A Surrogate model of gravitational waveforms from numerical relativity simulations of precessing binary black hole mergers

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

[1]*Theoretical Astrophysics 350-17, California Institute of Technology, Pasadena, California 91125, USA*
[2]*Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, New York 14853, USA*
[3]*Mathematics Department, University of Massachusetts Dartmouth,*
Dartmouth, Massachusetts 02747, USA
[4]*LIGO Laboratory, California Institute of Technology, MS 100-36, Pasadena, California 91125, USA*

(Received 4 January 2017; published 17 May 2017)

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

DOI: 10.1103/PhysRevD.95.104023

I. INTRODUCTION

With two confident detections of gravitational waves (GWs) from binary black hole (BBH) systems [1,2], an exciting era of gravitational wave astronomy has begun. Once a signal has been detected, the masses and spins of the black holes (BHs), and their uncertainties, can be determined by comparing the signal to waveforms predicted by general relativity (GR) [3]. Similarly, by comparing the signal to predictions, tests of GR can now be performed in the regime of strong-field dynamics with relativistic velocities [4].

Parameter estimation and tests of GR typically require the computation of predicted gravitational waveforms for a large set of different source parameters (e.g. black hole masses and spins). A typical Bayesian parameter estimation analysis, for example, evaluates millions of waveforms [5]. Therefore, in order to obtain reliable results on realistic time scales, the GW model must be fast to evaluate. Additionally, the waveform model must be accurate not only during the weak-field perturbative binary inspiral, but also in the strong-field, large-velocity regime. Otherwise the model may introduce biases in parameter estimation and inaccuracies in tests of GR. Waveform accuracy will become increasingly important in future GW measurements, because higher signal-to-noise-ratio detections are anticipated as detector technology improves.

Numerical relativity (NR) is now in a sufficiently mature state that there are a number of codes [6–12] capable of accurately simulating the late inspiral, merger and ringdown of a BBH system, and the resulting GWs, even for somewhat extreme spins [13,14] and high mass ratios [15,16]. While the resulting waveforms are quite accurate, the simulations can take weeks or months, thereby precluding them from being directly used in most data analysis studies. Therefore, data analysis studies currently use approximate NR-tuned waveform models that are fast to evaluate [17–24].

For the analysis of GW150914 [3,25], the first GW detection by Advanced LIGO [26], waveform models built within the effective-one-body (EOB) [22–24,27–29] and the phenomenological (Phenom) [19–21] frameworks were used [3,25]. All models necessarily introduce some systematic error, however small, which are often quantified either by comparing to NR simulations directly [20,30–33] or by performing parameter estimation with many different waveform models and monitoring the discrepancies. In the case of GW150914, the systematic error for the black hole masses was estimated to be smaller than the statistical uncertainty. However, estimating a model's systematic error in this way is complicated by the fact that the waveform models make similar simplifications. For example, the models ignore spin-weighted spherical-harmonic (SWSH) modes with $\ell > 2$, which may be significant since the signal's power is dominated by the late inspiral and merger. Recent studies continue to investigate this systematic parameter estimation bias through the use of newer waveform models including additional physics [25] and by comparing to NR waveforms [33].

In this paper, we use a surrogate model, which we call NRSur4d2s, to compute waveforms approaching the accuracy of NR simulations. A surrogate model [29,34–36] is a way to substantially accelerate the evaluation of a slower but accurate waveform model (in our case, NR), while largely retaining the accuracy of the original model. This is done by through an expensive *offline* stage, where we perform many accurate NR simulations for different input parameter values and subsequently build and validate the surrogate model on this set of simulations. The waveforms from these simulations are then "interpolated" in parameter space in an inexpensive *online* stage. The resulting model can be used in place of performing additional NR simulations. Surrogates can be used to accelerate other analytical models and have been used to successfully speed up nonspinning EOB models with multiple SWSH modes [35] and spin-aligned EOB models that include only the $\ell = 2$ modes [29,36]. Most recently, surrogates have been used to speed up nonspinning BBH waveforms from NR simulations including 40 SWSH modes [34].

The surrogate model we develop here is based on NR simulations using the Spectral Einstein Code (SpEC) [8,37–42]. It extends previous NR surrogate models [34] to include precessing binaries. The number of NR simulations required to build a surrogate model increases with parameter space size, and NR simulations become more expensive as the mass ratio and spin magnitudes grow. To reduce the computational cost, we restrict to a subspace of the full precessing parameter space. The initial spin direction of the smaller black hole is restricted to be parallel to the orbital angular momentum. We also restrict the mass ratio of the black holes to $1 \leq q \leq 2$ and the dimensionless spin magnitudes to be at most 0.8. The duration of each NRSur4d2s waveform is equal to that of the NR simulations, which begin $4500M$ before merger, corresponding to $\sim$30 gravitational wave cycles.

It has been shown that waveforms from precessing systems closely resemble waveforms from nonprecessing systems when viewed in a suitable noninertial, coprecessing frame [43,44]. We use this relationship to simplify the construction of the surrogate model by decomposing each precessing waveform into a simpler waveform measured in a coprecessing frame [45–47], plus a time-dependent rotation that characterizes the precession. Additional simplification is achieved by further decomposing each waveform into a set of functions that are slowly varying in parameter space and thus easier to model (cf. Fig. 6). The model is evaluated by "interpolating" these slowly varying functions to a desired point in parameter space, and then using the interpolated functions to reconstruct the waveform in the inertial source frame of the binary.

The NRSur4d2s surrogate model just described produces a waveform in the time domain and takes approximately 1 second to evaluate. While this is much faster than computing a waveform using NR, it is still too slow for many applications; furthermore many LIGO analyses are more easily performed in the frequency domain rather than the time domain. Therefore, we build a second surrogate model in the frequency domain, called NRSur4d2s_FDROM, using NRSur4d2s as input. NRSur4d2s_FDROM does not employ complicated decompositions of its input waveforms, so it requires significantly more waveforms to build (an offline cost), but because of its simplicity it is significantly faster and can be evaluated in about 50 ms.

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

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

II. SURROGATE MODELING METHODS

Compared to previous work [29,34–36,48–50], which focused on surrogates of analytical waveform models or on surrogates of simpler NR waveforms, surrogate models of precessing NR waveforms pose a number of new, unique challenges. First, the complicated waveform morphologies characteristic of precessing systems [51,52] suggest that a substantially larger training set may be necessary for these systems than for simpler cases considered previously. On the other hand, NR waveforms require the solution of computationally intensive time-dependent partial differential equations; current hardware and binary black hole evolution codes are capable of producing only roughly $\mathcal{O}(1,000)$ simulations in about a year.

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

$$h(t,\theta,\phi;\boldsymbol{\lambda}) = h_+(t,\theta,\phi;\boldsymbol{\lambda}) - ih_\times(t,\theta,\phi;\boldsymbol{\lambda}) \qquad (1)$$

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

When numerically generating a waveform by solving partial differential equations, one solves an initial-boundary value problem for a fixed $\boldsymbol{\lambda}$, thereby generating a waveform on a dense temporal grid. In this paper we seek to build an accurate and fast-to-evaluate surrogate gravitational-wave strain model $h_S(t,\theta,\phi;\boldsymbol{\lambda})$ by numerically solving the Einstein equations for judicious choices of $\boldsymbol{\lambda}$. Surrogate evaluations require only simple function evaluations, matrix-vector products and coordinate transformations. In Sec. VII we also build a frequency-domain surrogate model, using our time-domain surrogate model as input data, with the purpose of accelerating the evaluation of model waveforms. Evaluation of the frequency-domain model is about 20 times faster than the corresponding time-domain surrogate. Except for Sec. VII our discussion will focus exclusively on time-domain surrogates.

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

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

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

A. The basic surrogate modeling approach

1. Problem statement

Our surrogate modeling methods build on those outlined in [35], which we briefly describe here. Consider a physical system parametrized by $\boldsymbol{\lambda} \in \mathcal{T}$, where $\mathcal{T}$ is a compact region in the space of possible parameters. We seek quick-to-evaluate time-dependent functions $X(t;\boldsymbol{\lambda})$ that describe this system for times $t \in [t_{\min}, t_{\max}]$. In our case, $\boldsymbol{\lambda}$ will be the black hole masses and spins for a single BBH system, and $\mathcal{T}$ will extend to some maximum spin magnitude and maximum mass ratio for which we choose to compute NR waveforms. The functions $X(t;\boldsymbol{\lambda})$ will be obtained from decomposing $h^{\ell m}(t;\boldsymbol{\lambda})$ as described in Sec. IV, but here we discuss building a surrogate model for a single such function.

We already have a slow method of generating $X(t;\boldsymbol{\lambda})$, so we seek a faster *surrogate model*, denoted as $X_S(t;\boldsymbol{\lambda})$, which approximates $X(t;\boldsymbol{\lambda})$. The surrogate model $X_S(t;\boldsymbol{\lambda})$, whose construction is summarized in this section culminating in Eq. (13), is built to achieve small approximation errors $\|X(\cdot;\boldsymbol{\lambda}) - X_S(\cdot;\boldsymbol{\lambda})\|$. In our case, the slow method is performing a NR simulation, extracting $h^{\ell,m}(t;\boldsymbol{\lambda})$, and decomposing it to obtain $X(t;\boldsymbol{\lambda})$. A solution $X(t;\boldsymbol{\lambda})$ for a fixed $\boldsymbol{\lambda}$ is represented as a single (dotted red) vertical line in Fig. 1, which diagramatically represents the surrogate model.

2. Discovering representative binary configurations

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

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

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

One strategy (described in more detail in [35]) to find the greedy parameters begins by evaluating the slow method on a densely sampled *training set*, $\mathcal{T}_{\mathrm{TS}} \subset \mathcal{T}$. This training set is input to a *greedy algorithm* (hence the name greedy

[1] More precisely, we work with the distance-independent dimensionless strain Rh/M, where R is the distance from the binary's center-of-mass and M is the total Christodoulou mass [53] measured after the initial burst of junk radiation [54] has passed. In this paper we choose units so that $c = G = 1$.

