initial velocities and with an impact parameter of order the total energy of the system or less.

One reason for this classification is that we might expect very different qualitative behavior of the spacetime in these two cases. Consider two black holes of mass $m_1$ and $m_2$ with net ADM [34] energy $E$ in the center of mass frame. All astrophysical mergers are expected to take place in the rest-mass dominated regime where $(m_1 + m_2)/E \approx 1$, while in the scattering problem the kinetic energy of the black holes will dominate so that $(m_1 + m_2)/E \approx 0$. In the latter regime the geometry of each black hole also gets length contracted into a pancake-like region, with the actual black holes occupying an ever smaller region of the non-trivial geometry as the boost factor increases. In fact, eventually it does not matter that it was a black hole that was boosted to large energies—any compact source will produce the same geometry in the limit. The ultra-relativistic limit might also be an interesting place to look for violations of cosmic censorship—the collision of plane-fronted gravitational waves generically leads to the formation of naked singularities [33], and though not exactly analogous to high energy black hole collisions, there are enough similarities that it would be worth while to explore this regime of the two body problem in some detail. Note that the particular value of $E$ is not relevant; there is no intrinsic length scale in the field equations of general relativity, and any solution with energy $E_0$ can trivially be re-scaled to a “new” solution with arbitrary energy $E$.

---

2 Here we consider $m_1$ and $m_2$ to be the total BH mass including spin energy, so not the irreducible mass.
1. Astrophysical binaries

Here the astrophysical merger scenario is broken down into four stages: Newtonian, inspiral, plunge/merger and ringdown.

**Newtonian:** In this stage the two black holes are sufficiently far apart that gravitational wave emission will be too weak to cause the binary to merge within a Hubble time $t_H$ \(2\). Thus, to have any hope of observing mergers of binaries formed in this stage, other “Newtonian”, non-two-body processes need to operate. For example, in the stellar mass range, it is unlikely that a close black hole binary could be formed as the end point of the evolution of a massive binary star system. The reason is that at the requisite separations for a subsequent gravitational wave driven inspiral within $t_H$, the stars will most likely evolve through a common envelope phase, and recent results have suggested this will cause a merger of the stellar cores before a binary black hole system could be formed \(30\). A likely mechanism then to produce hard binaries is through n-body interactions that occur in dense cluster environments \(37, 38\). For supermassive black hole binaries, which are thought to form during galaxy mergers in the hierarchical structure formation scenario \(39, 40\), gas interactions, dynamical friction and other n-body processes are thought cable of driving most black holes close enough so that gravitational wave emission can take over and cause a merger \(41, 42\). If such processes did not operate efficiently it would be in apparent contradiction with the observations that most galaxies harbor supermassive black holes at their centers \(2, 3\).

**Inspiral:** In the inspiral regime gravitational wave emission becomes the dominant process driving the black holes to closer separation, though the orbital time scale is still much shorter than the time scale over which orbital parameters change. The majority of non-extreme mass ratio binaries are expected to “form” with sufficiently large semi-major axis that the orbit will circularize via gravitational wave emission long before the binaries merge \(43, 44\). Some exceptional cases might be stellar and intermediate mass binaries in dense star clusters, where numerous interactions with neighbors could frequently perturb the orbit, or triple systems where the Kozai mechanism operates \(45, 46, 47, 48, 49, 50\). On the other hand, the majority of extreme mass ratio systems that will merge within the Hubble time are expected to have sizable eccentricities at merger \(51, 52\). Note however that much of the theory behind the formation mechanisms and environments of binary black holes are not well known (indeed, no candidate binary black hole system has yet been observed), which offers gravitational wave detection a fantastic opportunity to help decipher some of these interesting questions.

The inspiral phase is well modeled by post-Newtonian (PN) methods \(53\). For initially non-spinning, zero eccentricity binaries the higher order PN approximations and effective one body (EOB) resummations \(54\) give waveforms that are surprisingly close to full numerical results even until very close to merger, well beyond when naive arguments suggest they should fail \(55, 56, 57, 58, 59\); comparisons for more generic scenarios have yet to be made. Extreme mass ratio inspirals (EMRIs) can also be well described by geodesic motion in a black hole background together with prescriptions for computing the gravitational wave emission and effects of radiation reaction \(60, 61, 62, 63, 64, 65\). Generic (non-equatorial) orbits about a Kerr black hole will not lie in a plane due to precession and frame-dragging effects, and thus during the lengthy course of an EMRI, which could be in LISA-band for thousands of cycles, the small black hole will “sample” much of the geometry of the background spacetime. The structure of the corresponding gravitational waves emitted will therefore contain a map of this geometry, and so EMRIs offer a remarkable opportunity to probe the geometry of a black hole, and will be able to confirm whether it is indeed described by the Kerr metric \(52, 72\).

**Plunge/merger:** Here, for non extreme-mass-ratio systems, gravitational wave emission becomes strong enough that the evolution of the orbit is no longer adiabatic, and the black holes plunge together to form a single black hole. Understanding this phase requires full numerical simulations, and it is only within the last couple of years that such simulations have become available. The interesting picture that is now emerging is that this phase is very short, lasting on the order of one to two gravitational wave cycles. To get an idea for the couple of years that such simulations have become available. The interesting picture that is now emerging is that this phase is very short, lasting on the order of one to two gravitational wave cycles. The reason is that at the requisite separations for a subsequent gravitational wave driven inspiral within $t_H$, the stars will most likely evolve through a common envelope phase, and recent results have suggested this will cause a merger of the stellar cores before a binary black hole system could be formed \(30\). A likely mechanism then to produce hard binaries is through n-body interactions that occur in dense cluster environments \(37, 38\). For supermassive black hole binaries, which are thought to form during galaxy mergers in the hierarchical structure formation scenario \(39, 40\), gas interactions, dynamical friction and other n-body processes are thought cable of driving most black holes close enough so that gravitational wave emission can take over and cause a merger \(41, 42\). If such processes did not operate efficiently it would be in apparent contradiction with the observations that most galaxies harbor supermassive black holes at their centers \(2, 3\).

The plunge/merger phase requires full numerical simulations, and it is only within the last couple of years that such simulations have become available. The interesting picture that is now emerging is that this phase is very short, lasting on the order of one to two gravitational wave cycles. The reason is that at the requisite separations for a subsequent gravitational wave driven inspiral within $t_H$, the stars will most likely evolve through a common envelope phase, and recent results have suggested this will cause a merger of the stellar cores before a binary black hole system could be formed \(30\). A likely mechanism then to produce hard binaries is through n-body interactions that occur in dense cluster environments \(37, 38\). For supermassive black hole binaries, which are thought to form during galaxy mergers in the hierarchical structure formation scenario \(39, 40\), gas interactions, dynamical friction and other n-body processes are thought cable of driving most black holes close enough so that gravitational wave emission can take over and cause a merger \(41, 42\). If such processes did not operate efficiently it would be in apparent contradiction with the observations that most galaxies harbor supermassive black holes at their centers \(2, 3\).

**Ringdown:** Here the astrophysical merger scenario is broken down into four stages: Newtonian, inspiral, plunge/merger and ringdown.

**Newtonian:** In this stage the two black holes are sufficiently far apart that gravitational wave emission will be too weak to cause the binary to merge within a Hubble time $t_H$ \(2\). Thus, to have any hope of observing mergers of binaries formed in this stage, other “Newtonian”, non-two-body processes need to operate. For example, in the stellar mass range, it is unlikely that a close black hole binary could be formed as the end point of the evolution of a massive binary star system. The reason is that at the requisite separations for a subsequent gravitational wave driven inspiral within $t_H$, the stars will most likely evolve through a common envelope phase, and recent results have suggested this will cause a merger of the stellar cores before a binary black hole system could be formed \(30\). A likely mechanism then to produce hard binaries is through n-body interactions that occur in dense cluster environments \(37, 38\). For supermassive black hole binaries, which are thought to form during galaxy mergers in the hierarchical structure formation scenario \(39, 40\), gas interactions, dynamical friction and other n-body processes are thought cable of driving most black holes close enough so that gravitational wave emission can take over and cause a merger \(41, 42\). If such processes did not operate efficiently it would be in apparent contradiction with the observations that most galaxies harbor supermassive black holes at their centers \(2, 3\).

**Inspiral:** In the inspiral regime gravitational wave emission becomes the dominant process driving the black holes to closer separation, though the orbital time scale is still much shorter than the time scale over which orbital parameters change. The majority of non-extreme mass ratio binaries are expected to “form” with sufficiently large semi-major axis that the orbit will circularize via gravitational wave emission long before the binaries merge \(43, 44\). Some exceptional cases might be stellar and intermediate mass binaries in dense star clusters, where numerous interactions with neighbors could frequently perturb the orbit, or triple systems where the Kozai mechanism operates \(45, 46, 47, 48, 49, 50\). On the other hand, the majority of extreme mass ratio systems that will merge within the Hubble time are expected to have sizable eccentricities at merger \(51, 52\). Note however that much of the theory behind the formation mechanisms and environments of binary black holes are not well known (indeed, no candidate binary black hole system has yet been observed), which offers gravitational wave detection a fantastic opportunity to help decipher some of these interesting questions.

The inspiral phase is well modeled by post-Newtonian (PN) methods \(53\). For initially non-spinning, zero eccentricity binaries the higher order PN approximations and effective one body (EOB) resummations \(54\) give waveforms that are surprisingly close to full numerical results even until very close to merger, well beyond when naive arguments suggest they should fail \(55, 56, 57, 58, 59\); comparisons for more generic scenarios have yet to be made. Extreme mass ratio inspirals (EMRIs) can also be well described by geodesic motion in a black hole background together with prescriptions for computing the gravitational wave emission and effects of radiation reaction \(60, 61, 62, 63, 64, 65\). Generic (non-equatorial) orbits about a Kerr black hole will not lie in a plane due to precession and frame-dragging effects, and thus during the lengthy course of an EMRI, which could be in LISA-band for thousands of cycles, the small black hole will “sample” much of the geometry of the background spacetime. The structure of the corresponding gravitational waves emitted will therefore contain a map of this geometry, and so EMRIs offer a remarkable opportunity to probe the geometry of a black hole, and will be able to confirm whether it is indeed described by the Kerr metric \(52, 72\).

**Plunge/merger:** Here, for non extreme-mass-ratio systems, gravitational wave emission becomes strong enough that the evolution of the orbit is no longer adiabatic, and the black holes plunge together to form a single black hole. Understanding this phase requires full numerical simulations, and it is only within the last couple of years that such simulations have become available. The interesting picture that is now emerging is that this phase is very short, lasting on the order of one to two gravitational wave cycles. To get an idea for the time scale of this regime, the Keplerian orbital angular frequency $\omega$ for an equal mass quasi-circular inspiral is

$$\frac{\omega}{2\pi} = \frac{1}{2\pi} \sqrt{\frac{M}{R^3}} \approx 11\text{kHz} \frac{M_\odot}{M} \left(\frac{R_\odot}{R}\right)^{3/2},$$

where $M$ is the total mass of the binary with corresponding Schwarzschild radius $R_\odot$, $R$ is the separation, and the plunge/merger happens as $R_\odot/R \to 1$. Note that the frequency of the dominant quadrupolar ($\ell = 2, m = 2$) component of the gravitational wave that is emitted is twice the orbital frequency. The structure of the wavefront is quite simple, however, this is the time of strongest gravitational wave emission, with the luminosity approaching on the order of one-hundredth of the Planck luminosity of $10^{59}$ergs/s, making black hole mergers by far the most energetic events in the post-big-bang era of the universe. Furthermore, the frequency of the
emitted wave rapidly grows to that of the dominant quasinormal mode frequency of the final black hole, causing the spectrum of the plunge/merger phase to occupy a large region of the frequency domain. For equal mass systems upwards of 3% of the rest mass energy of the system is radiate away here. A more detailed discussion of this phase is give in Sec. IV.

If cosmic censorship holds (and there are no signs that it is violated in any merger simulations to date), then Hawking’s “no-bifurcation” theorem [73] states that a single black hole must result as the consequence of a merger. The uniqueness, or “no hair” theorems [30, 31, 32] further imply that the newly formed black hole must eventually settle down to the Kerr solution in the so-called ringdown phase.

Ringdown: The ringdown is the phase where the remnant black hole can be described as a perturbed Kerr spacetime. A more precise definition might be the time at which the gravitational waves emitted from the merger can, to good precision, be written entirely as a superposition of quasi-normal modes (QNMs) [74, 75, 76, 77, 78, 79] of the final black hole. Given appropriate initial conditions, the ringdown phase could be calculated using perturbative techniques, in particular using the so-called close limit approximation [82]. As discussed more in Sec. IV, the early simulation results suggests this description is already adequate very shortly after formation of the common apparent horizon, which roughly coincides with the time of peak luminosity. Very shortly after ringdown begins, the waveform (|h| \propto e^{-t/\tau} \sin(\omega t)) is dominated by the least damped (fundamental harmonic) quadrupolar QNM, with angular frequency $\omega_2$ and decay time $\tau_2$, given approximately by the following fitting formulas [80, 81]

\[
\frac{\omega_2}{2\pi} \approx \frac{1}{2\pi M} \left[ 1 - 0.63(1 - j)^{0.3} \right] \approx 32\text{kHz} \frac{M_{\odot}}{M} \left[ 1 - 0.63(1 - j)^{0.3} \right]
\]

\[
\tau_2 \approx \frac{4M(1 - j)^{-0.45}}{1 - 0.63(1 - j)^{0.3}} \approx 20\mu \text{s} \frac{M}{M_{\odot}} \frac{1}{1 - 0.63(1 - j)^{0.3}},
\]

where $M$ and $J = jM^2$ are the total mass and angular momentum of the final black hole (with $|j| \leq 1$). The dominant ringdown frequency is several times higher than the orbital frequency in the last few inspiral cycles, and the decay time is quite short, so the majority of the energy lost during ringdown (1%–2% of the rest mass) is emitted quite rapidly. Waves propagating in a curved spacetime like Kerr are back-scattered off the curvature, producing so-called power law tails [83]. They decay by integer powers of time, and so even though initially of much smaller amplitude than the ringdown waves, they will eventually dominate the late-time structure of the gravitational wave. Given their small amplitude it is unlikely that the tails could be detected by ground-based detectors.

The relative simplicity of the plunge/merger waveform, together with how short this phase is, suggests it may be possible to build effective analytic template banks of merger waveforms by stitching together PN inspiral waveforms with ringdown waveforms. Numerical simulations of the plunger/merger phase can provide instruction on exactly how this stitching should be performed, i.e., how long the transition region is, which set of quasi-normal modes are excited, how does the waveform interpolate between inspiral and ringdown modes, etc. In fact, this kind of prescription for constructing templates for mergers was already proposed several years ago by proponents of the effective-one-body (EOB) approach to binary dynamics [54, 61], and was recently demonstrated to work well for the extreme mass ratio problem [62] and a range of non-spinning comparable mass mergers [58]. Why might such a “simple” approach work so well for the merger phase, which was anticipated to be a showcase of the complexity and non-linearity of the field equations? First of all, recall that the PN approaches (including the EOB) are hardly simple, having required the dedicated effort of many researchers over a couple of decades to push to the orders presently known [53]—($v/c)^7$ beyond Newtonian order for non-spinning binaries, and ($v/c)^5$ if spin is included. At such high orders in $v/c$ it is not too surprising that much of the essential physics is already being captured, and the only question becomes how far the approach can be trusted. As the velocity of the black holes increase toward the merger one would expect the expansions to become increasingly inaccurate.

3 The QNM spectrum is not complete, and so it is conceivable that they might not be able to exactly describe the wave structure.

4 For the quasi-circular equal mass inspirals the coordinate velocities of the apparent horizons only approach around $v = 0.3$ prior to formation of the common horizon.
could also be expected to capture much of the late time physics of a merger.

2. The black hole scattering problem

Consider the collision of two black holes in the center of mass frame with masses $m_1$ and $m_2$, Kerr spin parameters $a_1$ and $a_2$, and initially moving toward each other with impact parameter $b$ and (large) Lorentz $\gamma$-factors $\gamma_1$ and $\gamma_2$. At present very little is known about all the possible outcomes as a function of $(b, \gamma_1, m_1, a_1, 2)$, though one can speculate about several distinct stages, that will be classified here as Lorentz, collision/ringdown, scatter and threshold. Note that in contrast to the rest-mass dominated case, there is not necessarily such a straight-forward progression through the phases. In particular, there could be a range of impact parameters where the black holes do not merge during the initial encounter, but have lost sufficient energy that they now form a bound system. Then subsequent evolution of the system will follow the stages of the astrophysical binaries outlined in the preceding section.

