Fast and Accurate Prediction of Numerical Relativity Waveforms from Binary Black Hole Coalescences Using Surrogate Models

Jonathan Blackman,1 Scott E. Field,2 Chad R. Galley,1 Béla Szilágyi,1 Mark A. Scheel,1 Manuel Tiglio,3 and Daniel A. Hemberger1

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

(Received 2 March 2015; revised manuscript received 5 August 2015; published 18 September 2015)

Simulating a binary black hole coalescence by solving Einstein’s equations is computationally expensive, requiring days to months of supercomputing time. Using reduced order modeling techniques, we construct an accurate surrogate model, which is evaluated in a millisecond to a second, for numerical relativity (NR) waveforms from nonspinning binary black hole coalescences with mass ratios in [1, 10] and durations corresponding to about 15 orbits before merger. We assess the model’s uncertainty and show that our modeling strategy predicts NR waveforms not used for the surrogate’s training with errors nearly as small as the numerical error of the NR code. Our model includes all spherical-harmonic $l=2$ $Y_{lm}$ waveform modes resolved by the NR code up to $\ell = 8$. We compare our surrogate model to effective one body waveforms from 50$M_\odot$ to 300$M_\odot$ for advanced LIGO detectors and find that the surrogate is always more faithful (by at least an order of magnitude in most cases).

DOI: 10.1103/PhysRevLett.115.121102 PACS numbers: 04.25.dg, 02.60.-x, 04.30.-w, 04.30.Db

Since the breakthroughs of 2005 [1–3], tremendous progress in numerical relativity (NR) has led to hundreds of simulations of binary black hole (BBH) coalescences [4–10]. This progress has been driven partly by the data analysis needs of advanced ground-based gravitational wave detectors like LIGO [11] and Virgo [12]. Recent upgrades to these detectors are expected to yield the first direct detections of gravitational waves (GWs) from compact binary coalescences [13].

Despite the remarkable progress of the NR community, a single high-quality simulation typically requires days to months of supercomputing time. This high computational cost makes it difficult to directly use NR waveforms for data analysis, except for injection studies [4,9], since detecting GWs and inferring their source parameters may require thousands to millions of accurate gravitational waveforms. Nevertheless, a first template bank for nonspinning waveforms using an EOB model [18],a s

In this Letter, we present an ab initio methodology based on surrogate [27,28] and reduced order modeling techniques [29–33] that is capable of accurately predicting the gravitational waveform outputs from NR without any phenomenological assumptions or approximations to general relativity. From a small set of specially selected nonspinning BBH simulations performed with the Spectral Einstein code (spec) [34–36], we build a surrogate model that can be used in place of performing spec simulations. The techniques are general, however, and directly apply to other NR codes or even analytical waveform models. The surrogate model constructed here generates nonspinning BBH waveforms with mass ratios $q \in [1, 10]$, contains 25–31 gravitational wave cycles before peak amplitude, and includes many spherical-harmonic modes (see Table II and its caption). These choices are made based on available NR waveforms and are not limitations of the method. Our surrogate model has errors close to the estimated numerical error of the input waveforms. An example comparing the surrogate output to a NR waveform can be seen in Fig. 1. This simulation took 9.3 days using 48 cores but only ~0.01 sec for the surrogate evaluation of the (2,2) mode.

Previous work [27,37] built surrogates for EOB waveforms; building and assessing surrogate models of NR waveforms have unique challenges associated with input waveforms that are expensive to compute. We summarize next the construction of our model, focusing on steps not addressed in [27] but that are required for NR surrogates.