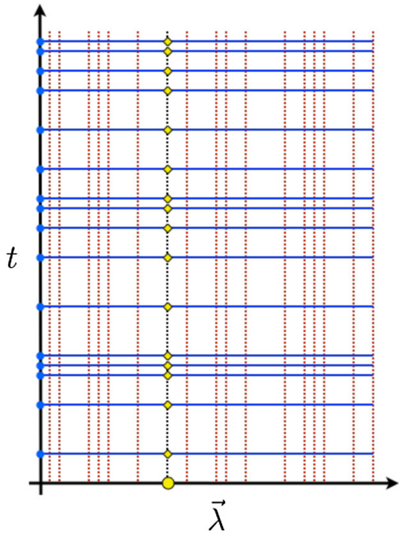

FIG. 1. A schematic of the method for building a surrogate model for a function $X(t;\boldsymbol{\lambda})$. The red dotted lines show $X(t)$ evaluated at a selected set of greedy parameters $\boldsymbol{\Lambda}_i$ used to build a linear basis, and the blue dots show the associated empirical nodes in time from which $X_S(t;\boldsymbol{\lambda})$ can be reconstructed by interpolation with high accuracy. The blue lines indicate fits for $X(t;\boldsymbol{\lambda})$ as a function of $\boldsymbol{\lambda}$ at each of the empirical time nodes. The cyan dot shows a generic parameter $\boldsymbol{\lambda}_0$ that is not in the set of greedy parameters. To compute $X_S(t;\boldsymbol{\lambda}_0)$, each fit is evaluated at $\boldsymbol{\lambda}_0$ (the yellow diamonds), and then the empirical interpolant is used to construct $X_S(t;\boldsymbol{\lambda}_0)$ at arbitrary times (the dotted black line).

parameters) that works as follows. First, the greedy algorithm is initialized by arbitrarily selecting the first few greedy parameters which are sometimes called the algorithm's *seed*.[2]To select the next greedy parameter The set of greedy parameters is then extended iteratively by first building an orthonormal linear basis $B_n = \{e^i(t)\}_{i=1}^n$ spanning the n current greedy solutions, such that

$$X(t;\boldsymbol{\Lambda}_j) = \sum_{i=1}^{n} c_i(\boldsymbol{\Lambda}_j)e^i(t). \tag{3}$$

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

$$X(t;\boldsymbol{\lambda}) \approx \sum_{i=1}^{n} c_i(\boldsymbol{\lambda})e^i(t), \qquad \boldsymbol{\lambda} \in \mathcal{T}_{\mathrm{TS}} \tag{4}$$

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

$$E_n(\boldsymbol{\lambda}) = \left\| X(\cdot;\boldsymbol{\lambda}) - \sum_{i=1}^{n} c_i(\boldsymbol{\lambda})e^i(\cdot)\right\| \tag{5}$$

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

This method unfortunately requires evaluating the slow method on each point in the (large) training set, so we make modifications as described in Secs. II B and III B.

3. Temporal compression

We have built a linear basis B_N which can represent $X(t;\boldsymbol{\lambda})$ for any $\boldsymbol{\lambda} \in \mathcal{T}$ using Eq. (4), up to some small projection error. This reduces the problem of determining $X(t;\boldsymbol{\lambda})$ to determining the basis coefficients $\{c_i(\boldsymbol{\lambda})\}_{i=1}^N$. The most straightforward method of doing so would be to fit or interpolate the basis coefficients c_i over the parameter space $\mathcal{T}$ as is done in [29,36]. We have more intuition for the behavior over $\mathcal{T}$ of the solutions $X(T;\cdot)$ evaluated at a fixed time T than we do for the basis coefficients. We will therefore pursue an *empirical interpolation* approach, described in detail in [35], which enables us to avoid fitting the basis coefficients.

An empirical interpolant makes use of the orthogonal linear basis $B_N = \{e^i(t)\}_{i=1}^N$ such that the errors given by Eq. (5) are small, so Eq. (4) continues to provide a good approximation despite using a different method to compute the coefficients. During the construction of the empirical interpolant, N empirical time nodes $\{T_j\}_{j=1}^N$ will be used. An algorithm to find these special time nodes will be described later on.

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

$$I_N[f](t) = \sum_{i=1}^{N} c_i e^i(t). \tag{6}$$

We then write a linear system of equations

[2] The final set of greedy parameters selected by the greedy algorithm will depend on that choice of seed. However, the number and distribution of greedy parameters is expected to be robust to the choice of seed [55,56].

$$\sum_{i=1}^{N} c_i e^i(T_j) = f(T_j), \qquad j = 1, \ldots, N \tag{7}$$

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

$$c_i = (V^{-1})_{ij} f(T_j). \tag{8}$$

We can then substitute the coefficients back into Eq. (6) to obtain

$$I_N[f](t) = \sum_{i=1}^{N} (V^{-1})_{ij} f(T_j) e^i(t). \tag{9}$$

If we then define

$$b^j(t) = \sum_{i=1}^{N} (V^{-1})_{ij} e^i(t), \tag{10}$$

we obtain

$$I_N[f](t) = \sum_{j=1}^{N} f(T_j) b^j(t). \tag{11}$$

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

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

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

$$X_j(\boldsymbol{\lambda}) = X(T_j;\boldsymbol{\lambda}) \tag{12}$$

by fitting the available data $\{X(T_j;\boldsymbol{\Lambda}) : \boldsymbol{\Lambda} \in G\}$ over the parameter space $\mathcal{T}$. We call these *parametric fits*, and denote the fitted approximation for $X_j(\boldsymbol{\lambda})$ by $X_{jS}(\boldsymbol{\lambda})$. The parametric fits are represented by the blue horizontal lines in Fig. 1. The explicit form of our *surrogate model for X* is then given by

$$X_S(t;\boldsymbol{\lambda}) = \sum_{j=1}^{N} X_{jS}(\boldsymbol{\lambda}) b^j(t). \tag{13}$$

B. Modifications to the basic surrogate modeling approach

A drawback of the algorithm presented in Sec. II A, and of many previous surrogate modeling efforts, is the assumption that the original slow model can be evaluated an arbitrary number of times to build a dense training set. Because of the significant computational expense, this is not feasible for waveforms found by numerically solving the Einstein equations. We can neither build NR surrogate models from dense training data nor can we assess the surrogate's quality at arbitrarily many randomly chosen validation points. In previous work that used computationally inexpensive waveform models [35], thousands of nonspinning waveforms comprised the training set, yet the final surrogate required only a very small subset of greedy parameters G. If we could have predicted G in advance then dense training sets would not be required.

Since we cannot evaluate an arbitrarily large number of NR waveforms, we instead first construct a temporary *mock surrogate* using a simpler waveform model that is both fast to evaluate and is defined in the training region of interest. In this paper, for the purpose of discovering the most relevant parameter values, we build a mock surrogate using the precessing TaylorT4 post-Newtonian (PN) waveform model as implemented in GWFrames [57,58]. We determine the PN greedy parameters G^{PN} using a training set containing many thousands of these PN waveforms, as described in Sec. III B. If we then assume that the distribution of parameters selected using PN waveforms roughly mimics the distribution we would have obtained had NR waveforms been available, then G^{PN} should be a suitable set of greedy parameters for building a NR surrogate. This was found to work well for the nonspinning surrogates of Ref. [59] and, as judged by our validation studies, continues to remain applicable to the precessing waveforms considered here. Instead of PN, we could have used a different analytical waveform model [19–24,27–29] that contains merger and ringdown. However, these other models either omit $\ell = 3$ modes, omit precession, or yield waveforms that do not vary smoothly as a function of $\boldsymbol{\lambda}$. We find that these other considerations outweigh the inclusion of merger and ringdown.

This entire process just described is shown in the first stage of the surrogate workflow diagram (Fig. 2) as the "PN-sampler". Once the points G^{PN} have been selected, the corresponding NR waveforms are generated, and the surrogate building proceeds as in Fig. 2. We emphasize

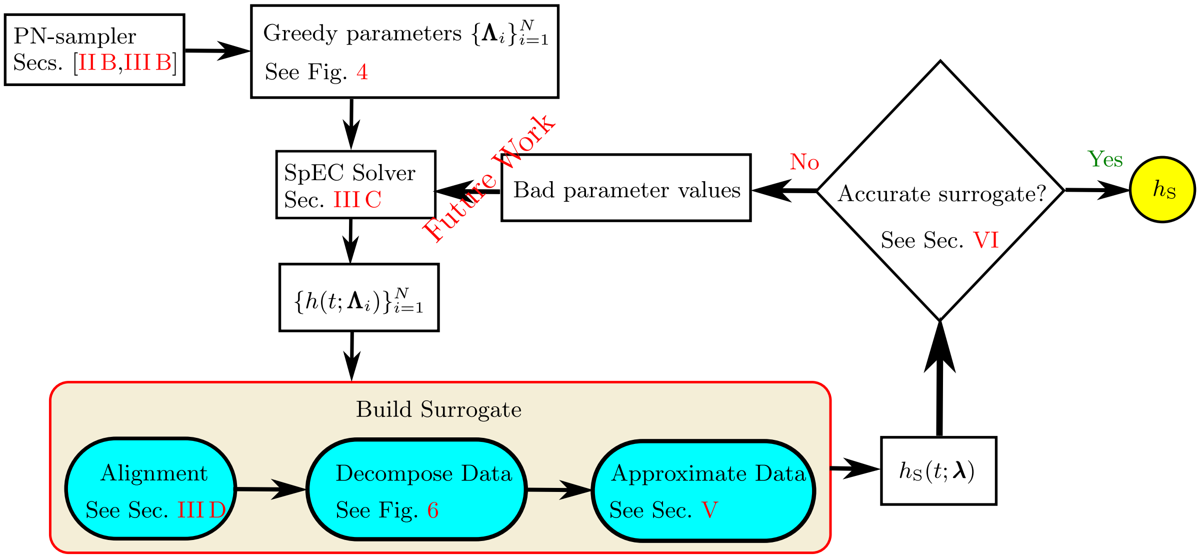

FIG. 2. Surrogate workflow. A greedy "PN-sampler" selects the most informative parameter values $\{\boldsymbol{\Lambda}_i\}_{i=1}^N$ for a fixed parametric and temporal range. For each selected value Λ_i, SpEC generates a gravitational waveform. A surrogate model building algorithm (cf. Fig. 1) is applied to a set of suitably aligned and decomposed (cf. Fig. 6) numerical relativity waveforms thereby producing a trial surrogate. A handful of validation tests are performed to assess the surrogate's quality. If the surrogate performs poorly for some parameter values, one could produce additional numerical relativity waveforms near those values and rebuild a more accurate surrogate.

that *no* PN waveforms are used to build the resulting NRSur4d2s surrogate; the PN model is used only to find the greedy parameters $G = G^{\mathrm{PN}}$.