Lorentz: With sufficiently large $\gamma$ factors the initial non-trivial geometry of each black hole is Lorentz contracted into a thin “pan-cake” (or plane-wave) transverse to the direction of propagation, and close to Minkowski spacetime on either side\(^5\).

collision/ringdown: As suggested by studies of colliding black holes in the infinite $\gamma$ limit\(^6\), if the impact parameter is close to zero there will not be any phase analogous to inspiral; rather an encompassing apparent horizon forms at the moment of collision, and this will presumably settle down to a Kerr black hole. Estimates based on the size of the initial apparent horizon place an upper limit of 30% on the net energy of the spacetime that could be radiated in a head-on collision, though these limits weaken as the impact parameter increases. For the head-on collision case, perturbative studies suggest the energy radiated is close to 16%\(^8\).

scatter: For larger values of the impact parameter there will be a deflection of the two black hole trajectories, accompanied by a burst of radiation, after which they will move apart and the spacetime near each black hole will settle down to the Lorentz phase again. It has been suggested that there may even be a regime where a third or more black holes are formed during the interaction of the two black holes before they scatter, essentially due to the strong focusing of gravitational waves by the shock-fronts representing the boosted black holes\(^9\). This would be an astonishing addition to the phenomenology of the two body problem in general relativity if the scenario can be realized.

threshold: At intermediate values of the impact parameter there could be threshold-type behavior as seen when fine-tuning eccentric orbits in the rest-mass dominated regime\(^9\). Namely, approaching a critical value $b = b^*$ of the impact parameter, the two black holes settle into the analogue of an unstable circular geodesic orbit, whirl around for an amount of time proportional to $-\ln |b - b^*|$, then either fly apart or plunge together (this is described in more detail in Sec.\(^\text{V.D}\)). During this phase copious amounts of energy could be radiated in gravitational waves; in fact, at threshold it is conceivable that essentially all of the kinetic energy of the system is radiated as gravitational waves in $O(10)$ orbits. If the black holes merge after the whirl phase, there will be a plunge/merger and ringdown phase similar to astrophysical binaries. If they separate and have lost enough kinetic energy to form a bound system they will enter the inspiral phase of an astrophysical binary, otherwise they will fly apart as in the scatter phase above. It is tempting to speculate that exactly at threshold, $b = b^*$, the spacetime may approach a self-similar solution (see the discussion in Sec.\(^\text{V.B}\)).

III. CONTEMPORARY SUCCESSFUL NUMERICAL SOLUTION METHODS

This section describes the two methods of formulating the field equations presently known that are amenable to stable numerical integration of binary black hole spacetimes\(^5\), namely generalized harmonic coordinates with constraint damping (GHC) and the Baumgarte-Shapiro-Shibata-Nakamura (BSSN) formalism with moving punctures\(^9\). It is beyond the scope of this article to discuss either method in much detail, or all the variations and details of particular codes; rather the equations will be presented and briefly discussed to

---

\(^5\) In the the limit $\gamma \to \infty$ and $m \to 0$ with $m\gamma = E$ kept constant, one obtains the Aichelburg-Sexi solution\(^6\), and the spacetime becomes exactly Minkowski on either side of a propagating $C^0$ kink in the geometry.

\(^6\) At least for the regions of parameter space studied to date, which are all in the rest-mass dominated regime.
provide the reader with some appreciation for the similarities and differences between them. Note also that if a code produces an apparently stable, convergent solution, it is much more likely that the method actually is stable, compared to the opposite situation where a simulation “crashes” and from which one would like to conclude that the method is unstable. This is simply because bugs are easy to make, more difficult to find, and almost never “help” in any interpretation of the word. The point of this discussion is that there have been numerous good ideas and formulations of the field equations proposed over the past several years (see for example [100, 101]), and only in a few cases were they studied with sufficient detail to conclude that they were unstable; thus that we now know of two methods that are stable does not imply that all earlier methods are not. A case-and-point might be the \(Z_4\) formalism [102, 103] proposed several years ago, which is quite similar in some respects to generalized harmonic coordinates, and to which the same constraint damping mechanism can be applied [104]. On the other hand, the fact that “zero’s” need to be added to the equations in just the right way to make things stable also tells us that the Einstein equations are even more subtle and intricate than previously thought.

A. Historical background

The first attempt at a numerical solution to the field equations for a binary black hole spacetime was carried out by Hahn and Lindquist [105] in 1964. This was even before the word “black hole” had been coined by Wheeler, and they evolved what was then called the “worm hole” initial data of Misner [106]. They considered a time-symmetric scenario in axisymmetry, and reported a run performed on an IBM 7090 computer, using a 51x151 mesh. It took about 4 hours to complete 50 time steps, after which they concluded that errors had grown too large to warrant further evolution. This corresponded to a time of \(m/2\), with \(m = \sqrt{A/16\pi}\), \(A\) being the area of the throat. Given the short run time, not much physics could be extracted from the simulation, yet even so there was no motivation to explore gravitational wave emission. In 1975 Smarr [107], and shortly afterwards Eppley [108], again simulated the head-on collision of two black holes now, with one of the primary goals being to compute the gravitational waves emitted in the process. Despite still being an axisymmetric simulation and almost a decade after Hahn and Lindquist, it was still beyond the capabilities of computers of the time to integrate the field equations with sufficient resolution to obtain very accurate results. Nevertheless, they were able to extract gravitational waveforms from the solutions, calculating (with uncertainties of a factor of 2) that upwards of 0.1% of the rest mass energy is released in gravitational waves in a time-symmetric case where the initial proper separation between the two throats is \(9.6M\) [109]. Primarily because of the stringent computational requirements for numerical solution, no further work on the problem was carried out until the early 1990’s, when prospects for the construction of LIGO became solid. LIGO was the impetus for returning to the two body problem as it was realized early on [10] that given a practical design for the instrument, together with the estimated density of sources in our universe, matched-filtering would be an essential data analysis tool to allow a decent detection rate within a several year time-frame. Matched-filtering looks for known signals in a noisy data stream by convolving theoretical templates of the signals with the data. To be successful it is therefore imperative to understand the gravitational wave emission properties of the source with sufficient detail to construct the template libraries.

The early expectations following a revisit of the head-on collision case [110] was that although certain issues about the generic merger problem had yet to be fully addressed and could be complicated, such as having well behaved coordinates and providing astrophysically relevant initial conditions, a fair consensus was that the most significant hurdle to the problem was lack of computer power [111, 112]. Certainly a portion of the difficulties encountered may be traceable to attempting to find solutions with insufficient resolution, though it turns out that a host of additional issues had to be “discovered”, understood and overcome to reach the state where the field is today.

The review of the history of the numerical two body problem now continues, though switching to a non-traditional format: instead of trying to follow events in chronological order the ingredients needed for a successful simulation will be summarized, noting contributions that offered insights or solutions to the various problems.

---

7 And my apologies in advance to authors that I have missed here.
B. Historical background continued—ingredients to assemble a successful numerical 2-body code

The list of ingredients described below certainly “makes sense”, and so one might wonder, why not satisfy all of them to begin with? First of all, many of the issues, such as choosing well behaved coordinates, are quite complicated, and one does not expect general solutions applicable to all spacetimes of interest. Second, it was perhaps not fully appreciated how vast the landscape of free-evolution schemes are, i.e. systems of equations that give solutions to the Einstein equations for only a restricted subset of initial and boundary conditions, and how important the behavior of these equations are for initial and boundary conditions that do not exactly satisfy those requirements. This is particularly so because it is numerical truncation error that sources “constraint violations”, which is not a priori a problem as truncation error is a well understood, part-and-parcel component of any numerical solution. The surprising thing then is that, in dynamical systems language, it appears as if for the vast majority of free evolution formulations of the Einstein equations, trajectories through phase space denoting solutions to Einstein’s equations form an unstable manifold in the space of all solution trajectories. A third reason is the ADM formulation, in the form popularized by York [113], certainly also “makes sense”, and seems to be a very reasonable and intuitive approach to an initial boundary value formulation of the field equations. Furthermore, given the success of the ADM equations in early evolutions of many symmetry reduced spacetimes, there was not much reason to suspect problems with it.

1. Fix the character of the differential equations

The Einstein field equations

\[ G_{ab} = 8\pi T_{ab}, \]  

when expanded verbatim in terms of the metric \( g_{ab} \)

\[ ds^2 = g_{ab}dx^a dx^b, \]  

results in a coupled system of 10, quasilinear, second order partial differential equations for the 10 independent components of the metric, depending on the four spacetime coordinates \( x^a \). However, these equations have no definite mathematical character—hyperbolic, parabolic or elliptic—and moreover, do not admit a well posed initial value problem. This in large part is due to the gauge invariance of the field equations: for a given, unique physical spacetime there are infinitely many different metric tensors describing it and all satisfying the same equations. The first step towards obtaining a well posed system of equations is to specify enough of the gauge to fix the character of each of the four spacetime coordinates \( x^a \). There are several possibilities, the most common being to choose one coordinate (\( t \)) to be timelike, and the remaining three (\( x, y, z \)) to be spacelike. After a bit more work, outlined in the following item, one comes up with a system of elliptic/hyperbolic equations. This “space-plus-time” (or 3+1) approach will be the exclusive focus of the remainder of the article, after briefly mentioned one alternative, characteristic or null coordinates (for more details, see for example [101, 114]). Here, one (single null) or two (double null) coordinates are chosen to be lightlike, and the rest of the coordinates spacelike. Single null evolution schemes have been very successful in evolving single black hole spacetimes [115]. Part of the reason for pursuing null evolution is that it is easy to extend the domain to future null infinity, which is ideally where one would want to measure gravitational waves. The difficulty with this approach is preventing or treating caustics that can form along the null coordinate in non-trivial, dynamical spacetimes, and no viable mechanism has yet been proposed that might be applied to a binary black hole spacetime. Hybrid null—3 + 1 schemes (often called Cauchy-characteristic matching) have been proposed (see [114] and the references therein), whereby a 3+1 scheme is used to evolve the spacetime in the vicinity of the binary, a characteristic scheme far from the binary, and the solutions mapped to one another in an intermediate zone where the coordinate systems overlap. The matching process is non-trivial, and to date the method has only been applied to single black hole spacetimes [116, 117, 118].

---

8 Again, using geometric units where the speed of light \( c = 1 \) and Newton’s constant \( G = 1 \).
To obtain a well-posed 3 + 1 formulation of the Einstein equations more work needs to be done than merely choosing one timelike and three spacelike coordinates. The choice of which set of fields to treat as the dependent variables of the system of PDEs, the gauge conditions that will be used, as well as modification of the equations by the addition of constraint terms (i.e. terms that are identically zero for any solution of the Einstein equations) all play an important role in determining the ultimate stability of the system. The “traditional ADM” approach, as outlined by York [113], is based on a Hamiltonian formulation of the field equations due to Arnowitt, Deser and Misner [34]. The end result is that the equations are rewritten in terms of quantities either intrinsic or extrinsic to \( t = \text{const.} \) slices of the geometry. First, the four dimensional metric is decomposed as
\[
d s^2 = -\alpha^2 d t^2 + h_{ij} \left( d x^i + \beta^i d t \right) \left( d x^j + \beta^j d t \right),
\]
where \( h_{ij} \) is the spatial metric of the \( t = \text{const.} \) hypersurface, \( \alpha \) is the lapse function measuring the rate at which proper time flows relative to \( t \) for a hypersurface-normal observer, and \( \beta^i \) is the spatial shift vector describing how the spatial coordinate label for such an observer changes with time \( t \). In other words, the time flow vector \( (\partial/\partial t)^a \) is related to the unit hypersurface normal vector \( n^a \) by
\[
\left( \frac{\partial}{\partial t} \right)^a = \alpha n^a + \beta^a
\]
In this way of describing the four-geometry the lapse and shift naturally represent the coordinate degrees of freedom in the theory. Second, the manner in which \( h_{ij} \) is embedded into the four dimensional space is described by the extrinsic curvature tensor \( K_{ij} \):
\[
K_{ij} \equiv -h^{ia}h^{jb}\nabla_b h_{ai}
\]
\[
= -\frac{1}{2\alpha} \left( \frac{\partial h_{ij}}{\partial t} - \mathcal{L}_\beta h_{ij} \right).
\]
In terms of the variables \((h_{ij}, K_{ij}, \alpha, \beta^i)\) the field equations can be written as a set of 12 independent hyperbolic evolution equations for \((h_{ij}, K_{ij})\), 4 constraint equations that do not contain any time derivatives of \( K_{ij} \), and need to be augmented with evolution equations for the gauge quantities \((\alpha, \beta^i)\) (see [119] for an overview of oft-used choices). A common way to proceed to solve these equations is by free evolution (see [119] for a discussion of the general alternatives): the constraints are only solved at the initial time, and the remaining equations are then used to evolve the variables with time. In a consistent discretization scheme the constraint equation will remain zero to within numerical truncation error, and hence, as mentioned before, that the constraints are not strictly enforced is not necessarily a problem. However, in general scenarios, i.e. when there are no symmetries that can be used to simplify the equations, it turns out that the standard ADM form of the equations just outlined is only weakly hyperbolic, and this implies that one cannot in general find a fully consistent, hence stable discretization scheme for the system [120]. This problem began to be appreciate by the numerical relativity community in the mid-90’s, which sprouted a cottage industry of finding symmetric-hyperbolic reductions or various more “ad-hoc” modifications of the field equations [103, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148]. Unfortunately, even though some of these methods were successfully applied to single black hole spacetimes, they offered only marginal improvements at best compared to ADM codes for the binary black hole problem [149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171], with the arguable exception of the BSSN formulation, which showed success in binary neutron star evolutions (see [172] and the works cited therein), and set the record for the longest binary black hole evolution [173] prior to the breakthroughs of 2005 [97, 98, 99]. One reason why some of these methods, even though provably stable, can still “fail” in practice, is if the truncation error grows too rapidly with time. The truncation error \( f_{te}(t, x^i) \) for a variable \( f \) will, to leading order in the mesh spacing \( h \) in an \( n^{\text{th}} \) order discretization scheme, have the form \( f_{te}(t, x^i) = e(t, x^i)h^n \). Formal stability only requires that the \( h \)-independent error term \( e(t, x^i) \) does not grow

\[9\] In the Hamiltonian picture the momentum \( \pi_{ij} \) canonically conjugate to \( h_{ij} \) is \( \pi_{ij} = \sqrt{h}(K_{ij} - Kh_{ij}) \), where \( h \) is the determinant of \( h_{ij} \) and \( K \) is the trace of \( K_{ij} \)
faster than exponential, though with sufficiently rapid exponential growth it might be impractical to give high enough resolution (small $h$) to keep the error term small for the desired run-time. Another (and somewhat related) reason why stable codes could fail, and which actually appears to be the problem in most free evolution schemes, are “constraint violating modes” discussed next.

3. Curb truncation-error-induced growth of constraints

Constraint violating modes (CVMs) are continuum solutions to the subset of Einstein equations that are evolved during a free evolution, but do not satisfy the constraint equations. These constraints could either be the Hamiltonian and momentum constraints inherent to the Einstein equations, or constraints arising from first order reductions or similar redefinitions of the underlying fields. Note that truncation error is not a CVM by this definition, however since in general truncation error will not satisfy the constraints it will be a source of CVMs in any free evolution scheme. For CVMs to be benign their growth rate must be sufficiently small to remain of comparable magnitude to truncation error during the evolution. At present only two formulations of the field equations appear to have this desired property “off the constraint manifold” for binary black hole evolutions—generalized harmonic coordinates with constraint damping [97], and variants of BSSN with appropriate gauge choices and methods for dealing with the black hole singularities [98, 99]. These two approaches will be described in more detail in Secs. III C and III D. Constraint damping adds a particular function of the constraint equations to the Einstein equations to try to curb the growth of CVMs for solutions close to the desired one. This is not a very new idea [102, 121, 122, 124, 125, 126, 132, 137, 174, 175, 176, 177, 178] though the particular method that works for harmonic evolution was only recently proposed by Gundlach et al. [104]. They were able to prove that their damping terms could curb all finite-wavelength constraint violating perturbations of Minkowski space. There is no mathematical proof that this should work for the binary black hole problem, and the evidence that the CVMs are adequately under control is entirely empirical. However, experience suggests there may never be a “black box” solver for the Einstein equations applicable to solving for arbitrary spacetimes; rather, schemes need to tailored to the particular scenario, and constraint damping is probably no exception. The nature of the evolution of the constraints in BSSN is even less well understood.