Parametric sampling.—Typically, a surrogate model is trained on a dense set of waveforms known as the training set. In the case of NR, we cannot afford to generate a large number of waveforms. Instead, we generate a dense set of nonspinning waveforms using an EOB model [18], as
implemented in [38], which contains the \((l', m) = \{(2, 2), (2, 1), (3, 3), (4, 4), (5, 5)\}\) spin-weight \(-2\) spherical-harmonic modes and captures robust features of NR waveforms. The EOB training set waveforms are computed harmonic modes and captures robust features of NR \([10, 19]\) (see Table I), and the next 17 (ordered) mass ratio input from the EOB model. The resulting NR waveforms are more accurate. Bottom: Phase waveform (see below). Our full surrogate, trained on the entire data set, is more accurate. Our method for generating the NR waveforms. We seeded the greedy algorithm with five publicly available \(\text{spec}\) simulations of nonspinning BBH mergers \([10, 19]\) (see Table I), and the next 17 (ordered) mass ratio values are the algorithm’s output based on the EOB model. The final \(\sim 10\) mass ratios are included to improve the surrogate if necessary, since we can assess the surrogate model’s accuracy only after it is built. Our method for building surrogates is hierarchical \([27, 40]\); additional NR waveforms can be included to improve the model.

Generating the NR waveforms.—Table I summarizes the 22 \(\text{spec}\) simulations used in this Letter. See, e.g., Ref. \([35]\) for the numerical techniques used in \(\text{spec}\). The numerical resolution is denoted by “Levi,” where \(i\) is an integer that controls the local truncation error in the metric and its derivatives allowed by adaptive mesh refinement (AMR) in \(\text{spec}\); larger numbers correspond to smaller errors (the error threshold scales like \(e^{-i}\)) and more computationally expensive simulations. The scaling of global quantities (e.g., waveform errors) with \(i\) is difficult to estimate \(a\ priori\). Between two and five levels of resolution are simulated for each mass ratio. To achieve quasicircular orbits, initial data are subject to an iterative eccentricity reduction procedure resulting in eccentricities \(\lesssim 7 \times 10^{-4}\) \([41–43]\).

\(\text{spec}\) numerically solves an initial boundary value problem defined on a finite computational domain. To obtain waveforms at future null infinity \(\mathcal{I}^+\), we use the Cauchy characteristic extraction method \([44–48]\). Using the \(\text{pittNull}\) code \([44–46]\), we compute the Newman-Penrose scalar \(\Psi_4\) at \(\mathcal{I}^+\) and finally obtain the gravitational wave strain \(h\) through two temporal integrations. We minimize the low-frequency, noise-induced “drifts” \([47]\) by using frequency cutoffs. [We integrate \(\Psi_4\) twice in the frequency domain by dividing \(-\Psi_4(f)\) by \(2\pi \max(f, 2f_0/3)^2\), where \(f_0\) is the initial GW mode frequency].

Figure 2 shows the convergence typically observed in our simulations when using AMR. Because AMR makes independent decisions for different Levi, a particular subdomain may sometimes have the same number of grid points for two different values of Levi at a given time, and the subdomain boundaries do not necessarily coincide for different Levi. Thus, plots like Fig. 2 sometimes show anomalously small differences between particular pairs of numerical resolutions (for instance Lev2 vs Lev3 near \(t = -3500M\) in the top panel of Fig. 2). See Sec. IIIB of [35]. Nevertheless, the waveform differences generally decrease quickly with increasing resolution. Let

\[
\delta h_1^f(q) = \frac{\|h_1^f(q; g) - h_2^f(q; g)\|^2}{\sum_{f, m}\|h_2^m(q; g)\|^2}
\]