While the PN waveforms are much cheaper to evaluate than NR waveforms, building a dense training set remains costly for high dimensional parameter spaces. In Ref. [59], it was found that an accurate basis can be achieved using small, sparse training sets if each iteration i of the greedy algorithm uses an independent randomly sampled training set $\mathcal{T}_{\mathrm{R}}^i$. We extend this methodology by also including in our training sets a fixed set of parameters $\mathcal{T}_{\mathrm{B}}$ on the boundary of $\mathcal{T}$ (for example, the maximum mass ratio allowed in $\mathcal{T}$). This is motivated by the fact that the boundary of $\mathcal{T}$ carries significant weight both when building a linear basis and when performing parametric fits. At the ith greedy iteration, we then have

$$\mathcal{T}_{\mathrm{TS}}^i = \mathcal{T}_{\mathrm{B}} \cup \mathcal{T}_{\mathrm{R}}^i \tag{14}$$

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

Another issue with the standard greedy algorithm is that it considers only a single function X. For modeling waveforms, we will decompose each waveform $h^{\ell m}(t;\boldsymbol{\lambda})$ into many such functions, which we call *waveform data pieces* (cf. Sec. IV). Rather than generate a separate set of greedy parameters G_X^{PN} for each X, we construct a single set of greedy parameters G^{PN} that can be used for all waveform data pieces X. We do so by replacing the projection errors for a single waveform data piece given in Eq. (5) with a single error including contributions from all waveform data pieces. This will be described explicitly in Sec. III B after the waveform decomposition and error measures have been introduced.

The standard greedy algorithm guarantees that the basis yields small projection errors given by Eq. (5). Therefore, if we have perfect parametric fits [so that $X_{jS}(\boldsymbol{\lambda}) = X_j(\boldsymbol{\lambda})$ for all $\boldsymbol{\lambda} \in \mathcal{T}$] then the surrogate model X_S given by Eq. (13) will agree with X in the sense that the L^2 norm of $X_S(t;\boldsymbol{\lambda}) - X(t;\boldsymbol{\lambda})$ will be small for all $\boldsymbol{\lambda} \in \mathcal{T}$. There is, however, no corresponding guarantee that the greedy points are sufficient for producing accurate parametric fits X_{jS}. In the one-dimensional models built in Refs. [34,35], the parametric fits performed well using the samples produced from the standard greedy algorithm. As the dimensionality of the parameter space increases, the number of greedy parameters required for an accurate basis grows slowly [60], but the number of samples required for accurate parametric fits can grow rapidly. We therefore anticipate that the standard greedy algorithm alone may lead to underresolved parametric fits in problems with high dimensionality.

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

C. Handling noise in the NR waveforms

The presence of numerical noise in the input NR waveforms complicates the construction of surrogates. The situation is simpler when building a surrogate of a waveform model that is mostly noise-free, such as post-Newtonian or EOB models that require the solution of ordinary differential equations (which can be evaluated to almost arbitrary accuracy) but not partial differential equations. For example, Ref. [35] demonstrates in their Fig. 15 that EOB surrogates can be made to have arbitrarily small errors, and Refs. [29,36] use interpolation across the parameter space without needing to avoid potential pitfalls such as overfitting the noise. We do not expect this to be the case for numerical relativity waveforms which are beset by numerous error sources, some of which cannot be made arbitrarily small with current computing technology.

Systematic as well as numerical errors can influence the quality of the NR waveform. For example, when attempting to model noneccentric binaries, the NR simulations will always have some small but nonzero orbital eccentricity. In this paper we will mostly focus on numerical truncation error. This is typically the dominant source of error in SpEC waveforms [61], and the other sources of error are expected to be significantly smaller than truncation error and smaller than the surrogate error (see Fig. 3 of Ref. [34]). The numerical error can be quantified through standard convergence tests [61]. Following Ref. [34], we will (i) characterize SpEC waveform error across the parameter space and, if necessary, remove poorly resolved waveforms (Sec. IV E) (ii) avoid overfitting the noise sources (Appendix A), and (iii) set surrogate accuracy goals based on our answer to the first question. In future work it would be interesting to study the impact of other noise sources.

D. Decomposing NR waveforms into simpler components

The detailed time dependence of an NR waveform is generally too complicated to model directly with an acceptable degree of accuracy. Instead, each NR waveform is decomposed into *waveform data pieces* (cf. Sec. IV), which are simpler, more slowly varying functions that can be modeled more easily. A surrogate model is then built for each waveform data piece (cf Sec. V), and then these models are recombined to produce a full surrogate waveform. This process is shown in the "surrogate build" step shown in Fig. 2.

Selecting the waveform data pieces is a critically important step. For example, in nonspinning [34,35] and spin-aligned [29,36] surrogate models, the waveform data pieces are either the real and imaginary parts of the SWSH modes, $h^{\ell m}$, or the amplitude and phase decompositions of these modes $A_{\ell m}$ and $\phi_{\ell m}$, where $h^{\ell m} = A_{\ell m} \exp(-\mathrm{i}\phi_{\ell m})$. The idea is that it is easier to model every $A_{\ell m}$ and $\phi_{\ell m}$, which are smooth and slowly varying functions of time, than it is to directly model the complicated waveform $h(t, \theta, \phi; \boldsymbol{\lambda})$, Eq. (2).

Because of the complexity of precessing waveforms, we have needed to pursue a somewhat more complicated decomposition scheme than in the nonprecessing case. Figure 6 summarizes the decomposition scheme used here. Briefly, each waveform is transformed into a coordinate frame in which the binary is not precessing [45–47,57,62]; specifically, we use the minimal-rotation coprecessing frame of Boyle [47]. The waveform modes in this frame have a simpler structure than their inertial frame counterparts. Additional simplifications occur by applying further transformations (described in detail in Sec. IV) to the coprecessing-frame waveform modes. The result of these steps is a set of waveform data pieces. If $X(t, \boldsymbol{\lambda})$ is a single waveform data piece, then for that piece we build a surrogate $X_{\mathrm{S}}(t, \boldsymbol{\lambda}) \approx X(t, \boldsymbol{\lambda})$. Here X can stand for any of the decomposed waveform data pieces depicted as cyan ellipses in Fig. 6, for example A_{+}^{22}, ϕ_{-}^{32}, φ_p, etc. The full NRSur4d2s surrogate waveform model is defined by the individual data piece surrogates, $X_{\mathrm{S}}(t, \boldsymbol{\lambda})$, and the inverse transformations required to move back up the data decomposition diagram (Fig. 6) and reconstruct the waveform from all of the $X_{\mathrm{S}}(t, \boldsymbol{\lambda})$.

E. Tools for surrogate model validation

Here we describe a useful framework for assessing the surrogate's predictive quality when only a limited number of waveforms are available. This is a different setting from the EOB surrogates of [29,36] where *out-of-sample* validation studies could be performed at arbitrarily many parameter values. The primary tool we shall use is *cross-validation* [63], which was also used in [34]. Cross-validation happens after the surrogate is built and determines whether or not more SpEC waveforms are needed to improve the accuracy of the model (see Fig. 2).

We consider the case where our full data set is composed of N SpEC waveforms. From the full data set, we select nonintersecting sets of trial and verification waveforms with sizes N_{t} and N_{v}, such that $N_{\mathrm{t}} + N_{\mathrm{v}} \leq N$. In the cross-validation step, a new *trial surrogate* is built solely from N_{t} trial waveforms. The remaining N_{v} verification waveforms serve as an *exact and independent* error measure of the *trial* surrogate's prediction. The key assumption, which we believe to be true in practice, is that the surrogate built from all N waveforms will have an accuracy similar to the trial surrogates, if not better. Indeed, each step of the surrogate building algorithm will be more accurate so long as parametric overfitting is kept under control. Hence, the trial surrogate's error should serve as a useful estimate of the error associated with the full surrogate built from all N

waveforms. We note, however, that when $N_{\rm v}$ is small or the surrogate error is dominated by some systematic source of error, the improved accuracy when including all N waveforms may not be enough to overcome the variance in the accuracy of the parametric fits seen in Fig. 5. In that case the full surrogate error may in fact be slightly larger than a trial surrogate error.

Two variants of cross-validation are considered. *Random* cross-validation proceeds by selecting the verification waveform set randomly. When $N_{\rm v} = 1$ this is known as the *leave-one-out* strategy. In Ref. [34], all possible leave-one-out studies were performed. In our case, N is sufficiently large and surrogate-building is sufficiently expensive that we opt to choose $N_{\rm v} = 10$. We can perform many resamplings of the validation subset to infer an error profile across the parameter space.