Given all the problems with the constraints, an obvious alternative would be constrained evolution, whereby the constraints are solved at each time step in lieu of a subset of evolution equations. Such methods have worked very well in symmetry reduced situations, though with the exception of [179] have not yet been attempted in 3D. Part of the reason is that solving the constraints involves solving elliptic equations, which many people in the community have been reluctant to attempt. Also, it is not clear in a general 3D setting which degrees of freedom to constrain, and which to freely evolve. In [179], a spherical polar coordinate system is used, which does allow for a “natural” decomposition into free vs. constrained variables; such a coordinate system is not well adaptive to studying binary black hole spacetimes. Several years ago Andersson and Moncrief [180] discussed an elliptic-hyperbolic formulation of the field equations that appears to be ideally suited for 3D constrained evolution, though to date no implementations of this system have been carried out. A related idea is constraint projection (similar to “divergence cleaning” in the solution of Maxwell’s equations), whereby a free evolution system is used, then periodically the constraints are re-solved, modifying a subset of the variables accordingly. This technique was shown to have promise in a single black hole spacetime [168], though in that code excision boundary problems (apparently) prevented long-time stable evolution. In [181] a Langrange multiplier method was proposed to optimally project out the constraints; it was successfully implemented for scalar field evolution, though has not yet been applied to the full Einstein equations. One might think that another option for dealing with the constraints is at the numerical level via something akin to the constrained transport [182] scheme used in some magneto-hydrodynamic codes, however Meier [183] showed that similar finite-difference based techniques will not work for the Einstein equations due to the non-linearity of the equations.

4. Provide well behaved dynamical coordinates conditions

It almost goes without saying that covering the spacetime manifold with a well behaved, non-singular coordinate system is a necessary condition for stable evolution. The difficulty is that in a Cauchy evolution the dynamics of the fields describing the geometry are intimately linked to the coordinates, and thus when solving for a new spacetime where the future geometric structure is unknown, the future behavior of the coordinates is just as uncertain. A large number of analytic solutions discovered throughout the history of relativity have, in their original form, been riddled with coordinate pathologies (the most famous example of course being the
event horizon of the Schwarzschild solution); dealing with them involved first understanding the nature of the pathology, then constructing a coordinate transformation to remove it. *In principle* this is an approach that could be applied in a numerical evolution: evolve to the point of a coordinate singularity, understand it, apply a coordinate transformation to remove it, and continue the evolution. However, given the nature of a numerical solution, i.e. discrete meshes of numbers representing either field values or coefficients of basis functions, this would be a very challenging endeavour in all but the simplest spacetimes. Thus, the universal approach in numerical relativity to try to avoid coordinate problems has been to devise coordinate conditions that typically either make the coordinates satisfy certain properties (e.g. constant mean curvature, or CMC slicing where \( t = \text{const} \) is a space of constant mean curvature, or harmonic coordinates described in Sec. III C), or conditions that force the variables to satisfy certain constraints (e.g. the unit-determinant condition on the conformal metric in the BSSN approach discussed in Sec. III D). Coordinate conditions usually take the form of algebraic or differential operators acting on the “gauge” fields of the formalism, which are most commonly the lapse and shift. It is beyond the scope of this article to describe the numerous coordinate conditions proposed over the years related to the binary hole problem (see [101, 109, 113] for more details), though in Secs. III C and III D we will described those that have been instrumental in the recent successful binary black hole simulations.

5. Specify good outer boundary conditions

“Good” outer boundary conditions for evolved fields in a simulation must have three properties: (1) be mathematically well posed, (2) be consistent with the constraints, and (3) be consistent with the physics being modeled, which here is asymptotic flatness\(^{10}\) and no incoming gravitational radiation. The class of boundary conditions (3) will form a subset of (2), which in turn is a subset of conditions (1). A common approach is to apply either exact or some approximation to *maximally dissipative* boundary conditions, where the incoming characteristics of all fields normal to the boundary are set to zero. Though mathematically well-posed this in general is neither consistent with the constraints nor prevents outgoing waves of the solution from being reflected back. Much effort has been spent over the past several years devising constraint preserving boundary conditions (CPBCs) for various formulations [176, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201]. By themselves CPBCs do not alleviate the problem of spurious incoming radiation, and more recently research in CPBCs has begun to focus on subsets of CPBCs that do address this issue.

An alternative approach to outer boundary conditions is to extend the computational domain to infinity, where the exact Minkowski spacetime boundary conditions can be placed. As mentioned in Sec. III B 1 Cauchy-characteristic matching effectively extends the domain to future null infinity. The matching procedure is non-trivial however, and this technique has yet to be applied to a binary black hole merger scenario. Another option is to compactify the coordinates to spatial infinity, which is the approach used in the generalized harmonic evolutions in [97]. This rather straight-forwardly solves all issues (1)-(3), though introduces potential numerical complications in that outgoing waves suffer an ever increasing blue shift as they travel toward the outer boundary \(^{11}\). Either increasing resolution and hence computational resources must be used to resolve the waves, or once they have passed the desired wave extraction radius be allowed to blue-shift to coarse resolution. With the latter option the numerical technique must therefore be robust to the introduction of high-frequency solution components; in the generalized harmonic evolution code this is achieved using Kreiss-Oliger style dissipation [202]. Note that this kind of dissipation is not akin to artificial viscosity sometimes used in hydrodynamical simulations, as Kreiss-Oliger dissipation modifies the difference equations at the level of the truncation error terms, and thus converges away in the continuum limit.

A couple of alternative methods of compactification include conformal compactification [140, 203, 204], and using asymptotically hyperboloidal or null slices [205, 206].

\(^{10}\) For the purposes of modeling the local geometry of a merger and extracting the resultant gravitational waves in the far-zone there is little practical distinction between an asymptotically flat versus Friedman-Robertson-Walker universe. The effects of a wave propagating across cosmological distances in an expanding universe can readily be accounted for analytically, as the wave amplitude decays as \( 1/D_L \) and its wavelength increases by a factor \( 1 + z \), where \( z \) is the redshift and \( D_L \) the luminosity distance to the source—see for example [182].

\(^{11}\) The blue shift is infinite in the limit, though it would take an infinite amount of time for the waves to reach the boundary.
6. Deal with black hole singularities

By the singularity theorems of Hawking and Penrose [24, 25] we know that all black holes contain true (geometric) singularities, which in a simulation will manifest as various field quantities diverging as the spacetime slice approaches the singularity. Infinities cannot be dealt with in a numerical code, and must be “regularized” in some fashion. The two contemporary successful approaches to deal with the singularities are excision and punctures. Both techniques rely on the causal property of a black hole spacetime that no information can flow out of the event horizon, and that cosmic censorship is valid, namely that all the geometric singularities that might exist in the spacetime are always inside the event horizon. If cosmic censorship were violated in a particular evolution, the codes would “crash”; thus a stable evolution is confirmation that in that instance cosmic censorship was not violated.

With excision, a 2-sphere inside the black hole and enclosing the singularity is designated as a boundary of the computational domain. By the assumed causal properties of the black hole there will always exist a class of such boundaries where all characteristics of the fields are directed toward the boundary, i.e. out of the computational domain. Mathematical theory (and common sense) says one is only allowed to place boundary conditions on the incoming components of fields satisfying hyperbolic equations. Thus no boundary conditions are specified on the excision surface; rather, the difference equations are simply solved there. The formal definition of a black hole event horizon is the boundary of the causal past of future null infinity, which is not a local property of spacetime and cannot be found during evolution. Instead, the apparent horizon—a marginally outer-trapped surface, which is surface from which “outward” traveling photons have zero expansion—is used to determine where to excise. Excision surfaces on or inside the apparent horizon can also satisfy the requirement that all field-characteristics are directed outside of the domain, since, if cosmic censorship holds, the apparent horizon will always be inside the event horizon (see for example [207]).

Originally, a puncture was the singular point inside a maximally extended vacuum black hole spacetime representing the spatial infinity reached by a conformally mapped slice passing through an Einstein-Rosen bridge into a second asymptotically flat universe [208]. Therefore a puncture is a coordinate rather than geometric singularity. The manner in which the metric diverges approaching the puncture is known analytically, and can be factored out. Punctures were originally used to construct initial data for the binary black hole problem [208], though soon afterwards it was shown that punctures can be used in evolution [156]. The metric at the puncture was regularized by diving out a time-independent conformal factor, however the extrinsic curvature was not regularized. Thus, to avoid problems this was anticipated to cause, the punctures were placed “between” grid points, and coordinate conditions were chosen to make the shift vector zero at the punctures so that derivatives of the extrinsic curvature across the punctures would not be needed. The vanishing of the shift vector implies the puncture locations are fixed in the grid. Maximal slicing was used for the lapse. The breakthrough in puncture evolutions discovered recently is to relax the condition on the shift vector, allowing the punctures to move through the domain. At the same time the slicing condition is altered to force the lapse to zero at the puncture, essentially “freezing” evolution there (see Sec. [111] for more on these coordinate conditions). This, remarkably, causes the codes to remain stable despite the irregular nature of the solution about the punctures. There have been several studies since attempting to understand geometrically what a moving puncture represents [204, 210, 211, 212, 213]: a couple of competing viewpoints at present are that 1) the puncture remains attached to spatial infinity of the alternate universe [213], and 2) the spacetime slice quickly evolves so that the alternate universe is “pinched-off”, and the puncture effectively becomes a single excised point inside the black hole [204, 210, 212]. Though from the perspective of seeking solutions to the field equations exterior to the event horizons of the black holes, the question of what a puncture represents is academic.

7. Provide consistent and relevant initial data

The initial data problem for binary black holes is not trivial. First, the initial conditions must satisfy the constraints, which typically evolves solving systems of coupled, non-linear elliptic equations. Second, providing astrophysically relevant initial data is quite challenging, as for practical considerations the evolution must

12 In a finite difference code this implies replacing centered derivative operators with “sideways” operators as appropriate to avoid referencing regions of the domain inside the excision boundary.
begin within several or tens of orbits before merger. This implies that there should already be a non-negligible amount of gravitational radiation from the prior inspiral of the black holes present in the initial data. Also, the closer the black holes are the more difficult it becomes to unambiguously associate relevant orbital parameters to the spacetime, for example the orbital eccentricity, binary separation, orbital frequency, etc. It is beyond the scope of this article to describe these issues—see [214, 213, 216] for review articles on contemporary initial data construction methods, and [217, 218, 219, 220] for suggestions to incorporate realistic initial conditions motivated by post-Newtonian expansions.

C. Generalized harmonic evolution

*Generalized harmonic evolution,* as its name implies, is an evolution scheme based on a generalization of harmonic coordinates. Harmonic coordinates are a set of gauge conditions that require each spacetime coordinate \(x^a\) to independently satisfy the covariant scalar wave equation:

\[
\Box x^a = \frac{1}{\sqrt{-g}} \partial_b \left( \sqrt{-g} g^{bc} \partial_c x^a \right) \equiv 0, \tag{12}
\]

where \(g\) is the determinant of the spacetime metric \(g\). The use of these coordinate conditions have a long and celebrated history in relativity, including DeDonder’s analysis of the characteristic structure of general relativity [221], Fock’s study of gravitational waves [222], and proofs of uniqueness and existence of solutions to the field equations by Choquet-Bruhat [223] and Fischer and Marsden [224]. In fact, as early as 1912 Einstein used harmonic coordinates, then known as isothermal coordinates, in his search for a relativistic theory of gravitation [225]. One of the key properties of harmonic coordinates that make them so useful in these studies is that when (12) is substituted into the field equations, the principal part of the resultant equation for each metric element becomes a scalar wave equation for that particular metric element, with all non-linearities and couplings between the equations relegated to lower order terms. This has obvious benefits for formal analysis of the field equations, and is also a natural system to study the radiative degrees of freedom in the theory. Also, given that there are simple and effective numerical solution techniques available to solve wave equations, it would seem that harmonic coordinates would be a natural starting point for a numerical code.

However, only recently in numerical relativity have harmonic coordinates been used as the basis for numerical evolution schemes [226, 227, 228, 229, 230, 231], meaning discretizing the field equations after the harmonic conditions have been used to re-express the system as a set of wave-like equations. Prior to this harmonic coordinates had been advocated and used within the more traditional ADM space-plus-time formulation of the field equations [112, 124, 222, 223, 224, 225], where harmonic gauge (or variants of it) are imposed via conditions on the lapse function and shift vector. In such a decomposition the wave-like character of the field equations in not manifest, and the primary reason quoted for using harmonic gauge (in particular harmonic time slicing) was for its geometric “singularity avoiding” properties. However even within ADM evolutions harmonic coordinates were seldom used due to the notion that they would generically lead to the formation of “coordinate shocks” [235, 236, 237, 238, 239]. An in-principle solution to this problem noted by Garfinkle [226] (and see an earlier discussion of this by Hern [237]) was to use *generalized harmonic coordinates* (GHC), first introduced by Friedrich [240]. Here, a set of arbitrary source functions are added to (12):

\[
\Box x^a \equiv H^a. \tag{13}
\]

To see that this can avoid coordinate pathologies, note that (13) can be regarded as a definition of the source functions; in other words, take any metric in any (well behaved) coordinate system, and (13) tells one what the corresponding source functions for the metric in GHC are. When imposing GHC in a Cauchy evolution, the \(H^a\) must be treated as independent functions to allow (13) to reduce the principal parts of the field equations to the desired wave-like equations. Thus additional evolution equations must be supplied for \(H^a\) to close the system, and so the issue of finding well-behaved coordinates for a particular dynamical spacetime becomes one of finding the appropriate evolution equations for \(H^a\).

For concreteness, below an explicit form of the Einstein equations in GH form with constraint damping terms will be given, using the covariant metric elements and covariant source functions (\(H_a = g_{ab} H^b\)) as the fundamental variables. This is certainly not the only way to proceed—for a symmetric hyperbolic first order reduction see [229] (and see [241] for how the constraints introduced via auxiliary variables are kept under control), and versions using the densitized contravariant metric elements see [227, 230, 242]. Consider the
Einstein equations in trace-reversed form

\[ R_{ab} = 4\pi (2T_{ab} - g_{ab}T), \]  

(14)

where \( R_{ab} \) is the Ricci tensor

\[ R_{ab} = \Gamma^{d}_{ab,d} - \Gamma^{d}_{db,a} + \Gamma^{e}_{ab}\Gamma^{d}_{cd} - \Gamma^{e}_{db}\Gamma^{d}_{ca}, \]  

(15)

\( \Gamma^{g}_{ab} \) are the Christoffel symbols of the second kind

\[ \Gamma^{g}_{ab} = \frac{1}{2}g^{ge} [g_{ae,b} + g_{be,a} - g_{ab,e}], \]  

(16)

\( T_{ab} \) is the stress energy tensor with trace \( T \), and a comma is used to denote partial differentiation. Using the definition of GHC (13) and its first derivative (14) can be written out explicitly as

\[
\begin{align*}
1/2 g^{cd} g_{ab,cd} + (17) \\
g^{d}_{(a}g_{b)d,c} + H_{(a,b)} - H_{d}^{f}_{ab} + \Gamma^{d}_{bd}\Gamma^{d}_{ac} + \kappa [n_{(a}C_{b)} - \frac{1}{2}g_{ab}n^{d}C_{d}] \\
= -8\pi \left( T_{ab} - \frac{1}{2}g_{ab}T \right). 
\end{align*}
\]

(19)

(20)

Line (17) shows the principal, hyperbolic part of the equations, line (18) are the rest of the terms coming from (14) where all the couplings and non-linearities reside, line (19) are the constraint damping terms with adjustable parameter \( \kappa \) and unit timelike vector \( n^{a} \) normal to \( t = \text{const.} \) hypersurfaces\(^{13} \), and line (20) contains the coupling to matter. Here, the constraints are simply the definition of GHC

\[ C_{a} = g_{ab} (H^{a} - \Box a^{a}), \]  

(21)

and are thus zero for any solution of the field equations. The relationship between the GH constraints and the more familiar form of the constraints of the Einstein equations, written as a one-form \( \mathcal{M}_{a} \)

\[ \mathcal{M}_{a} = (R_{ab} - \frac{1}{2}g_{ab}R - 8\pi T_{ab})n^{b}, \]  

(22)

where the time-component \( \mathcal{M}_{a}n^{a} \) is the Hamiltonian constraint and the momentum constraints are the components of the spatial projection \( \mathcal{M}_{a}(\delta_{a}^{b} + n^{a}n_{b}) \), is

\[ \mathcal{M}_{a} = \nabla_{(a}C_{b)}n^{b} - \frac{1}{2}n_{a}\nabla_{b}C^{b}. \]  

(23)

Furthermore, one can show that if the metric is evolved using (17-20), the constraints will satisfy the following evolution equation

\[ \Box C^{a} = -R^{a}bC^{b} + 2\kappa \nabla_{b} \left[ n^{(b}C^{a)} \right]. \]  

(24)

From this it easy to see that, at the continuum level, a solution that initially satisfies the constraints will always do so if constraint-preserving boundary conditions are used during evolution. Part of the constraint damping modification in (24)—namely the term proportional to \( n^{b}\nabla_{b}C^{a} \)—is a wave-equation damping term, so one might reasonably then expect that (24) will not admit exponentially growing solutions given small (truncation-error-sourced) initial conditions. This “expectation” has been proven for small, finite-wavelength boundary conditions.

\(^{13} \) In [104] it was suggested that an arbitrary timelike vector can be used in the constraint damping terms, though in all situations studied to date \( n^{a} \) has been chosen as the hypersurface normal unit timelike vector.
constrain-violating perturbations of Minkowski spacetime \[104\], though not yet for general spacetimes.