be the disagreement between two waveform modes, \(h_1^f\) and \(h_2^m\), where \(\|h(q; g)\|^2 = \int dt|h(q; g)|^2\). We
the (intended to be circular) NR initial data, and a multimode peak alignment scheme described by Eq. (2) for simulation 180 (dominate all other sources of error for the (2,2) mode, except for simulation 198 in Table I. Top: Waveform output as directly estimated numerical truncation error of each mode when \( h_l \) and \( h_2 \) are waveforms computed at the two highest resolutions. The full waveform error for a given mass ratio is \( \delta h(q) = \sum_{l,m} \delta h_{l,m}^c(q) \). [Throughout, we exclude \( m = 0 \) modes because (nonoscillatory) Christodoulou memory is not accumulated sufficiently in current NR simulations [49].] We report numerical truncation errors after an overall simulation-dependent time shift and rotation (which we shall refer to as surrogate alignment, described in the next section), which are physically unimportant coordinate changes. The resulting estimated numerical truncation errors of the dominant (2,2) modes, using our surrogate alignment scheme, are shown in Fig. 3 (black circles).

Additional error sources are nonzero eccentricity in the (intended to be circular) NR initial data, and an imperfect procedure for integrating \( \Psi_{l,m}^c \) to obtain \( h_{l,m}^c \equiv A_l^{c,m} \exp(-i\omega_{l,m}^c t) \). These both cause small oscillations in the waveform amplitudes \( A_l^{c,m}(t) \) and phases \( \omega_{l,m}^c(t) \) [47,50] that we model following [50]. We also compute the error in the strain integration scheme by comparing \( \Psi_{l,m}^c \) to two time derivatives of \( h_{l,m}^c \), as well as estimates for numerical errors in the Cauchy characteristic extraction method [48]. For the (2,2) mode, these additional errors are negligibly small compared to SPEC truncation errors (cf. Fig. 3).

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

\[
A(t; q)^2 \equiv \int_{s} d\Omega |h(t, \theta, \phi; q)|^2 = \sum_{\ell,m} |h_{\ell,m}^c(t; q)|^2, \tag{2}
\]

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

**Building the surrogate.**—Each \( m > 0 \) mode, \( h_{l,m}^c(t; q) \), is modeled separately while (due to reflection symmetry about the orbital plane) \( m < 0 \) modes are evaluated using \( h_{l,m}^c(t; q) = (-1)^l h_{l,m}^c(t; q) \). We model all \( m \neq 0 \) modes but keep only those yielding smaller surrogate errors \( \delta h_{l,m}^c \) compared to setting the mode to zero. Table II lists our modeled modes and their errors.

Our complete surrogate waveform model is defined by

\[
h_S(t, \theta, \phi; q) = \sum_{\ell,m} h_{l,m}^c(t; q) Y_{\ell,m}(\theta, \phi) \]

where

\[
h_{l,m}^c(t; q) = A_{l,m}^c(t; q) e^{-i\omega_{l,m}^c(q) t},
\]

\[
X_{\ell,m}^c(t; q) = \sum_{i=1}^{N_x} B_{\ell,m}^{\ell,i}(t) X_i^c(q), \quad X = \{A, \phi\}. \tag{3}
\]

Unlike Ref. [27], we construct a reduced basis representation for the waveform amplitudes and phases separately, instead of the waveforms themselves [37]. Here, the \( \{B_{\ell,m}^{\ell,i}\}_{i=1}^{N_x} \) are computed off-line from the spec waveforms [27]. At a set of \( N_x \) specially selected times \( \{T_{X,i}\}_{i=1}^{N_x} \), which are the empirical interpolant nodes [27,51], the functions \( X_i^c(q) \approx X_{\ell,m}^c(T_{X,i}^c; q) \) approximate the parameter variation of the amplitudes and phases (via fitting).
TABLE II. Relative mode errors, reported as $10^3 \times \| h^{\ell,m}_S(q) - h^{\ell,m}(q) \|^2/\| h^{\ell,m}(q) \|^2$, from the leave-one-out surrogates. Only those modes which contribute greater than 0.02% to the full waveform’s time-domain power are used in the computation of the max and mean, except for “All” which is just $\delta h$. Our surrogate also includes the (3,1), (4,2,3], (5,3,4,5)], (6,4,5,6)], (7,5,6,7)], and (8,7,8)] modes. Weaker modes typically have relative errors between 1% and 35%.