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

F. Waveform error measurements

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

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

$$\delta h(t) = \sqrt{\sum_{\ell,m} |\delta h^{\ell m}(t)|^2}, \qquad (15)$$

from the individual mode differences

$$\delta h^{\ell m}(t) = h_1^{\ell m}(t;\boldsymbol{\lambda}_1) - h_2^{\ell m}(t;\boldsymbol{\lambda}_2). \qquad (16)$$

We use the time-domain inner product

$$\langle a, b\rangle_t = \frac{1}{T}\int_{t_{\rm min}}^{t_{\rm max}} a(t) b^*(t) dt, \qquad (17)$$

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

$$(\delta h)^2 = \int_{S^2} \|h_1(t,\theta,\phi;\boldsymbol{\lambda}_1) - h_2(t,\theta,\phi;\boldsymbol{\lambda}_2)\|^2 d\Omega \qquad (18)$$

$$= \sum_{\ell m} \|\delta h^{\ell,m}\|^2 \qquad (19)$$

$$= \frac{1}{T}\int_{t_{\rm min}}^{t_{\rm max}} \delta h(t)^2 dt \qquad (20)$$

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

$$\mathcal{E}[h_1, h_2] = \frac{1}{2}\frac{\delta h^2}{\|h_1\|^2}, \qquad (21)$$

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

$$\mathcal{O} = \frac{\langle h_1, h_2\rangle}{\sqrt{\langle h_1, h_1\rangle\langle h_2, h_2\rangle}}. \qquad (22)$$

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

Overlap errors are often computed in the frequency domain with a noise-weighted inner product [64]

$$\langle a, b\rangle_f = 4{\rm Re}\int_{f_{\rm min}}^{f_{\rm max}} \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df, \qquad (23)$$

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

III. POPULATING THE SET OF NR WAVEFORMS

A. Parameter space

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

$$A_{\mathrm{tot}}(t) = \sqrt{\sum_{\ell,m} |h^{\ell,m}(t)|^2}, \tag{24}$$

reaches its maximum value.

We restrict to a 5D subspace of the parameter space where $\vec{\chi}_2$ is aligned with the Newtonian orbital angular momentum $\hat{L}_N$ at the reference time (in practice the NR simulations give us small misalignments but ignore them; see Sec. III E). Let θ_χ and ϕ_χ be the polar and azimuthal angles of $\vec{\chi}_1$ at the reference time. Then our five parameters are q, $|\vec{\chi}_1|$, χ_2^z, θ_χ, and ϕ_χ (see Fig. 3). While NR simulations can be done for nearly extremal spins [13] and large mass ratios [15], they are computationally expensive and so we restrict to $|\vec{\chi}_1| \leq 0.8$, $|\chi_2^z| \leq 0.8$ and $q \leq 2$. These bounds were also motivated by the parameters of GW150914, which was close to equal mass and did not show strong evidence of large spin magnitudes [3].

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

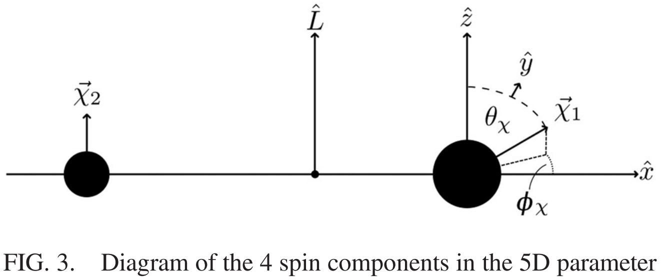

FIG. 3. Diagram of the 4 spin components in the 5D parameter subspace. We attempt to obtain $\phi_\chi = 0$ at $t_0 = 4500M$ before peak amplitude, but in practice the NR simulations have arbitrary values of ϕ_χ.

B. Selection of greedy parameters

We use $\phi_\chi = 0$ while determining the greedy parameters $G = \{\mathbf{\Lambda}_i\}$, and we use PN waveforms to identify the most relevant and distinct points in parameter space as outlined in Secs. II A and II B. We first seed G with the parameter space corner cases: $q \in \{1, 2\}$, $|\vec{\chi}_A| \in \{0, 0.8\}$, $\theta_\chi \in \{0, \pi\}$ and $\chi_2^z \in \{-0.8, 0.8\}$. As described in Eq. (14), we compute training sets $\mathcal{T}_{\mathrm{TS}}^i = \mathcal{T}_{\mathrm{B}} \cup \mathcal{T}_{\mathrm{R}}^i$ consisting of a set of boundary parameters $\mathcal{T}_{\mathrm{B}}$ as well as a set of randomly-sampled parameters $\mathcal{T}_{\mathrm{R}}^i$ that is resampled at each greedy iteration i. For $\mathcal{T}_{\mathrm{B}}$, we use a set of 216 points where two components of λ take on one of their extremal values and the other two are one of three intermediate (nonboundary) values. This results in features that can be seen in Fig. 4,

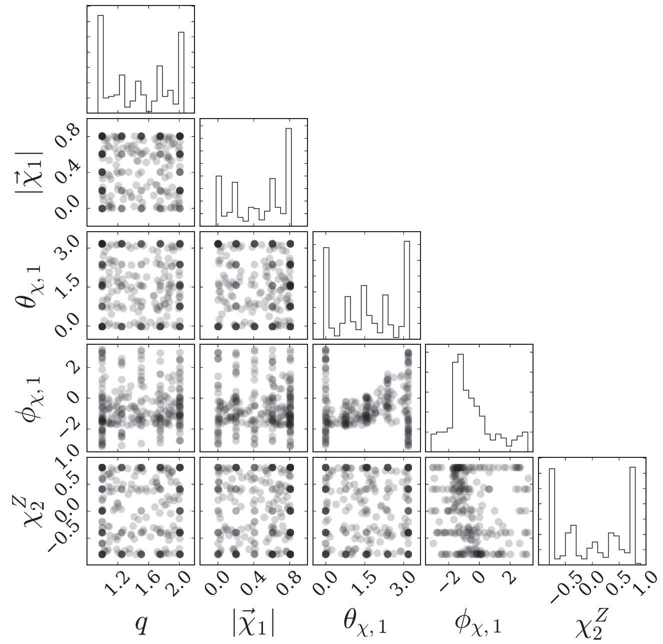

FIG. 4. A "triangle plot" showing all possible two-dimensional projections and one-dimensional histograms of the greedy parameters G selected by the procedure of Sec. III B. These are the parameters used for the numerical relativity simulations. Made using the Python package corner.py [65].

where the two boundary values and three intermediate values occur frequently. For example, because some 2D projections of these special points are selected multiple times, they appear as darker points around the boundary of some of the subplots in Fig. 4. In addition, subplots involving ϕ_χ show an uneven distribution of stripes that occur at these special points. For $\mathcal{T}^i_{\rm R}$, we randomly sample each parameter component uniformly in its range.

Next, we add parameters to G using an initial greedy algorithm that uses basis projection errors to select greedy parameters. Given a point $\boldsymbol{\lambda}$ as a candidate that might be added to G, we compute a PN waveform h corresponding to $\boldsymbol{\lambda}$, we decompose h into waveform data pieces (see Sec. IV), and we project each waveform data piece onto their respective bases. Then we recombine the projected waveform data pieces to produce a waveform $h_{\rm proj}$. We then compute an error $\mathcal{E}[h, h_{\rm proj}]$ using Eq. (21). The point in $\mathcal{T}^i_{\rm TS}$ with the largest such error is the next point added to G. This method is different than that of [59], in which projection errors of each waveform data piece were computed separately, and then these errors were combined in a weighted sum with coefficients determined by hand. Our new method avoids the need to determine these coefficients, and automatically ensures that the most significant waveform data pieces are resolved accurately. We use this initial greedy algorithm until the error is $\mathcal{E} \leq 10^{-5}$. At this point, the number of greedy points is approximately $|G| = 30$. Thus we have built a linear basis for each waveform data piece. For each iteration of this initial greedy algorithm, we choose the number of randomly sampled parameters to be $|\mathcal{T}^i_{\rm R}| = 10 + 2|G^i|$, where $|G^i|$ is the number of greedy parameters at the start of the ith iteration.

Finally, we add parameters to G using a second greedy algorithm that uses surrogate errors to select greedy parameters. At each iteration i, we construct a new trial PN waveform surrogate (as described in Appendix E), using the greedy parameters G^i, and then for each point $\boldsymbol{\lambda} \in \mathcal{T}^i_{\rm TS}$, we evaluate this surrogate and compare it to the corresponding PN waveform by computing $\mathcal{E}$. The parameter $\boldsymbol{\lambda}$ that maximizes this error is used as the next greedy parameter and is added to G. This error includes the errors in the parametric fits for each empirical node of all waveform data pieces; the parametric fits are shown as blue lines in Fig. 1 and are described in detail in Sec. V B and Appendixes A and E. For this step, we use $|\mathcal{T}^i_{\rm R}| = 6|G|$. The maximum errors found in each iteration of this second greedy algorithm are shown in Fig. 5 as a function of $|G|$. The noise is due to the random resampling of the training set, as well as the possibility of the parametric fits becoming worse by adding a data point. Because the parametric fits are restricted to a particular order, the surrogate error in Fig. 5 does not go below 10^{-3}. In principle one can reduce this error floor by increasing the order of the fits, but here we simply keep only the first 300 greedy parameters. We perform NR simulations for these 300 parameters, except for those parameters that can be obtained from other parameters by symmetry, for example by exchanging the black hole labels. These symmetry considerations reduce the number of simulations to 276.

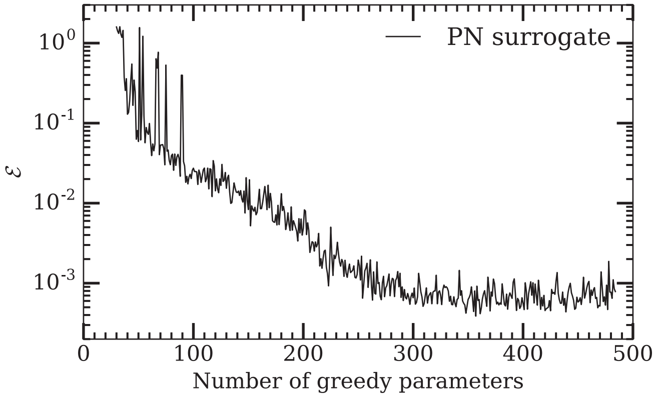

FIG. 5. Maximum surrogate errors found during the second greedy algorithm (see Sec. III B) for determining $\boldsymbol{\Lambda}_i$ using trial PN surrogates. The noise is due to the random resampling, as well as the possibility of the parameter space fits becoming worse by adding a data point. The finite order of the fits leads to an error floor of 10^{-3}, so we keep and perform NR simulations for only the first 300 greedy parameters.