1. Source function evolution

To close the system of equations \[17\]-\[20\], an additional set of evolution equations must be specified for the source functions, written schematically as

\[
\mathcal{L}_a H_a = 0 \quad \text{(no summation).} \tag{25}
\]

\(\mathcal{L}_a\) is a differential operator that in general can dependent upon the spacetime coordinates, the metric and its derivatives, and the source functions and their derivatives. The source functions directly encode the coordinate degrees of freedom of general relativity, as can be seen by writing the definition of GHC \[13\] in terms of ADM variables \[S\]:

\[
H_a n^a = -K - \partial_\nu (\ln \alpha) n^\nu \tag{26}
\]

\[
H_b h^{ab} = -\bar{\Gamma}^i_{jk} h^{ij} + \partial_j (\ln \alpha) h^{aj} + \frac{1}{\alpha} \partial_b \beta^a n^b, \tag{27}
\]

where \(\bar{\Gamma}^i_{jk}\) is the connection associated with spatial metric \(h_{ij} \equiv g_{ij} + n_i n_j\). Thus, the temporal source function \(H_a n^a\) is related to the time derivative of the lapse \(\alpha\), whereas a spatial source function \(H_b h^{ab}\) is related to the time derivative of the corresponding component of the shift vector \(\beta^a\). Not much research has been done on finding source function evolution equations to achieve a particular slicing or satisfy some coordinate conditions directly within the GH framework, though the above relationship between \(H^a\) and the lapse and shift allows many of the ideas developed over the years for ADM evolutions to be adopted in a GH evolution \[228, 243\]. We end this section by showing one example of a set of source evolution equations, used in \[97\]:

\[
\Box H_t = -\xi_1 \frac{\alpha - 1}{\alpha \eta} + \xi_2 H_{t,\nu} n^\nu, \quad H_t = 0. \tag{28}
\]

This equation for \(H_t\) is a damped wave equation with a forcing function designed to prevent the lapse \(\alpha\) from deviating too far from its Minkowski value of 1, which helps alleviate an apparent instability in the code of \[228\] that sets in when the lapse drops close to zero inside a black hole\[14\]. In \[28\] the parameter \(\xi_2\) controls the damping term, and \(\xi_1, \eta\) regulate the forcing term. Ranges of useful parameter values are discussed in \[95\].

D. BSSN with ‘moving punctures’

The BSSN formulation of the field equations \[177, 234, 244\] begins with the ADM \[S\] decomposition of spacetime, then continues by performing a York-Lichnerowicz-like conformal decomposition of the spatial metric and extrinsic curvature \[245\]. The conformal metric is defined via

\[
\hat{h}_{ij} \equiv e^{-4\phi} h_{ij} \tag{29}
\]

and is chosen to have unit determinant, so that

\[
e^{4\phi} = h^{1/3}, \tag{30}
\]

where \(h\) is the determinant of \(h_{ij}\). Continuing, the trace \(K\), and conformal, trace-free part of the extrinsic curvature \[10\]

\[
\hat{A}_{ij} \equiv e^{-4\phi} (K_{ij} - \frac{1}{3} h_{ij} K), \tag{31}
\]

\[14\] Note that \(\alpha\) is not an independent variable in the formalism—in the code it is replaced by its definition in terms of the metric \(g_{ab}\).
are treated as fundamental variables. The final ingredient in the BSSN formalism is to also evolve the conformal connection coefficients

$$\tilde{\Gamma}^i \equiv \tilde{h}^{jk} \tilde{\Gamma}^i_{jk} = -\tilde{h}^{ij}_{\phantom{ij}j}$$

independently, where \(\tilde{\Gamma}^i_{jk}\) is the Christoffel symbol of the conformal spatial metric. In summary then, \(\phi, \tilde{h}_{ij}, K, \tilde{A}_{ij}, \Gamma^i, \alpha, \beta\) are the fundamental variables of the BSSN formalism. The evolution equations for \(\phi, \tilde{h}_{ij}\) and \(\Gamma^i\) derive from their definitions

$$\frac{d}{dt} \phi = \frac{1}{6} \alpha K,$$

$$\frac{d}{dt} \tilde{h}_{ij} = -2\alpha \tilde{A}_{ij},$$

$$\frac{\partial}{\partial t} \tilde{\Gamma}^i = 2 \tilde{A}^{ij} \alpha, + 2\alpha \left( \tilde{\Gamma}^i_{jk} \tilde{A}^{kj} - \frac{2}{3} \tilde{h}^{ij} K_{j} - \tilde{h}^{ij} S_{j} + 6 \tilde{A}^{ij} \phi,_{j} \right)$$

$$+ \frac{\partial}{\partial x^j} \left( \beta^j \tilde{h}^{ij} \phi,_{i} - 2 \tilde{h}^{m(j} \beta^i_{i)},_m + 2 \tilde{A}^{ij} \beta^i_{j} \right),$$

and the evolution equations for \(K\) and \(\tilde{A}_{ij}\) come from the Einstein equations

$$\frac{d}{dt} K = -\tilde{h}^{ij} D_j D_i \alpha + \alpha \left( \tilde{A}_{ij} \tilde{A}^{ij} + \frac{1}{3} K^2 \right) + \frac{1}{2} \alpha (\rho + S),$$

$$\frac{d}{dt} \tilde{A}_{ij} = e^{-4\phi} \left[ - (D_i D_j \alpha)_{TF} + \alpha (R_{ij}^{TF} - S_{ij}^{TF}) \right] + \alpha (K \tilde{A}_{ij} - 2 \tilde{A}_{il} \tilde{A}^{il}_{j}),$$

with

$$R_{ij} = \tilde{R}_{ij} + R_{ij}^{\phi},$$

$$R_{ij}^{\phi} = -2 \tilde{D}_i \tilde{D}_j \phi - 2 \tilde{h}_{ij} \tilde{D}^l \tilde{D}_l \phi + 4 (\tilde{D}_i \phi)(\tilde{D}_j \phi) - 4 \tilde{h}_{ij} (\tilde{D}_l \phi)(\tilde{D}_l \phi),$$

$$\tilde{R}_{ij} = \frac{1}{2} \tilde{h}^{lm} \tilde{h}_{ij,lm} + \tilde{h}_{k(i,j)} \tilde{\Gamma}^{k} + \tilde{\Gamma}^{k} \tilde{\Gamma}_{(ij)k} + \tilde{h}^{lm} \left( 2 \tilde{\Gamma}^{k} \tilde{\Gamma}_{(ij)km} + \tilde{\Gamma}^{k} \tilde{\Gamma}_{klm} \right)$$

and matter projections

$$\rho = n_a n_b T^{ab},$$

$$S_i = -\tilde{h}_{ia} n_b T^{ab},$$

$$S_{ij} = h_{ia} h_{jb} T^{ab}.$$  

The gauge variables \(\alpha\) and \(\beta\) are freely specifiable. In the above the operator \(d/dt\) is defined to be

$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} - L_\beta,$$

where \(L_\beta\) is the Lie derivative with respect to the shift vector \(\beta^i\) (and note that \(\tilde{h}_{ij}\) and \(\tilde{A}_{ij}\) are tensor densities of weight \(-2/3\), \(D_i(\tilde{D}_j)\) is the covariant derivative operator with respect to \(h_{ij}(\tilde{h}_{ij})\), and \(TF\) denotes the trace-free part of the expression. The BSSN equations listed above were taken from [177]; some of the actual implementations use slightly different variables (for example \(\chi \equiv e^{-4\phi}\) is used instead of \(\phi\) in [98]), differ in whether and/or how certain algebraic constraints in the formalism are enforced (such as the trace-free nature of \(\tilde{A}_{ij}\) or that \(\tilde{h}_{ij}\) has unit determinant), replace undifferentiated occurrences of \(\tilde{\Gamma}^i\) with its definition [32], or adds multiples of the constraints inferred by [32] to the evolution equation for \(\Gamma^i\) by [162, 246, 247, 248].

There are several reasons often quoted as motivation behind the BSSN formalism. First, the conformal decomposition in part separates the extrinsic curvature into “radiative” versus “non-radiative” degrees of freedom (though within the York-Lichnerowicz formalism it is the transverse trace-free part of the extrinsic curvature that represents the radiative degrees of freedom). Second, the constraint equations are used to eliminate certain terms from the “bare” evolution equations (in particular the Hamiltonian constraint is used to eliminate a Ricci scalar term from the evolution equation for \(K\), and the momentum constraints to eliminate a divergence of \(\tilde{A}_{ij}\).
term from the evolution equation of $\tilde{\Gamma}^i$, and so in a sense this is a partially constrained evolution system \cite{249}. Third, with appropriate gauge conditions the BSSN system of equations is hyperbolic \cite{145, 250, 251, 252, 253}. An important step in achieving hyperbolicity is treating the connection functions $\tilde{\Gamma}^i$ as independent quantities, which makes the principle part of the differential operator acting on the conformal metric in (40) elliptic. Incidentally, this is exactly what would be done if one were to express the spatial conformal metric in generalized harmonic form, with $\tilde{\Gamma}^i$ being the source functions.

1. Moving punctures

An important element in achieving stable evolution of binary black hole spacetimes with the BSSN formulation is using coordinates that allow the punctures hiding the black hole singularities to move through the grid, yet do not allow any evolution at the puncture point itself (i.e., the lapse is forced to go zero at the puncture, though not the shift vector, hence the “frozen” puncture can be advected through the domain). The conditions that have so far proven successful are modifications to the so-called 1+log slicing and Gamma-driver shift conditions \cite{123, 254}:

$$\frac{d}{dt} = -2\alpha K$$
$$\partial_t \beta^i = \xi B_i, \quad \partial_t B_i = \chi \partial_t \tilde{\Gamma}^i - \eta B^i - \zeta \beta^j \partial_j \tilde{\Gamma}^i.$$ (46)

In the above, $\xi, \chi, \eta$ and $\zeta$ are parameters (that are required to be within certain ranges for stable evolution, though do not require fine-tuning); a couple of examples for typical choices: ($\xi = 3\alpha/4, \chi = 1, \eta = 4, \zeta = 1$) \cite{255} and ($\xi = 1, \chi = 1, \eta = 1, \zeta = 0$) \cite{256}. Common initial conditions are $\beta^i = B^i = 0$, and $\alpha = 1/\psi_{BL}^2$, where $\psi_{BL} = 1 + \sum_i m_i/2|\vec{r} - \vec{r}_i|$ is the Brill-Linquist conformal factor for the initial data containing black holes with mass parameter $m_i$ at coordinate location $\vec{r}_i$.

It is uncertain exactly why these coordinate conditions work as well as they do (similar to why the rather ad-hoc equations used in the generalized harmonic scheme shown in \cite{258} improve the evolution); an alternative way of phrasing this is that it is not known why fixed puncture evolutions are prone to instabilities. Recently in \cite{239} it was suggested that 1+log slicing could generically lead to the formation of gauge shocks near the punctures as anticipated in \cite{235}, and that these have not yet been observed in current 3D simulations due to poor resolution about the punctures (though again, as long as stability can be maintained this in theory is not problematic for studying the geometry exterior to the horizon).

E. Comparison of the two techniques

After discussion of the two evolution formalism, generalized harmonic and BSSN, a “required” section deals with a comparison of the methods. That section is here, though there really is not much to say on the matter. Personal preferences and aesthetics aside, both methods are capable of finding discrete solutions describing similar physical processes within the context of the same theory—general relativity—and thus are equivalent from a scientific perspective. In terms of technical issues, one could argue that moving punctures are much easier to get working than excision. However, this is more an issue of dealing with black hole singularities, and in principle either method could be implemented within either formalism. A technical issue of some relevance to numerical implementation is that there is (presently) no known fully first order, symmetric hyperbolic reduction of the BSSN equations, which would be a requirement for a spectral implementation using the methods of the Caltech/Cornell group.

F. Numerical algorithms

It is beyond the scope of this article to delve into computational issues involved in solving the field equations—here a few references to related material in the literature is given. With the exception of the Caltech/Cornell pseudo-spectral code \cite{257, 258}, all contemporary binary black hole evolution codes use finite-difference methods (for a broader view of the use of spectral methods in relativity see \cite{259}). The complexity of the field equations and the physical set-up of the binary black hole problem requires solution in a parallel computing environment,
and adaptive mesh refinement (AMR) to adequately resolve all the relevant length scales (the only code at present not employing AMR is the LazEv code \cite{260}, however there a non-linear “fisheye” coordinate transformation is used to resolve the length scales in the vicinity of the binary). Some of the parallel/AMR software presently used is the Cactus Computational Toolkit \cite{261} with the Carpet thorn for AMR \cite{262}, paramesh \cite{263}, \textsc{PAMR/AMRD} \cite{264}, \textsc{HAD} \cite{265} and \textsc{BAM} \cite{266}. Descriptions of some of the more computational aspects of the merger codes can be found in \cite{228, 229, 231, 255, 260, 266, 268, 269}.

**IV. RESULTS FROM BLACK HOLE MERGER SIMULATIONS**

In this section some results from recent merger simulations are discussed. This is a rapidly evolving field, and much of what is said could be dated in short-order. Also, in many respects this is still a very young field, and though there has been a flurry of early results, systematic, in-depth studies are sparse. We break the discussion up into the following classes of binary: a) equal mass, minimal eccentricity and spin, b) unequal mass, minimal eccentricity and spin, c) equal mass, non-negligible spin, minimal eccentricity, d) equal mass, large eccentricity, minimal spin, and e) generic.

**A. Equal mass, minimal eccentricity and spin**

The equal mass, minimal eccentricity and spin case is one of the simplest configurations in that there is no precession of the orbital plane, and no recoil imparted to the final black hole. Thus, the key parameters that characterize the merger are essentially only the final mass and spin of the remnant black hole. Recent simulations \cite{58, 98, 99, 248, 270, 271, 272, 273} of this scenario have used either Cook-Pfeiffer \cite{274} or Bowen-York \cite{275} initial data. These results indicate that the energy emitted during the last orbit, plunge/merger and ringdown is $3.5\% \pm 0.2\%$ of the total total energy of the system, resulting in a Kerr black hole with $a = 0.69 \pm 0.02$. Based on the binding energy of the initial data configurations, or PN/EOB estimates of the energy radiated during the inspiral (see for example \cite{54, 274}), an additional $1.5\% \pm 0.2\%$ of the available energy is radiated prior to this, implying that an equal mass, non-spinning inspiral beginning at infinite radial separation loses $5.0\% \pm 0.4\%$ of its total rest-mass energy to gravitational waves during the entire merger event\textsuperscript{15}. As mentioned in the discussion in Sec.\textsuperscript{III B 7}, these families of initial data do not exactly capture the conditions of the equivalent astrophysical scenario, and though it is difficult at present to estimate precisely what the effects of this are, systematic studies suggest the artifacts are small. In particular, the initial data lacks the correct initial gravitational radiation content, though within roughly an orbital light-crossing time this “junk” radiation leaves the vicinity of the orbit and is quickly replaced by radiation emitted by the binary motion. The energy content of the junk radiation also appears to be negligible within the quoted uncertainties. Other noticeable “artifacts” in some of the cited simulation results is a small amount of orbital eccentricity (due to the choice of initial data having zero initial radial momentum), and the black holes are initially co-rotating for the Cook-Pfeiffer data presently in use; again, the effect on the waveforms appear to be small, and can also be removed without much effort \cite{272, 273}. There have been several suggestions for how the correct radiation content can be inserted into initial fields \cite{217, 218, 219, 220}, though none of these methods have yet been implemented.

For illustrative purposes, Fig.'s \textsuperscript{1-14} show some of the simulation results, taken from evolution of Cook-Pfeiffer initial data, described in detail in \textsuperscript{52}. Fig. \textsuperscript{1} is a plot of the orbital trajectory, Fig. \textsuperscript{2} shows the real component of the Newman-Penrose scalar $\Psi_4$ in the orbital plane (which far from the source represents the second time derivative of the “plus” polarization of the usual gravitational wave strain), Fig. \textsuperscript{3} shows the plus and cross polarizations of the waveform extracted on the axis normal to the orbital plane, and Fig. \textsuperscript{4} shows the instantaneous gravitational wave frequency (divided by 2) and energy flux versus time together with labels depicting some phases of the merger.