<table>
<thead>
<tr>
<th>$(\ell, m)$</th>
<th>Surrogate</th>
<th>NR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2,2)</td>
<td>0.36</td>
<td>0.36</td>
</tr>
<tr>
<td>(1,1)</td>
<td>29</td>
<td>4.1</td>
</tr>
<tr>
<td>(3,3)</td>
<td>53</td>
<td>11</td>
</tr>
<tr>
<td>(3,2)</td>
<td>100</td>
<td>1.7</td>
</tr>
<tr>
<td>(4,4)</td>
<td>7.4</td>
<td>20</td>
</tr>
<tr>
<td>All</td>
<td>0.42</td>
<td>0.40</td>
</tr>
</tbody>
</table>

A thorough discussion of surrogate model building steps is presented in [27]. When evaluating the surrogate at a particular mass ratio, the fits are evaluated first to determine the amplitudes and phases at their respective interpolating times $\{T_{\ell,m}^{\ell,m}(X,h)\}_{\ell=1}$. The remaining operations yield the surrogate model prediction, $h_S(t, \theta, \varphi; q)$.

To find each $X_{\ell,m}^\nu(q)$ we perform least-squares fits to the 22 data points, $\{X^{\ell,m}(T_{\ell,m}^{\ell,m}(X,h);q)\}_{\ell=1}$. All fits except odd $m$ mode amplitudes use fifth degree polynomials in the symmetric mass ratio, $\nu = q/(1 + q)^2$. For odd $m$ modes, the amplitude approaches 0 and its derivative with respect to $\nu$ diverges as $q \to 1$ (or $\nu \to 1/4$). Consequently, we use $X_{\ell,m}^\nu(\nu) = \sum_{n=1}^{3} a_{n}^m (1 - 4 \nu)^n$ to account for this behavior. The waveform phases of odd $m$ modes at $q = 1$, which are undefined, are excluded when fitting for each $q_i^{\ell,m}(q)$.

**Assessing surrogate errors.**—We next assess the surrogate’s predictive quality. To quantify the error in the surrogate model itself, as opposed to its usage in a data analysis study, we do not minimize the errors over relative time and phase shifts here.

A first test is a consistency check to reproduce the 22 input SpEC waveforms used to build the surrogate. These errors are shown in Fig. 4 (red squares) and are comparable to or smaller than the largest SpEC truncation errors (black circles).

A more stringent test is the leave-one-out cross-validation (LOOCV) study [52]. For each simulated mass ratio $q_i$, we build a temporary trial surrogate using the other 21 waveforms, evaluate the trial surrogate at $q_i$, and compare the prediction with the SpEC waveform for $q_i$. Hence, the trial surrogate’s error at $q_i$ should serve as an upper bound for the full surrogate trained on all 22 waveforms. Repeating this process for all possible 20 LOOCV tests results in Fig. 4 (blue triangles). (We omit the smallest and largest mass ratios here as the corresponding trial surrogates would extrapolate to their values.)

Despite the $i$th trial surrogate having no information about the waveform at $q_i$, the errors remain comparable to the largest SpEC truncation errors. The LOOCV errors are typically twice as large as the full surrogate ones, confirming the former as bounds for the latter. Relative errors for selected modes are shown in Table II. While weaker modes have larger relative errors, their power contribution is small enough that the error in the full surrogate waveform, $\delta h$, is nearly identical to the SpEC resolution error.

A third test is to compare the surrogate waveforms to those of a second surrogate, built from the second highest resolution SpEC waveforms. The resulting comparison is shown in Fig. 4 (cyan line). These errors are comparable to SpEC waveform truncation errors (black circles). We find that the surrogate building process is robust to resolution differences. Furthermore, the surrogate can be improved using NR waveforms of higher accuracy.