C. Numerical relativity simulations

To build our time-domain model, we use the 276 NR waveforms computed by the SXS collaboration with the SpEC described in Ref. [61]. Each NR simulation is performed at three different numerical resolutions, labeled “Lev1”, “Lev2”, and “Lev3”, in order of increasing resolution; Lev$i$ has an adaptive-mesh-refinement (AMR) error tolerance that is a factor of 4 smaller than Lev$i-1$. For each resolution, the waveform is extracted at multiple finite radii from the source, and then the waveform is extrapolated to future null infinity [66]. The extrapolation is done using an Nth order polynomial in $1/r$, where r is a radial coordinate. To estimate errors in extrapolation, we perform extrapolation with several values of N [66]. Similarly, to estimate numerical truncation error, we compare simulations that are identical except for resolution [61]. However, for building surrogates, we always use the highest available resolution (Lev3) simulations, and use the $N = 2$ extrapolated waveforms. The simulations begin at a time of ~5000M before merger where $M = m_1 + m_2$ and m_i are the Christodoulou masses of each black hole. We ignore the small eccentricities present in the simulations, which have a median of 0.00029 and a maximum of 0.00085 for the highest resolution simulations. The masses we use are those measured after the initial burst of junk radiation [54] leaves the computational domain.

The BH spin vectors are measured on the apparent horizons of the BHs during the evolution of the NR simulation. The spin directions are therefore gauge-dependent. The potential concern is that when the surrogate model is evaluated, the spin directions must be provided with the gauge used to build the model, so that the spin directions obtained in gravitational wave parameter estimation can be interpreted correctly. However, it has been found that the time-dependent spin and orbital angular momentum vectors in the damped harmonic gauge used by SpEC agree very well with the corresponding vectors in PN theory [67]. Therefore, this is of no more concern than the interpretation of spin directions with PN-based gravitational wave models.

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

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

D. Waveform alignment

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

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

To align the waveforms, we shift them in time such that the peak of the total waveform amplitude as given in Eq. (24) occurs at $t = 0$. We then use a cubic spline to interpolate the real and imaginary parts of the waveform onto a uniformly spaced time series with $dt = 0.1M$. Next, we rotate the waveforms to align the orientation of the binary at $t_0 = -4500M$ in two steps: first we perform an approximate rotation using the black hole trajectories, and then we perform a small correction using only the waveform. For the initial approximate rotation, we use the horizon trajectory to align the Newtonian orbital angular momentum with $\hat{z}$ and rotate about $\hat{z}$ such that black hole 1 lies along the positive $\hat{x}$ axis. We then use the waveform modes to perform an additional rotation, aligning the principal eigenvector of the angular momentum operator [47] with $\hat{z}$ and equating the phases of $h^{2,2}$ and $h^{2,-2}$ at $t = t_0$. The first coarse alignment was used since the second alignment is ambiguous—we can change the sign of the principal eigenvector and/or rotate by an additional π about $\hat{z}$, which we resolve by choosing the smallest of the rotations, since the waveform is already nearly aligned.

We perform identical rotations on the spin directions and then measure them at t_0.

E. Postalignment parametrization

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

(i) $q = \frac{m_1}{m_2} \in [0.9999, 2.0005]$
(ii) $|\vec{\chi}_1| \in [0, 0.801]$
(iii) $\theta_\chi \equiv \cos^{-1}(\frac{\chi_1^z(t_0)}{|\vec{\chi}_1|}) \in [0, \pi]$
(iv) $\phi_\chi \equiv \arctan 2(\chi_1^y(t_0), \chi_1^x(t_0)) \in (-\pi, \pi]$
(v) $\chi_2^z(t_0) \in [-0.8, 0.800006]$

as shown in Fig. 3. We will often omit the time dependence of the last parameter and simply write χ_2^z.

IV. WAVEFORM DECOMPOSITION

This section describes how each input NR waveform is decomposed into a set of “waveform data pieces”, which are simple functions that can be modeled easily and can be recombined to produce the original waveform. This decomposition was outlined briefly in Sec. II D, and a flowchart of this process is shown in Fig. 6.

We write each input waveform as a set of modes $H = \{h^{\ell,m}(t)\}$, with $t \in [t_{\rm min}, t_{\rm max}]$. Here $t_{\rm min}$ and $t_{\rm max}$ are chosen to be the same for all waveforms and are selected in the following way: Recall that each waveform is time-shifted so that the maximum amplitude occurs at $t = 0$; this means that each time-shifted finite-length NR waveform H_i has a different beginning time $t_i^{\rm begin}$ and a different ending time $t_i^{\rm end}$. We choose $t_{\rm min} = \max_i(t_i^{\rm begin}) + 150M$ and $t_{\rm max} = \min_i(t_i^{\rm end})$. The value $150M$ is chosen to remove the worst of the “junk radiation” [54] that results from the failure of NR initial data to precisely describe a quasiequilibrium inspiral. Although the surrogate output will cover only the smaller time interval $[t_0, t_f = 70M]$, we use waveforms over the larger time interval $[t_{\rm min}, t_{\rm max}]$ in order to mitigate edge effects that can occur in later steps in the decomposition process (filtering and Hilbert transforms, described below). Selected modes of H are shown in Fig. 7.

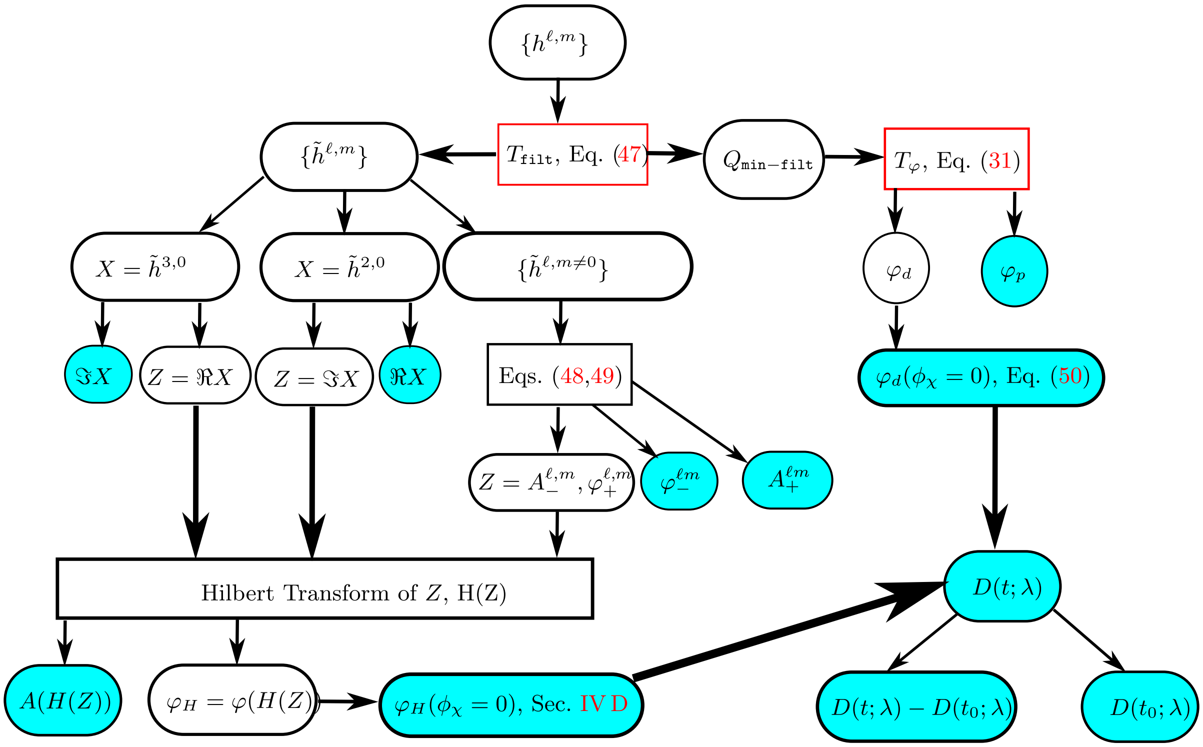

FIG. 6. Waveform decomposition schematic. A series of decompositions are applied to a set of NR waveform modes $\{h^{\ell m}\}$ yielding easier-to-approximate waveform data pieces (shown as cyan ellipses) for which we ultimately fit. Two types of objects are shown: time series data as an ellipse and operators/maps as rectangles. A red outlining border identifies an object which uses a modeling approximation which will *not* go away with additional NR waveforms. These *decomposition errors* are quantified and shown to be smaller than other sources of error in Sec. VI. An additional source of error that will not converge away with more NR waveforms results from the assumption that each data piece transform in a simplistic way with changes of ϕ_χ.

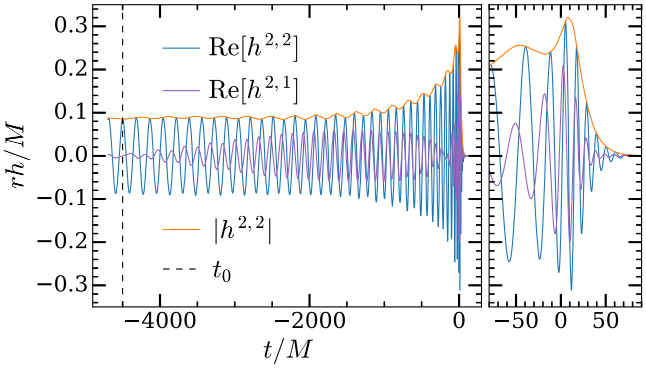

FIG. 7. Waveform modes in the inertial frame for SXS: BBH:0338 with $q = 2$, $|\vec{\chi}_1| = 0.8$, $\theta_\chi = 1.505$, $\phi_\chi = -1.041$ and $\chi_2^z = 0.8$. For precessing systems, all $\ell = 2$ modes contain significant power in the inertial frame. The NR waveform is aligned to have the canonical orientation at $t = t_0$.