\textsuperscript{15} The uncertainties reflect the authors best conservative “guess” based on the various results published in the literature to date; the uncertainty in the PN inspiral value is that the results usually quoted are for the integrated energy up to the ISCO (innermost stable circular orbit), which only approximately corresponds to the “last” orbit of the numerical results.
Decomposed into a spin-weight 2 spherical harmonic basis, the waveform throughout the evolution is dominated by the quadrupole ($\ell = 2, |m| = 2$) component. The next leading order component ($\ell = 4, |m| = 4$) has an amplitude less than $1/10^{th}$ the quadrupole mode during the inspiral phase, growing briefly to $1/5^{th}$ of it near the peak of the emitted energy flux (and note that the energy content of a mode is proportional to the square of its amplitude) \cite{55,276}. Moreover, the quadrupole formula seems to describe the physics of gravitational wave production quite accurately throughout the orbital phase, in that the coordinate motion of the apparent horizons taken from the simulation and plugged verbatim into the quadrupole formula for two point-masses shows remarkable agreement with the full numerical waveform \cite{55,96}. Several studies have now also shown that the higher order PN and EOB methods can reproduce, to within the various uncertainties of the comparison (including numerical error, mapping parameters between the two descriptions, when to begin the comparison, etc.), up until close to merger \cite{55,56,57,58,59}. The most accurate study to date \cite{58} of exactly when the waveform from a particular PN approach begins to deviate from the numerical signal to within the errors of the simulation—0.3% in the phase and 1% in the amplitude over 18 cycles (9 orbits) of inspiral—showed that despite a relatively large amplitude disagreement of 7% between the restricted 3.5PN Taylor waveforms, the cumulative phase difference was 0.15 after 13 cycles, suggesting for the given numerical accuracy only the last 41.5 orbits of the inspiral would require numerical solution.

The transition from inspiral to ringdown does not last very long, only on the order of $10 - 20M$. There is also no noticeable “plunge” in the orbital motion from the time there are two distinct black holes to a single one (see Fig.1). However, in the Fourier transform of the waveform there seems to be a distinct change in the slope of the spectrum from the leading order PN prediction of $-7/6 \approx -1.2$ to somewhere between $-0.6$ and $-0.8$ before asymptoting to the dominant ringdown frequency \cite{55,60}.

The ringdown portion of the waveform is dominated by the the fundamental harmonic ($n = 0$) of the
FIG. 2: A depiction of the gravitational waves emitted during the merger of two equal mass black holes (specifically “d=19” Cook-Pfeiffer initial data [55]). Shown is a color-map of the real component of the Newman-Penrose scalar \( \Psi_4 \) multiplied by \( r \) along a slice through the orbital plane, which far from the blackholes is proportional to the second time derivative of the plus polarization (green is 0, toward violet (red) positive (negative)). The time sequence is from top to bottom, and left to right within each row. Each imagine is 25M apart, and a common apparent horizon is first detected at \( t = 529M \) (i.e., the “merger”), which is a little after the frame in row 5, column 3. In the first several frames the spurious radiation associated with the initial data, and how quickly it leaves the domain, is clearly evident. The width/height of each box is around 100M.
FIG. 3: The plus (left) and cross (right) polarizations of the waveform (multiplied by coordinate distance $r$ from the source, and by the total mass $M$ of the spacetime to non-dimensionalize) from the simulation shown in Fig.1, though here measured along the axis normal to the orbital plane. $t_{CAH}$ is the time when a common apparent horizon is first detected.

quadrupole moment ($\ell = 2, m = 2$) of the final black hole’s quasi-normal modes (QNMs) [55, 276]. The first two overtones of the quadrupole mode have amplitudes close to the fundamental mode, though they decay rapidly and are thus only discernible early on during the inspiral. Higher order multiple modes are also present, though as with the waveform itself at a much reduced amplitude compared to the quadrupole mode. An interesting property of the waveform is that from the moment of the peak in the flux onwards it can quite accurately be represented as a sum of QNMs. One reason why this is interesting is that here one would expect to be furthest into the regime where “non-linear effects” are most apparent, yet the wave can be described as coming from a linearly perturbed black hole. Proponents of the EOB approach predicted this behavior, and in fact have further suggested that with a sufficient number of QNM overtones and harmonics that the entire post-inspiral portion of the waveform may be described as a ringdown. This prescription has been carried out quite successfully for the extreme mass ratio case [62], and a range of non-spinning near equal mass mergers (with mass ratios from 1 : 1 to 4 : 1) [59], though may not be as straight-forward (or possible at all) for general configurations with spin.

B. Unequal mass, minimal eccentricity and spin

Relaxing the condition of equal mass from the configuration discussed in Sec.IVA1 several qualitative features of the merger and corresponding waveform change [255, 276, 277, 278]. First, the equal mass case maximizes the total energy emitted and also maximizes the final spin of the remnant black hole. To a good approximation the total energy radiated decreases by a factor $(\eta/\eta_1)^2$, and the final spin decreases linearly in $\eta$ via $a/M_f \approx 0.089 + 2.4\eta$, where the symmetric mass ratio $\eta = q/(1 + q)^2$, $\eta_1 = 1/4$, $q = M/m$ with $q \geq 1$, and $M_f = m + M$ [276, 278]. The second difference is that although the quadrupole mode still dominates in the waveform, certain higher multipole modes, in particular the $\ell = 3, |m| = 3$ component, become non-negligible [276]. The simple explanation for this is quadrupole-formula physics again: the reduced quadrupole moment of the effective source energy distribution now has higher multipole modes when expressed in terms of a spherical harmonic decomposition, and this will be reflected in the structure of the gravitational waves emitted. The final significant difference is that there is an asymmetric beaming of the gravitational radiation in the orbital plane due to the mass difference. If not for the inspiral this would average to zero over a complete orbit, however the inspiral, combined with fact that the radiation eventually ceases due to merger, results in the asymmetry. This imparts a “kick”, or recoil of the final black hole within the orbital plane to compensate for the net linear momentum carried away by the radiation. The dependence of recoil speed $v$ on mass ratio can be
approximated by the Fitchett formula \( v = A\eta^2 \sqrt{1-4\eta(1+B\eta)} \), with \( A \approx 1.20 \times 10^4 \) and \( B \approx -0.93 \).

The maximum is upwards of 175 km/s achieved around a mass ratio \( \approx 3 : 1 \). Note that the direction of recoil, which in this case occurs within the orbital plane, depends on the “initial” phase of the orbit, and thus for astrophysical sources can be regarded as a uniform random variable.

**C. Equal mass, non-negligible spin, minimal eccentricity**

Simulations of binary black hole spacetimes where the initial black holes have spin angular momentum have to date largely focused on equal mass black holes, and with the spin vectors having “non-generic” alignments: either both black holes were given spins aligned and/or anti-aligned with the orbital angular momentum or the spin vectors were set equal in magnitude but
opposite in direction and lying within the orbital plane. In all these configurations the net angular momentum vector is aligned with the orbital angular momentum, and thus there will be no precession of the orbital plane during evolution (this, ignoring radiation-reaction, is a consequence of conservation of angular momentum, which incidentally is also why precession does occur in cases where the orbital and net angular momentum are misaligned). A couple of exceptional studies examining more generic spin configurations have been presented in \(^{282,286}\).

There have not yet been the kinds of detailed or systematic studies of inspiral with spin as described for the non-spinning case in Sec IV A in terms of understanding the multipole structure of the waves, comparison with PN inspiral waveforms, extraction of the QNMs, etc. Though at least one can describe certain qualitative features of the merger. Also, one of the more sought-after answers has been to the question of what the range of magnitudes of the recoil velocity are when spin is included, and many of the above cited papers have recently addressed this. In the next subsection we will outline the basics of what changes during merger with spin, and the subsection following that will describe the recoil results.

1. Qualitative features of a merger of spinning black holes

When the black holes are given spin, several aspects of the merger can changed compared to the non-spinning, equal mass case. First, the net amount of energy/angular momentum radiated can change significantly, and consequently the final spin and mass of the remnant black hole. If the component of the net spin in the direction of the orbital angular momentum has the same (opposite) sign as the orbital angular momentum, then typically more (less) energy and angular momentum will be radiated compared to the non-spinning case. One explanation for this comes from the PN description of the spin-orbit interaction(see for example \(^{293}\)), where in the aligned (anti-aligned) case this interaction term results in a repulsive (attractive) force between the black holes, thus causing them to orbit for a longer (shorter) amount of time emitting more (less) net radiation before merger. As an example, \(^{284}\) (see also \(^{285}\)) considered the merger of two equal mass black holes with spin parameters \(a = 0.76\); for the case where the two spin vectors were aligned with the orbital angular momentum \(\approx \) 6.7% of the rest-mass energy was radiated, leaving a black hole with a spin of \(\approx 0.89\), whereas in the anti-aligned case \(\approx 2.2\%\) energy was emitted, and the final black hole had a spin of only \(\approx 0.44\) (in the direction of the orbital angular momentum). Components of spin in the orbital plane have a much smaller effect on the dynamics of the orbit, and consequently the amount of energy emitted; for example, the configurations that result in the largest recoil velocities described in the next section are equal mass, have zero-net spin angular momentum with the spin vectors lying in the orbital plane, and in this case the total energy and angular momentum radiated is very close to the amount for the equivalent non-spinning case \(^{283,289}\).

A second significant effect of spin in a merger is that the spin vectors and orbital angular momentum vector will in general precess, and near the time of merger by potentially large enough amounts to cause spin and orbital plane “flips”. In PN-terms this can be thought of as due to spin-spin and spin-orbit interactions \(^{293}\). A more Newtonian way of thinking about these interactions is that a spinning black hole effectively has a quadrupole moment, and thus the exterior gravitational field of the second black hole will in general exert a torque on the first black hole (and vice-versa), causing precession of the spins, and hence the orbital plane to conserve angular momentum (ignoring radiation). The only full numerical study to date of these effects were presented in \(^{282,286}\), though it will certainly not be long before more systematic studies are available from several groups.

2. Recoil velocities

Any property of an orbit resulting in an asymmetric beaming pattern for the gravitational waves could, via conservation of linear momentum, impart a kick to the remnant black hole. As discussed in Sec. IV B unequal masses produce such an asymmetry, and so can individual black hole spins. An obvious example is the asymmetry that would be produced by precession of the orbital plane, and if the precession time scale is shorter than the orbital and inspiral time scales (which it is near merger) then there will not be enough time to average the momentum beamed in any one direction to zero before merger, thus resulting in a net momentum flux in some direction. For near-equal mass mergers this can produce larger kicks than an unequal mass ratio alone—typically around several hundred km/s. A less obvious source of asymmetry, though one resulting in the largest kicks of up to 4000 km/s \(^{283,286,288,289,292}\), are equal mass black holes with equal but opposite spin vectors lying within the orbital plane. At a first glance this is a rather surprising configuration
for producing a large recoil, as it is not obvious where the asymmetry in the energy emission is. Thus it will be instructive to spend a bit more time describing this configuration in the next couple of paragraphs, and see how the kick can be understood as a frame dragging (or gravito-magnetic) effect. More technical explanations of the source of the kick can be found in [291, 292]. A discussion of the astrophysical implications of such large recoil velocities is deferred to Sec. V.

Consider the orbit depicted in Fig. 5 and imagine what the effect of rotation of black hole 2 on the motion of black hole 1 would be. The rotation of black hole 2 causes spacetime to be “dragged” about it following the right hand rule: grasp the spin vector with your right hand so that the thumb points in the direction of spin; then the direction in which the rest of your fingers curl about the axis indicates how spacetime is whirled about due to the spin of the black hole. Using this image, notice that at phases (A) and (C) within the orbit black hole 2 can not impart any effective velocity to black hole 1. However, at phase (B) the dragging of the spacetime caused by black hole 2 will cause black hole 1 to move in the negative $z$ direction, where $z$ is in the direction of the orbital angular momentum (i.e. it will move into the paper in the illustration), and the opposite at phase (D). The same analysis of the effect of the rotation of black hole 1 on black hole 2 shows that with this particular configuration of spins both black holes will at each instant have the same velocity induced in the direction normal to the orbital plane by the other black hole. In otherwords, one can think of this as causing the entire orbital plane to oscillate normal to the plane with the orbital frequency, or equivalently, the trajectory of each orbit will be tilted by equal but opposite angles relative to the original orbital plane. This normal-motion by itself does not produce much radiation, however, it does cause the more copious amounts of radiation caused by the circular orbital motion to get blue-shifted in one direction normal to the orbital plane while at the same time being red-shifted in the other. Averaged over one orbit, and ignoring radiation reaction, the net Doppler shift in any one direction is zero. However, as the orbital radius begins to shrink due to radiation reaction, the flux and the magnitude of the doppler shift increases until about the time of merger. Up until this time the net momentum radiated in the $z$ direction will be a function depending sinusoidally on the phase of the orbit, and slowly increasing in amplitude. Depending on where in the orbit the merger occurs ultimately determines the magnitude and direction of the kick normal to the plane.

Of course, the structure of spacetime in the vicinity of two spinning black hole just about to merger will be quite non-trivial, preventing one from unambiguously localizing the positions of the black holes or, when and where the radiation is being produced, so the preceding description of the production of a kick is somewhat cartoonish. However, one can apply it at face value to come up with an order of magnitude estimate for the kick which is in the correct ball park, as well as account for the linear dependence of the kick on the spin $a$ of the black holes and sinusoidal dependence on initial orbital phase, as follows. When far apart, at any instant in time black hole 1 (2) will not have any component of orbital angular momentum in the direction of the spin axis of black (2) 1. Thus, one can approximate the instantaneous velocity imparted by frame dragging as that which a particle, dropped from rest at infinity, would have falling toward the black hole. This is a particle on a zero angular momentum orbit, which will have an instantaneous $z$ velocity of

$$v_z(r, \theta) \approx \frac{2ma \sin(\theta)}{r^2}$$

(47)
where \( \theta \) is the angle relative to the spin axis of the black hole with spin parameter \( a \) and (total) mass \( m \), and \( r \) is the distance to the black hole. We have used the Boyer-Lindquist form of the Kerr metric in the above, and only kept the term to leading order in \( r \). From this expression one immediately sees where the linear dependence in \( a \) and sinusoidal dependence on the phase arises. To estimate the maximum possible kick velocity, note that this would be produced if the doppler-shifting of the radiation ceases at maximum velocity, i.e. when \( \sin(\theta) = 1 \). Assume that this occurs when the black holes merge, and this happens when the two black holes “touch”, so when \( r \approx 4m \). Thus,

\[
v_{z,\text{max}} \approx \frac{j}{8},
\]

where \( j \equiv a/m \) is the dimensionless spin parameter. The energy density \( \epsilon \) of a gravitational wave is proportional to the square of the wave frequency; thus the doppler shifted energy density will be proportional to \( (1 \pm 2v_{z,\text{max}})\epsilon \), and so the net momentum density radiated at this moment will be \( \delta P = 4v_{z,\text{max}}\epsilon \). This accumulates over the last part of an orbit, during which a total of \( E = \epsilon M \) \((M = 2m)\) of energy is emitted in gravitational waves. Therefore the net momentum radiated in \( z \) would be \( \delta P = 4v_{z,\text{max}}E \), giving an estimate for the maximum recoil speed \( \delta P/M \) as

\[
v_{\text{recoil, max}} \approx \frac{je}{2}.
\]

For a concrete number, take \( \epsilon \approx 0.01 \) (which is not too unreasonable given that the net energy emitted in the last orbit/merger/ringdown is around 0.035), then for an extremal \((j = 1)\) black hole this gives \( v_{\text{recoil, max}} \approx 1500 \text{km/s} \).

Note that a similar line of argument can give an intuitive understanding of the effective repulsive (attractive) force between binaries with spin axis aligned (anti-aligned) to the orbital angular momentum, again due to frame dragging. For empirical formulas giving the net recoil for various spin and mass configurations see \([288, 289]\).

### D. Equal mass, large eccentricity, minimal spin

An equal mass, zero spin but sizable eccentricity case was in fact the first complete merger event simulated in \([97]\). Adding eccentricity to the orbit was not intentional, rather this was an artifact of the initial data method, which is as follows. Boosted, highly compact concentrations of scalar field energy are chosen for the initial conditions, which then quickly undergo gravitational collapse and form black holes. Any remnant scalar field energy quickly accretes into the black holes or radiates away from the vicinity of the orbit, leaving behind, for all intents and purposes, a vacuum black hole binary spacetime. For a given initial separation of the scalar field pulses, a single boost parameter \( k \) controls the initial data—one scalar field pulse is placed at \((x, y, z) = (d, 0, 0)\) and given a boost \( k^t = (0, k, 0)\), while the second is placed at \((-d, 0, 0)\) and given a boost \((0, -k, 0)\). It turns out for sufficiently close separation (as used in the simulations) the resultant black hole binary has non-negligible eccentricity regardless of \( k \). Furthermore, probably due to the scalar field dynamics and accretion, the effective vacuum binary black hole orbit that could be ascribed to the black holes has an apoapsis much further out that the initial scalar field pulses. The consequences are that for the smaller values of \( k \) which result in strong interaction of the black holes early on, the black holes have much more kinetic energy than what black holes on a slow, adiabatic inspiral at the same separation would have. This offers an explanation for why the interesting threshold “zoom-whirl” behavior \([294, 295]\) explained in the next paragraph could be observed using this class of initial data, though at the time this was puzzling as it was (incorrectly) assumed the binary was in the adiabatic inspiral regime where any “radiation-reaction” effects would always force the binaries to be closer on average from one orbit to the next.