We perform a final test and construct surrogates using the first $N$ selected mass ratios (from Table I) as input waveforms, leaving $22 - N$ mass ratios with which to test. We find the total surrogate error decreases exponentially with $N$ and is comparable to the SpEC truncation error after using 15 waveforms. Some modes [e.g., (2,2)] are fully resolved after as few as seven waveforms.

**Comparison to EOB.**—For data analysis purposes, we compare our surrogate with EOBNRV2 [19] and seOBNRV2 [21] models (generated from a current implementation in LAL [38]). In Fig. 5, we show the unfaithfulness

$$1 - \max_{\delta \varphi, \delta t} \Re \int_{16 \text{ Hz}} {h_1(f; \varphi, \psi) h_2(f; \varphi, \psi + \delta \varphi) e^{2 \pi i f \delta t}} S_n(f)$$

of the surrogate and the two EOB models against the NR waveforms. Here, $\hat{h}$ is the normalized Fourier transform of $h$ (such that a waveform’s unfaithfulness with itself gives 0), and $S_n(f)$ the Advanced LIGO zero-detuned high
FIG. 5 (color online). Unfaithfulness, from Eq. (4), comparing spec with our surrogate, EOBNRv2, and SEOBNRv2 models using all available $m \neq 0$ modes. Dashed lines show the unfaithfulness for $(2,2)$ modes only. All waveforms are Planck tapered [53] for $t \in [-2750, -2500]M$ and $t \in [50, 90]M$. For the full multimodal waveforms, we maximize the unfaithfulness over $\theta$ and $\phi$ for the worst-case scenario. We use the “+” polarization, which is nonzero for all $(\theta, \phi)$. Left: The shaded regions contain all 22 mass ratios, while the dashed lines maximize over mass ratio. The vertical grey line is the minimum total mass ($\approx 115M_\odot$) ensuring all $(2,2)$ modes start with $\leq 15$ Hz at the end of the first tapering window. Right: Unfaithfulness for a $115M_\odot$ binary.

power sensitivity noise curve [54]. The surrogate is more faithful than both EOB models for all cases considered. Since SEOBNRv2 only provides $(2, \pm 2)$ modes, it performs worst for large total masses where additional modes become important. All models predict the $(2,2)$ mode with unfaithfulness $< 1\%$ for $q \in [1, 10]$ at $115M_\odot$; however, the EOB models are limited by the availability of subdominant modes.

Discussion.—We have built a surrogate model for NR nonspinning BBH merger waveforms with which to calibrate, refine, and make comparisons. Building NR surrogates of precessing BBH waveforms with which to calibrate, refine, and make comparisons. Building NR surrogates of precessing BBH merger waveforms, which may be modeled from the parameters specially selected in [55], offers a promising avenue for modeling the full seven dimensional BBH parameter space. The surrogate model described in this Letter is available for download at [56,57].

We thank Mike Boyle, Alessandra Buonanno, Collin Capano, Jan Hesthaven, Jason Kaye, Geoffrey Lovelace, Lee Lindblom, Tom Loredo, Christian Ott, Yi Pan, Harald Pfeiffer, Rory Smith, and Nicholas Taylor for many useful discussions throughout this project. This work was supported in part by NSF Grants No. CAREER PHY-0956189, No. PHY-1068881, No. PHY-1005655, No. PHY-1440083, No. PHY-1404569, and No. AST-1333520 to Caltech, NSF Grants No. PHY-1306125 and No. AST-1333129 to Cornell University, NSF Grant No. PHY-1500818 to the University of California at San Diego, NSF Grants No. PHY-1208861 and No. PHY-1316424 to the University of Maryland (UMD), NSERC of Canada, and the Sherman Fairchild Foundation. Computations were performed on the Zwicky cluster at Caltech, which is supported by the Sherman Fairchild Foundation and by NSF Grant No. PHY-0960291. Portions of this research were carried out at the Center for Scientific Computation and Mathematical Modeling cluster at UMD.