A. Transforming to a coprecessing frame

The first step in the waveform decomposition is transforming to a rotating coordinate frame in which the binary is not precessing. Thus the original waveform is described by a (much simpler) waveform in this coprecessing frame, plus functions that describe the time-dependent rotation. We transform[3] H to the minimally-rotating coprecessing frame of Ref. [47], and thereby obtain the waveform modes $\tilde{H} = \{\tilde{h}^{\ell,m}(t)\}$ in this frame, as well as a time-dependent unit quaternion $q(t)$ that describes the rotation of the frame. Throughout this section we will use a tilde, i.e., $\tilde{h}^{\ell,m}(t)$, to denote a time-domain waveform mode in the coprecessing frame, as opposed to the Fourier transform of a waveform mode. Selected modes of $\tilde{H}$ are shown in Fig. 8. We denote this transformation by

$$T_C \colon H \to (\tilde{H}, q), \tag{25}$$

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

$$T_Q \colon (H', q) \to H \tag{26}$$

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

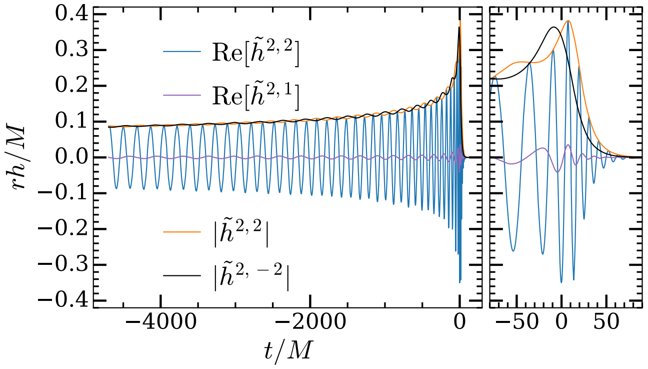

FIG. 8. Waveform modes in the coprecessing frame for SXS: BBH:0338. The mode power hierarchy is now the same as for a nonprecessing waveform, with the $(2, \pm 2)$ modes dominating, but small effects of precession are still present in the mode amplitudes and phases. The amplitudes of the $(2, \pm 2)$ modes have small nearly opposite oscillations.

The unit quaternion $q(t)$ has four components shown as solid lines in Fig. 9. However, the minimally rotating coprecessing frame constrains $q(t)$ so as to minimize the magnitude of the frame's instantaneous angular velocity (the "minimal rotation condition") [47]. This condition, combined with the unit norm, imply that $q(t)$ has only two independent components.

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

$$dq(t) = q^{-1}(t)q(t + dt) = 1 + 2\vec{\omega}(t)dt + \mathcal{O}(dt^2). \tag{27}$$

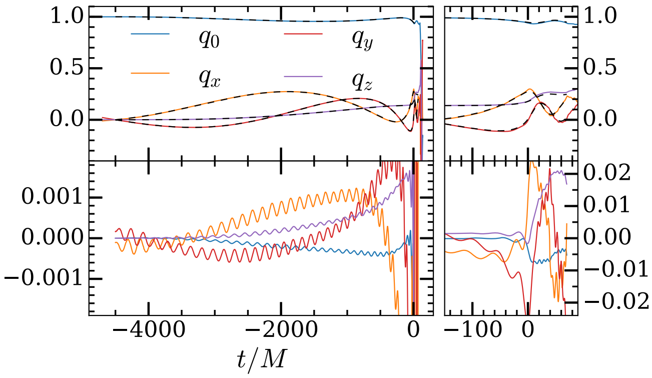

FIG. 9. Top: Quaternion q representing the time-dependent rotation from the coprecessing frame to the inertial frame (solid lines) and the filtered quaternion $q_{\mathrm{min-filt}}$ (dashed lines) for SXS: BBH:0338. Bottom: Differences between the filtered and unfiltered quaternions. This difference results in an error when reconstructing the waveform in the inertial frame, contributing to a "decomposition" error in the surrogate model.

[3] Throughout this work we use GWFrames [57,58] to enact our transformations.

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

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

$$\delta q(t) = q^{-1}(t)q(t+\delta t) = s(t) + \delta t\vec{u}(t), \tag{28}$$

where the scalar component $s(t)$ is $1 + \mathcal{O}(\delta t^2)$. Thus, for a given δt, Eq. (28) defines $\vec{u}(t)$ in terms of $q(t)$, and furthermore, $\vec{u}(t)$ approaches $\frac{1}{2}\vec{\omega}(t)$ as $\delta t \to 0$. We find that if we use $\delta t = 0.1M$, the $\vec{u}(t)$ we obtain is sufficiently close to this limit that the error we make is negligible compared to other errors; this error is included in the *decomposition error* discussed in Sec. VI. Finally, instead of using $\vec{u}(t)$ directly as independent components of $q(t)$, we define $\tilde{\omega}_p = \frac{1}{2}|\vec{u}(t)|$ and

$$\varphi_p(t) = \delta t \sum_{\tau<t} \tilde{\omega}_p(\tau) \tag{29}$$

$$\varphi_d(t) = \arg(u_x(t) + iu_y(t)). \tag{30}$$

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

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

$$T_\varphi\colon\ q \to (\varphi_p, \varphi_d). \tag{31}$$

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

$$\tilde{\omega}_p(t) = \frac{\varphi_p(t+dt) - \varphi_p(t)}{\delta t} \tag{32}$$

$$u_x(t) = 2\tilde{\omega}_p(t)\cos(\varphi_d(t)) \tag{33}$$

$$u_y(t) = 2\tilde{\omega}_p(t)\sin(\varphi_d(t)) \tag{34}$$

$$u_z(t) = 0 \tag{35}$$

$$s(t) = \sqrt{1 - (2\tilde{\omega}_p(t)\delta t)^2} \tag{36}$$

$$\delta q(t) = s(t) + \vec{u}(t)\delta t. \tag{37}$$

We include the $(\delta t)^2$ term in Eq. (36) so that the reconstructed $q(t)$ has unit norm. Because we assume $\delta q_z = 0$, the δq we compute in Eq. (37) is not exactly the δq we started with in Eq. (28); however, the error we make is only $\mathcal{O}(\delta t^3)$. Given $q(t)$ and δq, we can then compute $q(t+\delta t)$ using

$$q(t+\delta t) = q(t)\delta q(t), \tag{38}$$

which results in an $\mathcal{O}(\delta t^2)$ error in $q(t+\delta t)$. Because we have $q(t_0) = 1$ at the alignment time t_0, we can use the recurrence relation Eq. (38) to construct $q(t)$ at all times, given $\varphi_p(t)$ and $\varphi_d(t)$.

B. A "filtered" coprecessing frame

The quaternion $q(t)$ representing the coprecessing frame oscillates mostly on the slow precession time scale, which makes it easier to model. However, it also has small oscillations on the much faster orbital time scale, as shown by the purple curve in the bottom plot of Fig. 10. These oscillations are due to the nutation of the rotation axis of the coprecessing frame, relative to the inertial frame. These small oscillations can make it more difficult to fit φ_d across parameter space. Since the effect of the nutation on the inertial frame waveform is small, we filter out the nutation in the coprecessing frame. We use a Gaussian filter with a width of π radians of the orbital phase, which is computed from the angular velocity of the waveform as described in [68]. Near the edges of the domain, we truncate the filter on both sides to keep the filter centered. Specifically, if the (monotonic) orbital phase is given by $\varphi_{\mathrm{orb}}(t)$, then we can invert the relationship to find $t(\varphi_{\mathrm{orb}})$. For a given time τ with corresponding orbital phase $\varphi^* = \varphi_{\mathrm{orb}}(\tau)$ we then compute

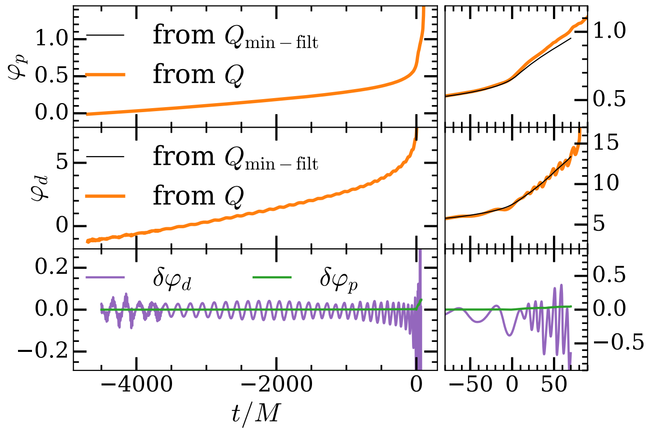

FIG. 10. Phases φ_p (top) and φ_d (middle) for SXS:BBH:0338. These phases represent the total amount of precession and the instantaneous direction of precession respectively. Shown are phases computed from the unfiltered coprecessing quaternion (thick orange lines) and the filtered quaternion (thin black lines). The orbital time scale oscillation in φ_d is suppressed after filtering. Bottom: Differences between the filtered and unfiltered phases.

$$\varphi_{\mathrm{min}} = \varphi_{\mathrm{orb}}(t_{\mathrm{min}}) \tag{39}$$

$$\varphi_{\mathrm{max}} = \varphi_{\mathrm{orb}}(t_{\mathrm{max}}) \tag{40}$$

$$\Delta\varphi = \min(4\pi, |\varphi^* - \varphi_{\mathrm{min}}|, |\varphi^* - \varphi_{\mathrm{max}}|) \tag{41}$$

$$\varphi_\pm = \varphi^* \pm \Delta\varphi \tag{42}$$

$$G(\varphi) = \exp\left[-\left(\frac{\varphi - \varphi^*}{\pi}\right)^2\right] \tag{43}$$

$$q_{\mathrm{init-filt}}(\tau) = \frac{\int_{\varphi_-}^{\varphi_+} q(t(\varphi))G(\varphi)d\varphi}{\int_{\varphi_-}^{\varphi_+} G(\varphi)d\varphi} \tag{44}$$

$$q_{\mathrm{filt}}(\tau) = \frac{q_{\mathrm{init-filt}}(\tau)}{|q_{\mathrm{init-filt}}(\tau)|}. \tag{45}$$

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

$$\tilde{H}', q_{\mathrm{min-filt}} = T_C(T_Q(\tilde{H}, q_{\mathrm{filt}})) \tag{46}$$

and use the frame corresponding to $q_{\mathrm{min-filt}}$, which is minimally rotating and has much less nutation than the frame corresponding to q. The components of the filtered quaternion $q_{\mathrm{min-filt}}$ are shown in Fig. 9 as dashed lines. We use $\tilde{H}$, and not $\tilde{H}'$, as the filtered coprecessing waveform, because $\tilde{H}'$ is not as slowly varying as $\tilde{H}$ and is therefore slightly more difficult to fit. We have verified that the error in the final model caused by choosing $\tilde{H}$ instead of $\tilde{H}'$ is small compared to other errors. Note that even if we choose $\tilde{H}'$, introducing a filter produces some information loss, and therefore results in some error in the final surrogate model. This *decomposition error* is discussed in Sec. VI and is plotted in Figs. 12 and 13. We thus denote the filtering transformation by

$$T_{\mathrm{filt}}\colon (\tilde{H}, q) \to (\tilde{H}, q_{\mathrm{min-filt}}). \tag{47}$$

Applying T_φ to $q_{\mathrm{min-filt}}$ results in less oscillatory behavior in φ_d than when T_φ is applied to q, as seen in Fig. 10. When evaluating the surrogate and reconstructing the inertial frame waveform, we do not attempt to invert T_{filt} which contributes to the decomposition errors shown in Fig. 13.