Imagine what should happen as the boost parameter \( k \) is varied. At one extreme, \( k = 0 \), there will be a head on collision; at the other, \( k \approx 1 \), the black hole trajectories will be deflected by some amount, though ultimately they will fly apart and separate. At intermediate values of \( k \) there should be a significant amount of close-interaction of the black holes, and then they will either merge or separate (to possibly merge at some time in the future). What appears to happen near the threshold value of \( k \) between these two distinct end-states is the black holes evolve toward an unstable near-circular orbit, remain in that configuration for an amount of time sensitively related to the initial conditions, then either plunge toward coalescence or separate. Specifically, for the single class of initial conditions examined in \([95, 96]\), the number of orbits \( n \) observed near threshold
FIG. 6: Left: The number of orbits $n$ versus logarithmic distance of the initial boost parameter $k$ from the immediate merger threshold $k^*$, for evolutions that did result in a merger. Results from three resolutions are plotted with characteristic mesh spacings $h$ (lowest resolution), $3h/4$ and $h/2$ (highest) to schematically illustrate convergence. For each resolution, a least-squares fit to the data is shown assuming the relation (50). Right: plots of the orbital motion from the two higher resolution simulations ($h/2$) tuned closest to threshold (only the coordinate motion of a single black hole—initially at positive $x$—is shown for clarity). The dashed curve is the case resulting in a merger, and the curve ends once a common apparent horizon is first detected, while for the solid curve the black holes separate again and here the curve ends when the simulation was stopped.

found to scale as

$$ e^n \propto |k - k^*|^{-\gamma} $$

with $\gamma \approx 0.34 \pm 0.02$—see Fig.'s 6 and 7 that depict this scaling relation for cases that merge, near-threshold orbits, a sample of the gravitational waves and the energy emitted as a function of $k$. Note that due to energy loss via gravitational radiation the threshold cannot be “sharp”, i.e. if the time $t_m(k)$ to merger is plotted as a function of $k$, this will not have a discontinuous step at $k = k^*$. There will be a maximum number of orbits $N$ for a given class of initial conditions, and what from a distance might appear like a step function will be resolved into a smooth transition over a region of size $\delta k \approx e^{-N/\gamma}$. Also note that the initial conditions need to be highly fine tuned to obtain even a few whirl-orbits. Thus, when close encounters of near equal mass black holes on hyperbolic or highly eccentric orbits occur in nature (which might occasionally happen in a dense environment such as a globular cluster), it is highly unlikely that it will be a near-immediate-threshold encounter.

The behavior just described is very similar to that of equatorial geodesic motion on a Kerr background, where geodesics near the threshold of capture approach the unstable circular geodesic orbits of the background spacetime, and exhibit similar scaling behavior of the number of orbits versus distance from threshold as in (50). In that case, the scaling exponent $\gamma$ is inversely proportional to the Lyapunov (or instability) exponent of the corresponding unstable circular geodesic [29], and in fact numerically has a value quite similar to that in the analogous equal mass scenario; for more information see the discussion in [96]. In contrast to near-equal mass binary black hole encounters in the universe, extreme mass ratio inspirals of a compact object into a supermassive black hole are expected to be numerous enough that a significant number of zoom-whirl type orbits will be seen with LISA [51, 52].
E. Generic

To date, there has not been any published systematic numerical studies of fully generic initial binary conditions, namely with varying mass ratios, spin magnitudes and orientations, and orbital eccentricity. The reason is simply that this field is still young, though given the rapid rate at which new results have been released over the past couple of years it should not be long before a rather detailed knowledge of a large class of astrophysically relevant merger spacetimes is available.

V. IMPLICATIONS, PROSPECTS AND QUESTIONS

This article concludes with a discussion of some of the implications of current results from the newly uncovered merger phase of the two body problem and what questions still need to be addressed. As before, the discussion is broken up into the rest-mass dominated regime of relevance to astrophysical black holes, and the kinetic energy dominated regime of the black hole scattering problem.

A. Black holes in our universe

Even though the merger phase has not yet presented any unexpected or bizarre phenomenology, that there are finally concrete numbers and waveforms associated with an ever growing set of initial conditions allows many consequences of the merger to be seriously explored. The key numbers are the amount of energy and net momentum lost to gravitational waves, and knowledge of the structure of the waves gives data analysts the information to build trust-worthy template banks. Note that the topics discussed in the next several sections hardly exhaust all the possible consequences and applications of black hole mergers, and certainly much of the future work on the “two body” problem in general relativity will include a thorough examination, numerical and otherwise, of mergers in astrophysical environments.
1. Consequences of radiated energy

An equal mass, non-spinning merger releases close to $\epsilon = 4\%$ of the rest-mass energy of the system into gravitational waves in the last plunge/orbit, merger and ringdown. With spins, depending on the relative alignment, this can increase or decrease by roughly a factor of two. For an unequal mass system with mass ratio $q = M/m$ ($q \geq 1$) the energy will be reduced by slightly less than a factor of $q$ for small $q$, though approaching a factor of $q^2$ for large $q$ (see Sec. IV.B). Thus for a “major merger”, with $q$ a few or less, a significant amount of the total gravitating energy of the system is effectively lost on a very short time-scale. To get an idea of just how short, define the light crossing frequency $f_{lc}$ of a system to be the frequency at which light could cross back and forth between the black holes separated by a distance $R$, i.e. $f_{lc} = c/2R$. Converting to the units given for the orbital frequency for an equal mass merger in [4], this is

$$f_{lc} \approx 51 \text{kHz} \left( \frac{R_s}{R} \right) \left( \frac{M_\odot}{M} \right)$$  \hspace{1cm} (51)

Note that this is the fastest frequency at which any causal process in the close vicinity of the binary could operate, and by comparison with [4] one can see that near coalescence ($R \rightarrow R_s$) the orbital frequency becomes a sizable fraction of this maximum possible frequency. The time-scale over which the final burst of energy is released will therefore be much shorter than almost any other astrophysical process that could be happening close to the binary. A likely non-vacuum environment for a binary is a circumbinary gas disk. Thus, a near-term effect of the passing gravitational waves on the particles in the disk is that essentially instantaneously the central mass they are orbiting will drop by a factor of [297]; said another way, if they were initially following circular orbits in a potential with mass $M$, they will suddenly be on eccentric orbits about a potential of mass $M(1 - \epsilon)$. Such a rapid perturbation of the disk could set up asymmetric waves and warping in the disk which could conceivably produce a weak but prompt electromagnetic counterpart to the merger event, though no detailed calculations of this have yet been performed. A related phenomena accompanies the secular evolution of the disk as the inspiral and merger occurs [297]. Early on in the inspiral phase the viscous timescale in the disk is shorter than the inspiral rate, allowing the inner edge of the disk with radius roughly twice the binary semi-major axis [298, 299] to “follow” the inspiral. However, eventually the inspiral time becomes much shorter than the viscous time, leading to an essentially non-accreting disk about the final black hole that is much further out than what a steady-state accretion disk would be (the innermost stable circular orbit). The subsequent inward migration of the disk and turn-on of accretion will produce a strong X-ray afterglow on a timescale of $\approx 7(1 + z)(M/10^8M_\odot)^{1.32}\text{yr}$ (with $z$ the cosmological redshift) that could be seen by future X-ray observatories [297, 300].

2. Consequences of radiated momentum

One of the more significant potential astrophysical consequences of a merger is when asymmetric radiation of linear momentum occurs, resulting in the recoil of the remnant black hole as discussed in section IV. The largest recoil speeds of several thousand kilometers per second for near-equal mass mergers are high enough to be able to eject the remnant from even the most massive galactic halo. If such large kicks are common, it would seem to be in contradiction with the observation of supermassive blackholes in most galaxies, and hierarchical structure formation scenarios [2, 3, 39, 40]. Note that the large recoils require each black hole to be spinning by a fair amount ($a > 0.3$) – X-ray observations of relativistic line broadening in a few AGN suggest spins are high, with $a > 0.9$ [304, 305, 306, 307]. As with the effects of energy loss discussed in the previous section, large recoils would also require major mergers, as the kick scales roughly as $1/q^2$ for large mass ratio $q$ [288, 289]. Recent estimates using the effective-one-body model, calibrated with some full numerical results, suggest that for mergers with spin magnitudes of $a = 0.9$, uniformly distributed spin configurations and mass ratios $1 \leq q \leq 10$, only about 3% (10%) of mergers result in kicks greater that $1000\text{km/s}$ ($500\text{km/s}$) [308]. Small kicks could still eject black holes from galaxies in the early universe when halos were much less massive, though the presence of supermassive black holes at the centers of most galaxies might be a more robust consequence of structure formation than naively thought in light of kicks. One study examined the effect of natal kicks in a scenario where supermassive black holes are formed in a primary halo via capture of intermediate mass black holes from surrounding secondary halos, and found that kick velocities imparted to mergers of the intermediate mass black holes had little affect on the growth of the supermassive black hole [301]. Another study following black hole formation through a simplified merger tree model found that even when a high probability of
ejection was assigned to merger events, still more than 50% of galaxies today retain their supermassive black holes. An examination of the effect of large recoil velocities on the predicted event rates for LISA to detect supermassive mergers suggested the event rate would drop by at most 60% if the seed black holes were light ($\approx 10^2 M_\odot$, from Population III stars) or by at most 15% if the seeds were heavier ($\approx 10^4 M_\odot$, from direct gas collapse of primordial disks).

The study mentioned in the preceding paragraph on the probability of kicks assumed a uniform probability distribution for the spin orientation in the progenitor binary system. However, the distribution of initial spins is likely highly non-uniform—have shown that in gas-rich mergers, torques from the surrounding gas will tend to align the black hole's spin vectors with the net angular momentum vector of the gas, a configuration which results in much more modest recoils of $< 200 \text{ km/s}$. Thus, only a small fraction of supermassive mergers would result in the black hole leaving the host galaxy. For those that do, there could be significant electromagnetic counterparts due to the recoil, as the black hole will drag the inner part (tens of thousands of Schwarzschild radii) of any accretion disk with it. However, given that such recoils would preferentially occur in gas-poor environments, accretion-related counterparts might also be correspondingly dim. A search for doppler shifted emission lines from quasars in the Sloan Digital Sky Survey, which would be a signature of an ejected accretion disk, placed upper limits of incidence of recoiling black holes in quasars at 4% (0.35%) for kicks greater that $500 \text{ km/s}$ ($1000 \text{ km/s}$) in the line-of-sight. Note that similar arguments also place doubt on a common explanation that X-shaped jets from radio-loud AGN are the result of spin re-alignment from a recent merger event. Kicks of smaller velocities that temporarily displace the black hole from the galactic center could also have interesting consequences, since this will transfer energy to stars in the nucleus, softening a steep density cusp. Also, modest recoil velocities will have a pronounced effect on the black hole population of globular clusters, effectively depleting the clusters of a large fraction of their black holes and leading to a “rogue” population of wandering black holes in the galactic halo.

### 3. Implication of waveform structure for detection efforts

The relative simplicity of the merger waveform, assuming the trend in new results continue and no complicated and lengthy structures in the merger phase occur generically, is on one hand a boon for gravitational wave astronomy, and on the other hand a curse. On the positive side is that the simple transition from inspiral to ringdown should allow the construction of high fidelity hybrid or fully analytic template banks, such as recently presented in. This will ensure, if the circumstances of the sources are consistent with the assumptions of the models, that the waves from the highest possible fraction of events passing earth during the operation of the various instruments will be detected. The downside is that, the less structure and shorter the waves, the more difficult it will be to discriminate between different events, and the less confidence with which one could claim observation of a particular event, statistics of source populations, etc. Indeed, in it was shown that high fitting factors can easily be achieved between a numerical source model and some member from a “wrong” template family. Note that this problem is only a significant issue for sources where the part of the waveform dominating the SNR comes from the last few cycles of inspiral, merger and ringdown, i.e. essentially the final burst (for LIGO this will be the merger of tens to hundreds of solar mass black holes, for LISA around $10^7 - 10^8 M_\odot$ supermassive black holes). When the inspiral portion of the event is visible to detectors it could be in band for hundreds to thousands of cycles, and in that case even small differences in the phase evolution relative to a given template could drastically affect the SNR. One way to deal with this problem for burst-like sources is to use a small, core template family to search for a given source, then have an expanded control group of template families with systematic deviations from the core family that will be used to place confidence levels and/or error-bars on any conclusions reached with the core family. For example, say one wants to test the hypothesis that all stellar mass mergers occur in environments where the orbits have circularized well before the time of merger. The core template family for this search would thus be a set of zero-eccentricity inspiral events, and the control group would be a set of similar inspirals with eccentricity. The point here is not to go into details of data analysis, rather it is to emphasize how important it will be to investigate non-standard, unexpected or unusual scenarios, not just for the hope of a serendipitous detection of a surprising event, but to strengthen the science that could be done with the usual and more mundane detections. This is perhaps most true for what will be the first triumph of detection of a binary black hole merger—confirmation of the existence.
of black holes. It is easy to lapse into a mental image of a black hole as this dark, concrete object with a surface, rather than the boundary demarking a region where the geometric structure of spacetime is undergoing gravitational collapse—an intrinsically dynamical regime where space and time itself are funneled to a singular state outside of the grasp of contemporary physical theories. To claim that such a remarkable scenario exists in the universe demands strong evidence, and part of this evidence would be being able to quantify exactly how distinctive the signatures of mergers of binary black holes within Einstein’s theory of gravity are. For example, could compact boson stars [321, 322], “gravestar” [323], or other exotic object mergers produce inspiral signals that would be detected using a black hole inspiral template family? How different would a metric theory of gravity have to be to produce observable differences from Einstein’s theory (yet be consistent in the weak field)? Could merger events contain the signatures of certain extra-dimension scenarios? The list of such questions is endless, though a reasonable subset will need to be addressed if only to bolster our confidence about what possible future detections could tell us about general relativity and how accurately it describes spacetime.

B. The black hole scattering problem

There is no-known natural mechanism in the universe that can accelerate black holes to ultra-relativistic velocities, and hence the black hole scattering problem is largely a thought experiment that can probe a very interesting regime of Einstein’s theory. However, over the past few years several ideas have emerged suggesting that an understanding of this problem could have relevance to high-energy particle physics experiments—this will briefly be discussed in the following section. As with rest-mass dominated collisions until recently, full solutions to the metrics describing ultra-relativistic collisions have eluded analytic attempts to obtain them. Most of what is known about this regime comes from studies of the collision of infinite-boosted Schwarzschild black holes, each described by the Aichelberg-Sexl metric [63]. In an impact with zero or small impact parameter, an apparent horizon has been found at the moment of impact [84, 85, 86, 87, 88, 89, 90, 91, 92, 93]; this is not a trivial statement, as the Aichelberg-Sexl metric does not contain an event horizon, and in this extreme case one might imagine that a naked singularity would form, in particular by drawing parallels to collisions of infinite plane gravitational waves [33]. For zero impact parameter, perturbative studies [84, 85, 87, 88] have given a description of the gravitational waves released in the process. Now, as is turning out with merger simulations in the rest-mass dominated regime, it may be that full (presumably numerical) solutions of the scattering problem will only add some details to our understanding of the process, though of course this will not be known until the solutions are discovered. Furthermore, as described in Sec. II A 2, the threshold of immediate merger could have very interesting behavior associated with it, where essentially all of the energy in the spacetime is converted to gravitational waves. Note that in the infinite boost limit the threshold of immediate merger also corresponds to a threshold of black hole formation. If, as conjectured by Choptuik [28], the threshold of black hole formation has a universal solution, then at critical impact parameter the structure of spacetime should be described by the Abrahams-Evans axisymmetric critical collapse vacuum solution [324].