C. Decomposition of coprecessing-frame waveforms

Once we have computed waveform modes $\tilde{h}^{\ell,m}(t)$ in the coprecessing frame, we decompose each of these modes (except for the $m = 0$ modes, which are discussed separately below) into an amplitude and a phase. However, these amplitudes and phases are difficult to model because they contain oscillations on the orbital time scale. These oscillations are due to asymmetries of waveforms from precessing systems and cannot be completely removed with a different choice of frame [62]. Figure 8 shows an example of these oscillations. To better model the amplitudes and phases of $\tilde{h}^{\ell,m}(t)$, we seek to further decompose them into simpler slowly varying functions. To do this, first note that the amplitudes of $\tilde{h}^{2,2}(t)$ and $\tilde{h}^{2,-2}(t)$ shown in Fig. 8 oscillate in opposite directions. The same is true for the phases, although it is not apparent in the figure, and it is also true for some (but not all) higher-order modes. This motivates the use of symmetric and antisymmetric amplitudes and phases

$$A_\pm^{\ell,m}(t) = \frac{1}{2}(|\tilde{h}^{\ell,m}(t)| \pm |\tilde{h}^{\ell,-m}(t)|) \tag{48}$$

$$\varphi_\pm^{\ell,m}(t) = \frac{1}{2}(\varphi(\tilde{h}^{\ell,m}(t)) \pm \varphi(\tilde{h}^{\ell,-m}(t))) \tag{49}$$

for $m > 0$ where $\varphi(x(t)) = \arg(x(t))$. The symmetric amplitude $A_+^{2,\pm2}(t)$ and the antisymmetric phase $\varphi_-^{2,\pm2}(t)$ contain almost no oscillations and are slowly varying, so we use these as waveform data pieces. However, the antisymmetric amplitude $A_-^{2,\pm2}(t)$ and the symmetric phase $\varphi_+^{2,\pm2}(t)$ of the $(2,\pm2)$ mode are small oscillatory real functions, so to model them we taper each of these functions in the intervals $[t_{\mathrm{min}}, t_0]$ and $[t_f, t_{\mathrm{max}}]$ with a Planck window [69] and take a Hilbert transform, thereby producing an amplitude and phase for each of these functions; these amplitudes and phases are slowly varying, so we use these as our waveform data pieces.

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

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

Now consider the general case with arbitrary $\vec{\chi}_1$. At the alignment time t_0, the orbits of all NR waveforms are aligned. Because of this, at time t_0 the phase of a given (ℓ, m) mode with $m > 0$ and even will be approximately equal for all NR waveforms, i.e. for all choices of parameters. Similarly, at time t_0 the phase of a given (ℓ, m) mode with m odd will either be approximately equal

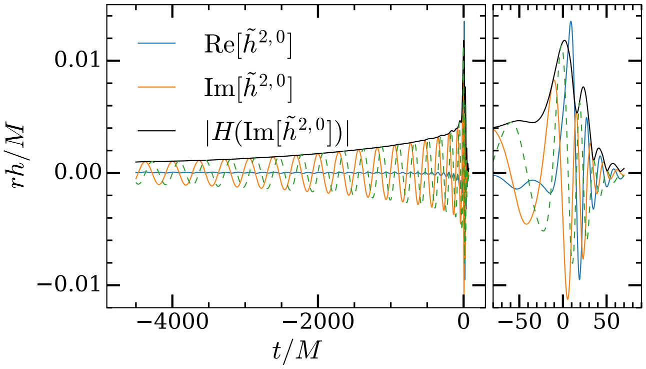

FIG. 11. For the real-valued oscillatory components X such as $\mathrm{Im}[\tilde{h}^{2,0}]$, we perform a Hilbert transform to obtain a complex signal $H(X)$ and extract an amplitude and phase. The dashed green line shows the imaginary part of $H(X)$.

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

Now we discuss modes $\tilde{h}^{\ell,0}(t)$, with $m = 0$. As seen in Fig. 11, the (2,0) mode has a nonoscillatory real part during the inspiral, while the imaginary part is small but oscillatory. The (3,0) mode is similar, with the roles of the real and imaginary parts reversed. Therefore, we do not decompose $\tilde{h}^{\ell,0}(t)$ according to Eqs. (48) and (49). Instead, we model the nonoscillatory component directly, and we take a Hilbert transform of the oscillatory component to obtain an amplitude and phase, after tapering that component in the intervals $[t_{\min}, t_0]$ and $[t_f, t_{\max}]$.

The decomposition of the NR waveforms is summarized in Fig. 6. The NR waveforms begin at the top of the diagram and are processed going downwards. Each blue end point represents one of the slowly varying waveform data pieces that we fit as a function of parameters $\boldsymbol{\lambda}$ at each of the empirical time nodes. To evaluate the surrogate, the fits and empirical interpolants are evaluated for each of the blue end points, and the waveform is reconstructed by going upwards in the diagram and undoing each decomposition, eventually yielding $h^{\ell,m}(t)$.

D. Removing the dependence on ϕ_χ

As discussed in Sec. III A, we attempt to start all NR simulations so that at the reference time $t = t_0$ we have $\phi_\chi = 0$, where ϕ_χ is the azimuthal angle of the spin of the larger black hole, as shown in Fig. 3. However, in practice we obtain NR simulations with nonzero values of ϕ_χ at $t = t_0$. In this section we describe how we analytically approximate the dependence of the waveform on ϕ_χ. The surrogate model is then built assuming $\phi_\chi = 0$, so that when the surrogate model predicts waveforms with $\phi_\chi \neq 0$, the ϕ_χ dependence is described fully by this analytical approximation. For an orbit-averaged PN waveform of any order that is decomposed into waveform data pieces as described above, it turns out that one can show from the equations (e.g. as written in [67]) that none of the waveform data pieces depend on the parameter ϕ_χ except for the phase $\varphi_d(t)$. This phase has a particularly simple dependence,

$$\varphi_d(t; \boldsymbol{\lambda}, \phi_\chi) = \varphi_d(t; \boldsymbol{\lambda}, 0) + \phi_\chi, \tag{50}$$

where $\boldsymbol{\lambda}$ describes all of the parameters except ϕ_χ. So we will make the approximation that Eq. (50) applies not only to orbit-averaged PN waveforms, but also to NR waveforms lying within the 5D parameter space. In addition, we find empirically for NR waveforms that the phases of the Hilbert transforms of $A_{-}^{\ell,m}$ and $\varphi_{+}^{\ell,m}$ also obey Eq. (50), but with the opposite sign on the last term.

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

To verify how well this procedure removes the dependence on ϕ_χ, we performed additional SpEC simulations with parameters identical to cases SXS:BBH:0346 and SXS:BBH:0346 but with different values of ϕ_χ. We then analytically remove the ϕ_χ dependence from all these waveforms, as described above, thereby generating

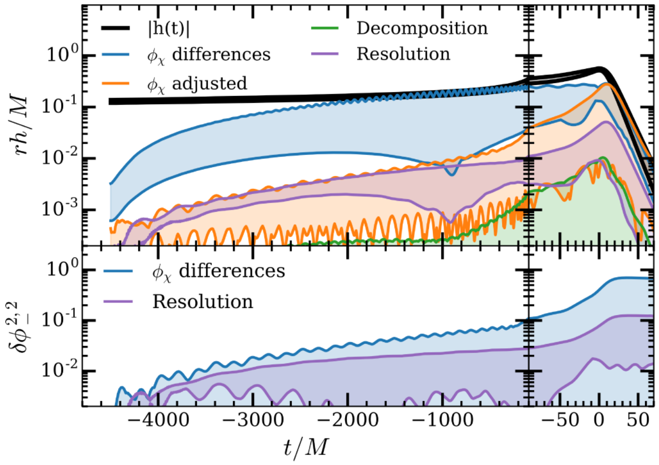

FIG. 12. Top: waveform differences $\delta h(t)$ investigating the removal of the ϕ_χ dependence on the waveform. Each colored band includes waveforms compared to SXS:BBH:0346 and SXS:BBH:0346 for several different values of ϕ_χ. Before making any adjustment, the errors (ϕ_χ differences) are large. After adjusting, the errors (ϕ_χ adjusted) are comparable to resolution errors during the inspiral but grow large at merger. The decomposition errors are negligible. Bottom: differences in $\varphi_-^{2,2}$. Our analytic approximation to remove the effect of ϕ_χ on the waveform does not affect $\varphi_-^{2,2}$, but here we see that the orbital phase at merger can vary by nearly a radian for different values of ϕ_χ, which is the most significant contribution to the ϕ_χ adjusted errors in the top figure.