The black hole scattering problem will be difficult to simulate numerically. First of all, it is unclear whether generalized harmonic or BSSN evolutions could be used without modification in this regime. Attempted evolution of boosted exact Schwarzschild solutions with Lorentz \(\gamma\)-factors above around 1.7 with the generalized harmonic code used in [228] suggested that at the very least new gauge conditions will be needed for long-term stable evolution. The BSSN code in [256] has been used to evolve boosted Bowen-York black holes with somewhat larger \(\gamma\) factors, though such initial data contains gravitational waves, and apparently a considerable amount with larger boosts. A further issue with evolving highly boosted black holes is the tremendous computational resources that would be required. Consider a single blackhole with total energy \(E\), which will be the characteristic length scale of the geometrically non-trivial portion of the spacetime. The black hole would have a rest mass of \(E/\gamma\), which would be the smallest length scale in the direction transverse to the boost direction. Length contraction in the direction of the boost will compress the horizon by an addition factor of \(\gamma\) in that direction. Furthermore, using a “naive” boost of the Schwarzschild solution, certain components of the metric get scaled by factors of \(\gamma^2\). Thus, in all, to obtain a numerical solution with a grid-based method will require a mesh spacing a factor of \(\gamma^4\) smaller than a similar rest-mass dominated problem and obtain similar accuracy. To be well within the kinetic energy dominated regime would require \(\gamma \approx 10 - 100\), implying around \(10^4 - 10^8\) times the computational resources. Of course, this is rather simplistic accounting, and certainly with some ingenuity a couple of orders of magnitude could be shaved off the estimated cost.
C. High energy particle experiments and black hole collisions

Recently proposed extra-dimension scenarios\(^{326, 327}\), offer the intriguing possibility that the Planck scale could be within reach of energies attainable by the Large Hadron Collider (LHC)\(^{328, 329, 330, 331, 332, 333}\). This implies that the LHC may be able to probe the quantum gravity regime, and that black holes could be produced in substantial quantities by the particle collisions. Similarly, cosmic ray collisions with the earth would produce black holes\(^{334}\), and this may be detected with current or near-future cosmic ray experiments\(^{335}\). In the collision of two particles with super-Planck kinetic energies, gravity dominates the interaction, and thus to a good approximation the collision can be modeled as the ultra-relativistic collision of two black holes\(^{17}\). Another intriguing application of ultra-relativistic black hole collisions is in 5 dimensional AdS spacetime, and how that might relate to the collision of gold ions at the Relativistic Heavy Ion Collider (RHIC). At RHIC, gold ions are accelerated to Lorentz gamma factors of around 100 before colliding. It is believed that during the collision a quark-gluon plasma (QGP) is formed. The present data supports this idea\(^{336, 337, 338, 339}\), though there are some puzzles, in particular why the QGP is strongly interacting, behaving almost like an ideal fluid (the energies of RHIC collisions are in the regime where the asymptotic freedom of QCD should be manifest, implying one should have a weakly interacting QGP). One suggested method for deriving properties of the QGP is via the AdS/CFT correspondence of string theory\(^{340, 341, 342}\). Specifically, the supposition is that \(N = 4\) super-Yang-Mills (SYM) theory at strong coupling, though different in many respects from QCD, can describe some of the features of a “real-world” QGP, and that a practical way of calculating the relevant SYM state is using the AdS/CFT map applied to the corresponding process in 5D AdS spacetime. It has been suggested that the AdS equivalent process is the collision of two black holes\(^{343}\), and in\(^{344}\) the quasinormal ringing of a perturbed 5D AdS black hole, which represents the final stage of a black hole collision and may describe aspects of thermalization and collective flow of the QGP, was used to provide in-the-ballpark estimates of the thermalization time and elliptic flow coefficients of an anisotropic heavy ion collision.

Thus there is considerable motivation to study black hole collisions in higher dimensional spacetimes, and in spacetimes with different asymptotic structures in regimes where the asymptotics are expected to affect the physics of the collision (in particular for collisions in AdS the black holes need to be “large” in terms of the length scale imparted by the cosmological constant). Full five (and higher) dimensional numerical simulations of collisions, in particular ultra-relativistic ones, will require computers several generations more powerful than current ones. However, if the analogy between geodesic behavior and the full problem at the threshold of immediate merger described in Sec.\(^{IV\,D}\) holds, then for an application to the LHC all that would be needed is a head-on collision simulation, which could be reduced to an 2D simulation\(^{18}\). Furthermore, if the threshold geodesic behavior of Myers-Perry solutions\(^{345}\) are an adequate description of the analogous problem in higher dimensions, it turns out that to a good approximation for dimensions greater than 5 the energy emitted in an ultra-relativistic collision will be given by\(^{346}\)

\[
E(b) = E_0 (\Theta(b) - \Theta(b - b^*)) ,
\]

where \(\Theta(b)\) is the unit step function, \(b\) the impact parameter, \(b^*\) the critical impact parameter for the geodesic (which is close to the Schwarzschild radius of the equivalent black hole), and \(E_0\) the energy emitted during the head-on collision (estimates of which can be found in\(^{92}\), for example). In otherwords, when black holes form, as much energy is lost to gravitational waves as in the head-on collision case, regardless of the impact parameter. This “missing energy”, in addition to Hawking radiation emitted when the black holes evaporate, could be used to detect this hypothesized scenario at the LHC.

VI. CONCLUSION

The two body problem in general relativity is a fascinating, rich problem that is just beginning to be fully revealed by recent breakthroughs in numerical relativity. At the same time, a new generation of gravitational wave detectors promise to offer us a view of the universe via the gravitational wave spectrum for the first time.

\(^{17}\) Though without a full theory of the physical laws that would operate in this regime, such statements are a bit hand-waving.

\(^{18}\) Of course, the “catch-22” here is that without doing the full n-dimensional problem it will not be known whether the analogy holds.
Black hole mergers are a promising source for gravitational waves, and detecting them would provide direct evidence for these remarkable objects, while providing much information about their environments. Suggestions that there might be more than four spacetime dimensions offers the astonishing possibility that black holes could be produced by proton collisions at TeV energies, which will be reached by the Large Hadron Collider, planned to begin operation within a year. Given all this, it is difficult not to be excited about what might be learnt about the universe from the smallest to largest scales during the next decade, and that black hole collisions could have something important to say at both extremes.

Acknowledgments: I would like to thank Alessandra Buonanno, Matthew Choptuik, Gregory Cook, Charles Gammie, David Garfinkle, Steven Gubser, Carsten Gundlach, Luis Lehner, Jeremiah Ostriker, Don Page, David Spergel and Ulrich Sperhake for many stimulating conversations related to some of the discussion presented here.
T. C. Quinn and R. M. 93, “An axiomatic approach to electromagnetic and gravitational radiation reaction of particles in curved spacetime,” Phys. Rev. D 56, 3381 (1997)

Y. Mino, M. Sasaki and T. Tanaka, “Gravitational radiation reaction to a particle motion,” Phys. Rev. D 55, 3457 (1997)

S. A. Hughes, “Computing radiation from Kerr black holes: Generalization of the Sasaki-Nakamura equation,” Phys. Rev. D 62, 044029 (2000) [Erratum-ibid. D 67, 089902 (2003)]

K. Glampedakis, S. A. Hughes and D. Kennefick, “Approximating the inspiral of test bodies into Kerr black holes,” Phys. Rev. D 66, 064005 (2002) [arXiv:gr-qc/0205033].

K. Glampedakis and S. Babak, “Mapping spacetimes with LISA: Inspiral of a test-body in a ‘quasi-Kerr’ field,” Class. Quant. Grav. 23, 4167 (2006)

J.R. Gair, D.J. Kennefick and S.L. Larson, “Semi-relativistic approximation to gravitational radiation from encounters with black holes”, Phys.Rev. D72, 084009 (2005), Erratum-ibid. D74, 109901 (2006)

F. D. Ryan, “Gravitational waves from the inspiral of a compact object into a massive, axisymmetric body with arbitrary multipole moments,” Phys. Rev. D 52, 5707 (1995).

S. W. Hawking, “Gravitational radiation from colliding black holes,” Phys. Rev. Lett. 26, 1344 (1971).

C.V. Vishveshwara, Nature 227, 936 (1970).

M. Davis, R. Ruffini, W. H. Press and R. H. Price, “Gravitational radiation from a particle falling radially into a schwarzschild black hole,” Phys. Rev. Lett. 27, 1466 (1971).

W. Press, Astrophys J. Letters 170, L105 (1971).

S. Chandrasekhar and S. Detweiler, Proc. R. Soc. Lond. A 344, (1975) 441.

E. W. Leaver, Phys. Rev. D 34, 384 (1986).

K. D. Kokkotas and B. G. Schmidt, “Quasi-normal modes of stars and black holes,” Living Rel. 2, 2 (1999)

E. Echeverria, Phys. Rev. D 40, 3194 (1997).

E. Berti, V. Cardoso and C. M. Will, “On gravitational-wave spectroscopy of massive black holes with the space interferometer LISA,” Phys. Rev. D 73, 064030 (2006)

R. H. Price and J. Pullin, Phys. Rev. Lett. 72, 3297 (1994) [arXiv:gr-qc/9402039].

R. H. Price, “Nonspherical perturbations of relativistic gravitational collapse. 1. Scalar and gravitational perturbations,” Phys. Rev. D 5, 2419 (1972).

R. Penrose, unpublished (1974)

P. D. D’Eath, “High Speed Black Hole Encounters And Gravitational Radiation,” Phys. Rev. D 18, 990 (1978).

P. D. D’Eath and P. N. Payne, “Gravitational Radiation In High Speed Black Hole Collisions. 1. Perturbation Treatment Of The Axisymmetric Speed Of Light Collision,” Phys. Rev. D 46, 658 (1992).

P. D. D’Eath and P. N. Payne, “Gravitational Radiation In High Speed Black Hole Collisions. 2. Reduction To Two Independent Variables And Calculation Of The Second Order News Function,” Phys. Rev. D 46, 675 (1992).

P. D. D’Eath and P. N. Payne, “Gravitational radiation in high speed black hole collisions. 3. Results and conclusions,” Phys. Rev. D 46, 694 (1992).

D. M. Eardley and S. B. Giddings, “Classical black hole production in high-energy collisions,” Phys. Rev. D 66, 044011 (2002) [arXiv:gr-qc/0203049].

H. Yoshino and Y. Nambu, “Black hole formation in the grazing collision of high-energy particles”, Phys.Rev. D67, 024009 (2003)

E. Berti, M. Cavaglia and L. Gualtieri, “ Gravitational energy loss in high-energy particle collisions: Ultrarelativistic plunge into a multidimensional black hole”. Phys. Rev. D69, 124011 (2004).

H. Yoshino and S. Rychkov, “Improved analysis of black hole formation in high-energy particle collisions”, Phys.Rev. D71, 104028 (2005)

V. Cardoso, E. Berti and M. Cavaglia, “What we (don’t) know about black hole formation in high-energy collisions”, Class.Quant.Grav. 22, L61-L84 (2005)

M.W. Choptuik, private communication

F. Pretorius, “Simulation of binary black hole spacetimes with a harmonic evolution scheme”, Class. Quant. Grav. 23, S529 (2006)

F. Pretorius and D. Khurana, “Black hole mergers and unstable circular orbits,” Class. Quant. Grav. 24, S83 (2007)

F. Pretorius, “Evolution of Binary Black Hole Spacetimes”, Phys. Rev. Lett. 95, 121101 (2005)

M. Campanelli, C.O. Lousto, P. Marronetti and Y. Zlochower, “Accurate Evolutions of Orbiting Black-Hole Binaries Without Excision”, Phys. Rev. Lett. 96, 111101 (2006)

J. G. Baker, J. Centrella, D. Choi, M. Koppitz and J. van Meter, “Gravitational Wave Extraction from an Inspiraling Configuration of Merging Black Holes”, Phys. Rev. Lett. 96, 111102 (2006)

O. A. Reula, “Hyperbolic Methods For Einstein’s Equations,” Living Rev. Rel. 1, 3 (1998).

L. Lehner, “Numerical relativity: A review,” Class. Quant. Grav. 18, R25 (2001)

C. Bona, T. Ledvinka, C. Palenzuela and M. Zacek, “General-covariant evolution formalism for Numerical Relativity,” Phys. Rev. D 67, 104005 (2003) [arXiv:gr-qc/0302083].

C. Bona, T. Ledvinka, C. Palenzuela and M. Zacek, “General-covariant constraint-free evolution system for Nu-
merical Relativity," arXiv:gr-qc/0209082.
[104] C. Gundlach, J. M. Martin-Garcia, G. Calabrese and I. Hinder, “Constraint damping in the Z4 formulation and harmonic gauge,” Class. Quant. Grav. 22, 3767 (2005).
[105] S.G.Hahn and R.W. Lindquist, Ann. Phys. 29, 304 (1964).
[106] C.W. Misner, “Wormhole Initial Conditions”, Phys. Rev. 118, 1110 (1960) Ann. Phys. 29, 304 (1964)
[107] L. Smarr, “The Structure of General Relativity with a Numerical Illustration: The Collision of Two Black Holes”, Univ. of Texas at Austin Ph.D. Thesis (1975)
[108] K.R. Eppley, Princeton Ph.D. Thesis (1977)
[109] L. Smarr, in Sources of Gravitational Radiation, ed. L. Smarr, Seattle, Cambridge University Press (1979)
[110] P. Anninos, D. Hobill, E. Seidel, L. Smarr and W. Suen, “Collision of Two Black Holes”, Phys. Rev. Lett. 71, 2851 (1993)
[111] P. Anninos, K. Camarda, J. Masso, E. Seidel, W. M. Suen, M. Tobias and J. Towns, “3-D numerical relativity at NCSA,” arXiv:gr-qc/9412059.
[112] P. Anninos, K. Camarda, J. Masso, E. Seidel, W. M. Suen and J. Towns, “Three-dimensional numerical relativity: The Evolution of black holes,” Phys. Rev. D 52, 2059 (1995)
[113] J.W. York, Jr., in Sources of Gravitational Radiation, ed. L. Smarr, Seattle, Cambridge University Press (1979)
[114] J. Winicour, “Characteristic Evolution and Matching,” arXiv:gr-qc/0102085.
[115] R. Gomez et al., “Stable characteristic evolution of generic 3-dimensional single-black-hole spacetimes,” Phys. Rev. Lett. 80, 3915 (1998)
[116] N. T. Bishop, R. Gomez, L. Lehner and J. Winicour, “Cauchy-characteristic extraction in numerical relativity,” arXiv:gr-qc/9705033.
[117] R. Gomez, R. L. Marsa and J. Winicour, “Black hole excision with matching,” Phys. Rev. D 56, 6310 (1997) [arXiv:gr-qc/9708002].
[118] W. Barreto, A. Da Silva, R. Gomez, L. Lehner, L. Rosales and J. Winicour, “The 3-dimensional Einstein-Klein-Gordon system in characteristic numerical relativity,” Phys. Rev. D 71, 064028 (2005)
[119] T. Piran, “Numerical Codes for Cylindrical Relativistic Systems”, J. Comp. Phys. 35, 254 (1980)
[120] B. Gustafsson, H. Kreiss and J. Oliger, Time Dependent Problems and Difference Methods, New York, John-Wiley & Sons, Inc. 1995
[121] C. Bona and J. Masso, “Hyperbolic evolution system for numerical relativity”, Phys. Rev. Lett. 68, 1097 (1992)
[122] S. Frittelli and O. A. Reula, Commun. Math. Phys. 166 221 (1994).
[123] C. Bona, J. Masso, E. Seidel and J. Stela, “A New formalism for numerical relativity,” Phys. Rev. Lett. 75, 600 (1995)
[124] A. Abrahams, A. Anderson, Y. Choquet-Bruhat and J. W. Jr., York, “Einstein And Yang-Mills Theories In Hyperbolic Form Without Gauge Fixing,” Phys. Rev. Lett. 75, 3377 (1995)
[125] H. Friedrich, “Hyperbolic Reductions For Einstein’s Equations,” Class. Quant. Grav. 13, 1451 (1996).
[126] S. Frittelli and O. A. Reula, “First-order symmetric-hyperbolic Einstein equations with arbitrary fixed gauge,” Phys. Rev. Lett. 76, 4667 (1996)
[127] A. Abrahams, A. Anderson, Y. Choquet-Bruhat and J. W. Jr., York, “Geometrical hyperbolic systems for general relativity and gauge theories,” Class. Quant. Grav. 14, A9 (1997)
[128] Y. Choquet-Bruhat and J. W. Jr., York, “Well Poss Reduced Systems for the Einstein Equations,” arXiv:gr-qc/9606001.
[129] C. Bona, J. Masso, E. Seidel and J. Stela, “First order hyperbolic formalism for numerical relativity,” Phys. Rev. D 56, 3405 (1997)
[130] M. S. Iriondo, E. O. Leguizamon and O. A. Reula, “Einstein’s equations in Ashtekar’s variables constitute a symmetric hyperbolic system,” Phys. Rev. Lett. 79, 4732 (1997)
[131] G. Yoneda and H. a. Shinkai, “Symmetric hyperbolic system in the Ashtekar formulation,” Phys. Rev. Lett. 82, 263 (1999)
[132] A. Anderson and J. W. Jr., York, “Fixing Einstein’s equations,” Phys. Rev. Lett. 82, 4384 (1999)
[133] G. Yoneda and H. a. Shinkai, “Constructing hyperbolic systems in the Ashtekar formulation of general relativity,” Int. J. Mod. Phys. D 9, 13 (2000)
[134] A. Arbona, C. Bona, J. Masso and J. Stela, “Robust evolution system for Numerical Relativity,” Phys. Rev. D 60, 104014 (1999)
[135] M. Alcubierre, B. Brugmann, M. A. Miller and W. M. Suen, “A conformal hyperbolic formulation of the Einstein equations,” Phys. Rev. D 60, 064017 (1999)
[136] S. Frittelli and O. A. Reula, “Well-posed forms of the 3+1 conformally-decomposed Einstein equations,” J. Math. Phys. 40, 5143 (1999)
[137] L. E. Kidder, M. A. Scheel and S. A. Teukolsky, “Extending the lifetime of 3D black hole computations with a new hyperbolic system of evolution equations,” Phys. Rev. D 64, 064017 (2001)
[138] H. a. Shinkai and G. Yoneda, “Adjusted ADM systems and their expected stability properties,” Class. Quant. Grav. 19, 1027 (2002)
[139] Y. Choquet-Bruhat, “Non strict and strict hyperbolic systems for the Einstein equations,” arXiv:gr-qc/0111017.
[140] H. Friedrich, “Conformal Einstein evolution,” Lect. Notes Phys. 604, 1 (2002)
Y. Choquet-Bruhat and T. Ruggeri, “Hyperbolicity Of The 3 + 1 System Of Einstein Equations,” Commun. Math. Phys. 89, 269 (1983).

O. Brodbeck, S. Fröllner, P. Hubner and O. A. Reula, “Einstein’s equations with asymptotically stable constraint propagation,” J. Math. Phys. 40, 909 (1999).

O. Sarbach and M. Tiglio, “Exploiting gauge and constraint freedom in hyperbolic formulations of Einstein’s equations,” Phys. Rev. D 66, 064023 (2002).

T. W. Baumgarte and S.L. Shapiro, “Numerical integration of Einstein’s field equations”, Phys. Rev. D59, 024007 (1999).

A. P. Gentle, N. D. George, A. Kheyfets and W. A. Miller, “The constraints as evolution equations for numerical relativity,” Class. Quant. Grav. 21, 83 (2004).

S. Bonazzola, E. Gourgoulhon, P. Grandclément and J. Novak, “A constrained scheme for Einstein equations based on Dirac gauge and spherical coordinates”, Phys.Rev. D70, 104007 (2004).

L. Andersson and V. Moncrief, “Elliptic-hyperbolic systems and the Einstein equations,” Annales Henri Poincare 4, 1 (2003).

M. Holst, L. Lindblom, R. Owen, H. P. Pfeiffer, M. A. Scheel and L. E. Kidder, “Optimal Constraint Projection for Hyperbolic Evolution Systems,” Phys. Rev. D 70, 084017 (2004) [arXiv:gr-qc/0407011].

C. R. Evans and J. F. Hawley, “Simulation of magnetohydrodynamic flows: A Constrained transport method,” Astrophys. J. 332, 659 (1988).

D. L. Meier, “Constrained Transport Algorithms for Numerical Relativity. I. Development of a Finite Difference Scheme,” Astrophys. J. 595, 980 (2003).

P.J.E. Peebles, Principles of Physical Cosmology, Princeton University Press (1993).

G. Calabrese, L. Lehner and M. Tiglio, “Constraint-preserving boundary conditions in numerical relativity,” Phys. Rev. D 65, 104031 (2002).

B. Szilagyi, B. G. Schmidt and J. Winicour, “Boundary conditions in linearized harmonic gravity,” Phys. Rev. D 65, 064015 (2002).

G. Calabrese, J. Pullin, O. Sarbach, M. Tiglio and O. Reula, “Well posed constraint-preserving boundary conditions for the linearized Einstein equations,” Commun. Math. Phys. 240, 377 (2003).

S. Frittelli and R. Gomez, “Boundary conditions for hyperbolic formulations of the Einstein equations,” Class. Quant. Grav. 20, 2379 (2003).

S. Frittelli and R. Gomez, “Einstein boundary conditions for the 3+1 Einstein equations,” Phys. Rev. D 68, 044014 (2003).

C. Gundlach and J. M. Martin-Garcia, “Symmetric hyperbolicity and consistent boundary conditions for second-order Einstein equations,” Phys. Rev. D 70, 044032 (2004).

C. Bona, T. Ledvinka, C. Palenzuela-Luque and M. Zacek, “Constraint-preserving boundary conditions in the Z4 Numerical Relativity formalism,” Class. Quant. Grav. 22, 2615 (2005).

O. Sarbach and M. Tiglio, “Boundary conditions for Einstein’s field equations: Analytical and numerical analysis,” J. Hyperbol. Diff. Equat. 2, 839 (2005).

L. E. Kidder, L. Lindblom, M. A. Scheel, L. T. Buchman and H. P. Pfeiffer, “Boundary conditions for the Einstein evolution system,” Phys. Rev. D 71, 064020 (2005).

D. N. Arnold and N. Trefethen, “Boundary conditions for the Einstein-Christoffel formulation of Einstein’s equations,” arXiv:gr-qc/0611010.

M. C. Babiuc, H. O. Kreiss and J. Winicour, “Constraint-preserving Sommerfeld conditions for the harmonic Einstein equations,” arXiv:gr-qc/0612051.

L. T. Buchman and O. C. A. Sarbach, “Towards absorbing outer boundaries in General Relativity,” Class. Quant. Grav. 23, 6709 (2006) [arXiv:gr-qc/0608051].

O. Rinne, “Stable radiation-controlling boundary conditions for the generalized harmonic Einstein equations,” Class. Quant. Grav. 23, 6275 (2006) [arXiv:gr-qc/0606053].

H. O. Kreiss, O. Reula, O. Sarbach and J. Winicour, “The Einstein equations, boundaries and integration by parts,” arXiv:0707.4188 [gr-qc].

L. T. Buchman and O. C. A. Sarbach, “Improved outer boundary conditions for Einstein’s field equations,” Class. Quant. Grav. 24, S307 (2007) [arXiv:gr-qc/0703129].

M. C. Babiuc, H. O. Kreiss and J. Winicour, “Constraint-preserving Sommerfeld conditions for the harmonic Einstein equations,” Phys. Rev. D 75, 044002 (2007).

O. Rinne, L. Lindblom and M. A. Scheel, “Testing outer boundary treatments for the Einstein equations,” arXiv:0704.0782 [gr-qc].

H. Kreiss and J. Oliger, “Methods for the Approximate Solution of Time Dependent Problems”, Global Atmospheric Research Programme, Publications Series No. 10. (1973)

C. W. Misner, “Over the Rainbow: Numerical Relativity beyond Scri+,” arXiv:gr-qc/0512167.

J. R. van Meter, D. R. Fiske and C. W. Misner, “Excising das All: Evolving Maxwell waves beyond Scri,” Phys. Rev. D 74, 064003 (2006).

C. W. Misner, “Hyperboloidal slices and artificial cosmology for numerical relativity,” arXiv:gr-qc/0409073.

G. Calabrese, C. Gundlach and D. Hilditch, “Asymptotically null slices in numerical relativity: Mathematical
analysis and spherical wave equation tests,” Class. Quant. Grav. 23, 4829 (2006)
[207] R.M. Wald, General Relativity, The University of Chicago Press (1984).
[208] S. Brandt and B. Brugmann, “Black hole punctures as initial data for general relativity,” Phys. Rev. Lett. 78, 3606 (1997)
[209] M. Hannam, S. Husa, D. Pollney, B. Brugmann and N. O’Murchadha, “Geometry and Regularity of Moving Punctures,” arXiv:gr-qc/0606099.
[210] M. Hannam, S. Husa, B. Brugmann, J. A. Gonzalez, U. Sperhake and N. O. Murchadha, “Where do moving punctures go?,” arXiv:gr-qc/0612097.
[211] J. Thornburg, P. Diener, D. Pollney, L. Rezzolla, E. Schnetter, E. Seidel and R. Takahashi, “Are moving punctures equivalent to moving black holes?,” arXiv:gr-qc/0701038.
[212] T. W. Baumgarte and S. G. Naculich, “Analytical Representation of a Black Hole Puncture Solution,” Phys. Rev. D 75, 067502 (2007)
[213] J. D. Brown, “Puncture Evolution of Schwarzschild Black Holes,” arXiv:0705.1359 [gr-qc].
[214] G. B. Cook, “Initial Data for Numerical Relativity,” Living Rev. Rel. 3, 5 (2000)
[215] H. P. Pfeiffer, “The initial value problem in numerical relativity,” arXiv:gr-qc/0412002.
[216] E. Gourgoulhon, “Construction of initial data for 3+1 numerical relativity,” arXiv:0704.0149 [gr-qc].
[217] S. Nissank, “Post-Newtonian freely specifiable initial data for binary black holes in numerical relativity,” Phys. Rev. D 73, 124002 (2006) [arXiv:gr-qc/0509128].
[218] N. Yunes, W. Tichy, B. J. Owen and B. Bruegmann, “Binary black hole initial data from matched asymptotic expansions,” Phys. Rev. D 74, 104011 (2006)
[219] N. Yunes and W. Tichy, “Improved initial data for black hole binaries by asymptotic matching of post-Newtonian and perturbed black hole solutions,” Phys. Rev. D 74, 064013 (2006) [arXiv:gr-qc/0601046].
[220] B. J. Kelly, W. Tichy, M. Campanelli and B. F. Whiting, “Black hole puncture initial data with realistic gravitational wave content,” arXiv:0704.0628 [gr-qc].
[221] T. DeDonder, La Gravifique Einsteinienne, Guntthier-Villars, Paris (1921); T. DeDonder, The Mathematical Theory of Relativity. Massachusetts Institute of Technology, Cambridge, MA (1927).
[222] V. Fock, The Theory of Space Time and Gravitation, Pergamon Press, New York (1959)
[223] Y. Four’es-Bruhat, “Theor’eme dexistence pour certains syst’emes dequations aux derivees partielles non lineaires”, Acta Math., 88 141 (1952); Y. Bruhat, “The Cauchy problem” in Gravitation: An Introduction to Current Research, ed. L. Witten, New York, Wiley (1962)
[224] A. E. Fischer and J. E. Marsden, “The Einstein evolution equations as a first-order quasi-linear symmetric hyperbolic system,” Commun. Math. Phys., 28 1 (1972).
[225] J. Renn and T. Sauer, “Heuristics and mathematical representation in Einsteins search for a gravitational field equation”, in History of General Relativity, volume 7 of Einstein Studies, eds. H. Goenner, J. Renn, and J. Ritter, Birkhauser, Cambridge, MA (1998).
[226] D. Garfinkle, “Harmonic coordinate method for simulating generic singularities”, Phys.Rev. D65, 044029 (2002)
[227] B. Szilagyi and J. Winicour, “Well-Posed Initial-Boundary Evolution in General Relativity”, Phys.Rev. D68, 041501 (2003)
[228] F. Pretorius, “Numerical Relativity Using a Generalized Harmonic Decomposition”, Class. Quant. Grav. 22 425, (2005)
[229] L. Lindblom, M. A. Scheel, L. E. Kidder, R. Owen and O. Rinne, “A New Generalized Harmonic Evolution System”, Class.Quant.Grav. 23, S447 (2006)
[230] M. C. Babic, B. Szilagyi and J. Winicour, “Harmonic Initial-Boundary Evolution in General Relativity,” Phys.Rev. D73, 064017 (2006)
[231] B. Szilagyi, D. Pollney, L. Rezzolla, J. Thornburg and J. Winicour “An explicit harmonic code for black-hole evolution using excision”, gr-qc/0612150.
[232] C. Bona and J. Masso, “Harmonic Synchronizations of Space-time,” Phys. Rev. D 38, 2419 (1988).
[233] C. Bona, J. Masso and J. Stela, “Numerical black holes: A Moving grid approach,” Phys. Rev. D 51, 1639 (1995)
[234] M. Shibata and T. Nakamura, “Evolution of three-dimensional gravitational waves: Harmonic slicing case”, Phys. Rev. D52, 5428 (1995).
[235] M. Alcubierre, “The appearance of coordinate shocks in hyperbolic formalisms of General Relativity,” Phys. Rev. D 55, 5981 (1997)
[236] M. Alcubierre and J. Masso, “Pathologies of hyperbolic gauges in general relativity and other field theories,” Phys. Rev. D 57, 4511 (1998)
[237] S. D. Hern, “Coordinate Singularities in Harmonically-sliced Cosmologies,” Phys. Rev. D 62, 044003 (2000) [arXiv:gr-qc/0001070].
[238] M. Alcubierre, “Are gauge shocks really shocks?,” Class. Quant. Grav. 22, 4071 (2005)
[239] D. Garfinkle, C. Gundlach and D. Hilditch, “Comments on Bona-Masso type slicing conditions in long-term black hole evolutions,” arXiv:0707.0726 [gr-qc].
[240] H. Friedrich, “On the Hyperbolicity of Einsteins’s and Other Gauge Field Equations”, Commun. Math. Phys 100, 525 (1985)
[241] R. Owen, “Constraint Damping in First-Order Evolution Systems for Numerical Relativity,” arXiv:gr-qc/0703145.
P. Laguna and D. Shoemaker, “Numerical stability of a new conformal-traceless 3+1 formulation of the Einstein equation,” Class. Quant. Grav. 19, 3679 (2002)

M. Alcubierre, B. Brugmann, P. Diener, M. Koppitz, D. Pollney, E. Seidel and R. Takahashi, “Gauge conditions for long-term numerical black hole evolutions without excision,” Phys. Rev. D 67, 084023 (2003)

U. Sperhake, “Binary black-hole evolutions of excision and puncture data,” arXiv:gr-qc/0606079.

L. E. Kidder, M. A. Scheel, S. A. Teukolsky, E. D. Carlson and G. B. Cook, “Black hole evolution by spectral methods,” Phys. Rev. D 62, 084032 (2000)

M. Boyle, L. Lindblom, H. Pfeiffer, M. Scheel and L. E. Kidder, “Testing the Accuracy and Stability of Spectral Methods in Numerical Relativity,” Phys. Rev. D 75, 024006 (2007)

P. Grandclement and J. Novak, “Spectral Methods for Numerical Relativity,” arXiv:0706.2286 [gr-qc].

Y. Zlochower, J. G. Baker, M. Campanelli and C. O. Lousto, “Accurate black hole evolutions by fourth-order numerical relativity,” Phys. Rev. D 72, 024021 (2005)

http://www.cactuscode.org

E. Schnetter, S. H. Hawley and I. Hawke, “Evolutions in 3D numerical relativity using fixed mesh refinement,” Class. Quant. Grav. 21, 1465 (2004)

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

PAMR (Parallel Adaptive Mesh Refinement) and AMRD (Adaptive Mesh Refinement Driver) libraries (http://laplace.physics.ubc.ca/Group/Software.html)

D. Neilsen, E. W. Hirschmann, M. Anderson and S. L. Liebling, “Adaptive mesh refinement and relativistic MHD,” arXiv:gr-qc/0702035.

P. Marronetti, W. Tichy, B. Brugmann, J. Gonzalez, M. Hannam, S. Husa and U. Sperhake, “Binary black holes on a budget: Simulations using workstations,” Class. Quant. Grav. 24, S43 (2007)

D. Brown, O. Sarbach, E. Schnetter, M. Tiglio, P. Diener, I. Hawke and D. Pollney, “Excision without excision: the relativistic turducken,” arXiv:0707.3101 [gr-qc].

B. Imbiriba et al., “Evolving a puncture black hole with fixed mesh refinement,” Phys. Rev. D 70, 124025 (2004) [arXiv:gr-qc/0403048].

S. Husa, J. A. Gonzalez, M. Hannam, B. Brugmann and U. Sperhake, “Reducing phase error in long numerical black hole evolutions with sixth order finite differencing,” arXiv:0706.0740 [gr-qc].

M. Campanelli, C. O. Lousto and Y. Zlochower, “The last orbit of binary black holes,” Phys. Rev. D 73, 061501 (2006) [arXiv:gr-qc/0601091].

J. G. Baker, M. Campanelli, F. Pretorius and Y. Zlochower, “Comparisons of binary black hole merger waveforms,” Class. Quant. Grav. 24, S25 (2007) [arXiv:gr-qc/0701016].

H. P. Pfeiffer, D. A. Brown, L. E. Kidder, L. Lindblom, G. Lovelace and M. A. Scheel, “Reducing orbital eccentricity in binary black hole simulations,” arXiv:gr-qc/0702106.

S. Husa, M. Hannam, J. A. Gonzalez, U. Sperhake and B. Brugmann, “Reducing eccentricity in black-hole binary evolutions with initial parameters from post-Newtonian inspiral,” arXiv:0706.0904 [gr-qc].

G. B. Cook and H. P. Pfeiffer, “Excision boundary conditions for black hole initial data,” Phys. Rev. D 70, 104016 (2004) [arXiv:gr-qc/0407078].

J. M. Bowen and J. W. York, “Time asymmetric initial data for black holes and black hole collisions,” Phys. Rev. D 21, 2047 (1980)

E. Berti, V. Cardoso, J. A. Gonzalez, U. Sperhake, M. Hannam, S. Husa and B. Brugmann, “Inspiral, merger and ringdown of unequal mass black hole binaries: A multipolar analysis,” arXiv:gr-qc/0703053.