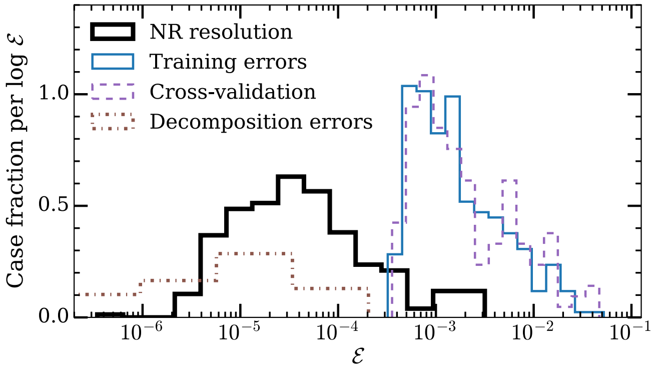

FIG. 13. Histograms of time domain waveform errors $\mathcal{E}$ relevant to the surrogate. Equal areas under the curves correspond to equal numbers of cases, and the curves are normalized such that the total area under each curve when integrated over $\log_{10}(\mathcal{E})$ is 1. Solid black: the resolution error comparing the highest and second highest resolution NR waveforms. Dotted brown: the error intrinsic to the surrogate's waveform decomposition. Filtering out nutation in the quaternions and neglecting the small but nonzero δq_z due to discrete time sampling leads to errors in the reconstructed waveforms. These errors are nearly zero for nonprecessing cases, and even for precessing cases they are smaller than the resolution errors. Thin solid blue: the errors when the full surrogate attempts to reproduce the set of waveforms from which it was built. Dashed purple: the errors when trial surrogates attempt to reproduce NR waveforms that were omitted during the surrogate construction.

$\phi_\chi = 0$ versions of these waveforms, which we compare with each other. The agreement (or lack thereof) of these $\phi_\chi = 0$ waveforms is a measure of the effectiveness of our analytical procedure for removing the ϕ_χ dependence. We find that while the dependence on ϕ_χ is removed well during the inspiral, $\varphi_-^{2,2}(t)$ varies by nearly a radian during the merger as we vary ϕ_χ, which leads to errors significantly larger than the SpEC resolution errors as shown in Fig. 12. Incidentally, we note that for two waveforms for which ϕ_χ originally differs by π, the corresponding $\phi_\chi = 0$ waveforms are nearly identical. Before removing ϕ_χ, the largest difference in the waveforms used in this test is $\mathcal{E} = 0.0285$, while after removing ϕ_χ, the largest difference is $\mathcal{E} = 0.00684$. While our ϕ_χ-removal procedure successfully accounts for most of the effect of ϕ_χ, the error associated with this procedure is larger than the median surrogate error (see Fig. 13 and Sec. VI) and indicates this approximation could be the dominant source of error in the surrogate model.

E. Handling undefined phases

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

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

V. BUILDING A SURROGATE MODEL FROM DECOMPOSED WAVEFORMS

We have decomposed each NR waveform into many functions $X(t;\boldsymbol{\lambda})$ that are smoothly varying as a function of parameters $\boldsymbol{\lambda}$. Here, X represents one of the many decomposed waveform data pieces such as φ_p or $A_+^{2,2}$. Note that while different waveform data pieces X will have different linear basis sizes, empirical time nodes, empirical interpolants, and parameter space fits, we will not always label the explicit X dependence of these quantities. For each X

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

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

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

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

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

A. Empirical interpolation

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

We require an orthonormal basis B spanning the space of solutions $\{X(t;\boldsymbol{\lambda}):\boldsymbol{\lambda}\in\mathcal{T}\}$. While the standard method is to use a reduced basis that was previously constructed when determining the greedy parameters G, in our case we used PN waveforms to find the greedy parameters and have not yet built a basis for NR solutions of X. Greedy and singular value decomposition (SVD) algorithms have been used within the gravitational wave surrogate modeling community [29,34–36] and will both provide an accurate basis provided any $X(t,\boldsymbol{\lambda})$ can be accurately approximated in the span of $\{X(t,\boldsymbol{\lambda}):\boldsymbol{\lambda}\in G^X\}$. A short discussion, including advantages and disadvantages of SVD and greedy algorithms in the context of surrogate waveform modeling, is given in Appendix B. Despite using a greedy sampling strategy to identify the set of greedy parameters, we use a SVD basis for the NR solutions, primarily for its ability to average out uncorrelated noise sources (see Appendix B).

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

$$X_S(t;\boldsymbol{\lambda}) = I_n[X_S](t) \tag{51}$$

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

B. Parametric fits

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

$$X_j(\boldsymbol{\lambda}) = X(T_j;\boldsymbol{\lambda}). \tag{52}$$

We build an approximate model for X_j denoted by X_{jS} by fitting it to the available data $\{X_j(\boldsymbol{\lambda}):\boldsymbol{\lambda}\in G^X\}$. We do so using a forward-stepwise least-squares fit [70] described in Appendix A, using products of univariate basis functions in q, $|\vec{\chi}_1|$, θ_χ and χ_2^z as the fit features. For each fit, the number of fit coefficients is determined through a cross validation study using 50 trials, each of which uses $N_v = 5$ randomly chosen validation points. The number of fit coefficients used is the one minimizing the sum in quadrature over the error in each trial, which is the maximum fit residual for the validation points.

C. Complete surrogate waveform model in inertial coordinates

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

$$X_S(t; \boldsymbol{\lambda}) = \sum_{j=1}^{n} X_{jS}(\boldsymbol{\lambda}) b^j(t). \tag{53}$$

We then obtain the inertial frame waveform $h_S^{\ell,m}(t)$ by combining the waveform data pieces and flowing upwards in Fig. 6. Explicitly,

$$Q(t) = T_\varphi^{-1}(\varphi_d(t; \boldsymbol{\lambda}) + \phi_\chi, \varphi_p(t; \boldsymbol{\lambda})) \tag{54}$$

$$\varphi_{\mathcal{I}}^{2,0}(t) = \varphi(H(\mathcal{I}\tilde{h}^{2,0}))(t; \boldsymbol{\lambda}) - \phi_\chi \tag{55}$$

$$\mathcal{I}\tilde{h}^{2,0}(t) = A(H(\mathcal{I}\tilde{h}^{2,0}))(t; \boldsymbol{\lambda}) \cos(\varphi_{\mathcal{I}}^{2,0}(t)) \tag{56}$$

$$\tilde{h}^{2,0}(t) = \mathcal{R}\tilde{h}^{2,0}(t; \boldsymbol{\lambda}) + i\mathcal{I}\tilde{h}^{2,0}(t) \tag{57}$$

$$\varphi_{\mathcal{R}}^{3,0}(t) = \varphi(H(\mathcal{R}\tilde{h}^{3,0}))(t; \boldsymbol{\lambda}) - \phi_\chi \tag{58}$$

$$\mathcal{R}\tilde{h}^{3,0}(t) = A(H(\mathcal{R}\tilde{h}^{3,0}))(t; \boldsymbol{\lambda}) \cos(\varphi_{\mathcal{R}}^{3,0}(t)) \tag{59}$$

$$\tilde{h}^{3,0}(t) = \mathcal{R}\tilde{h}^{3,0}(t) + i\mathcal{I}\tilde{h}^{3,0}(t; \boldsymbol{\lambda}) \tag{60}$$

$$\varphi^{\ell,\pm m}(t) = \varphi_+^{\ell,m}(t; \boldsymbol{\lambda}) \pm \varphi_-^{\ell,m}(t; \boldsymbol{\lambda}), \qquad m > 0 \tag{61}$$

$$A^{\ell,\pm m}(t) = A_+^{\ell,m}(t; \boldsymbol{\lambda}) \pm A_-^{\ell,m}(t; \boldsymbol{\lambda}), \qquad m > 0 \tag{62}$$

$$\tilde{h}^{\ell,\pm m}(t) = A^{\ell,\pm m}(t) \cos(\varphi^{\ell,\pm m}(t)), \qquad m > 0 \tag{63}$$

$$\{h^{\ell,m}(t)\} = T_Q(\{\tilde{h}^{\ell,m}(t)\}, Q(t)), \tag{64}$$

where we have included the dependence on $\boldsymbol{\lambda}$ explicitly for surrogate evaluations of waveform data pieces X_S. The full NRSur4d2s surrogate evaluation producing all $2 \leq \ell \leq 4$ modes for an array of times between $t_{\rm min}$ and $t_{\rm max}$ with spacing $\delta t = 0.1$ takes ~$1s$ on a single modern processor. Roughly half of this time is spent computing the transformation T_Q from the coprecessing frame to the inertial frame, Eq. (26).

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

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

VI. ASSESSING THE MODEL ERRORS

A. Time domain errors

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

$$\mathcal{E}[h_{\rm NR}, h_{\rm Sur}] = \frac{1}{2} \frac{\delta h}{\|h_{\rm NR}\|^2}, \tag{65}$$

where $h_{\rm NR}$ and $h_{\rm Sur}$ are the NR and surrogate waveforms, and δh is given by Eq. (19). This quantifies the surrogate error as a whole at one point in parameter space. For NR waveforms that were used to build the surrogate, we call Eq. (65) the *training error*. For NR waveforms that were not used to build the surrogate, but are used to test the accuracy of the surrogate model versus NR, we call Eq. (65) the *validation error*. Because we decompose each waveform into a set of slowly varying functions that are modeled independently (i.e., the waveform data pieces of Sec. IV), it is useful to consider the contribution to the surrogate error that arises from modeling a single waveform data piece. If X denotes the waveform data piece in question, then we compute this error contribution by decomposing the NR waveform $h_{\rm NR}$ into waveform data pieces, we replace the NR version of X with the surrogate model for X while leaving all waveform data pieces other than X untouched, and we recombine the waveform data pieces, thus producing a waveform we call $h_{\rm X}$. The error contribution from X is then $\mathcal{E}_X \equiv \mathcal{E}[h_{\rm NR}, h_{\rm X}]$. Values of $\mathcal{E}_X$ for various waveform data pieces X are listed in Table III. Note that if we decompose $h_{\rm NR}$ into waveform data pieces and then recompose the waveform data pieces, we do not recover $h_{\rm NR}$ exactly, but instead we get a different

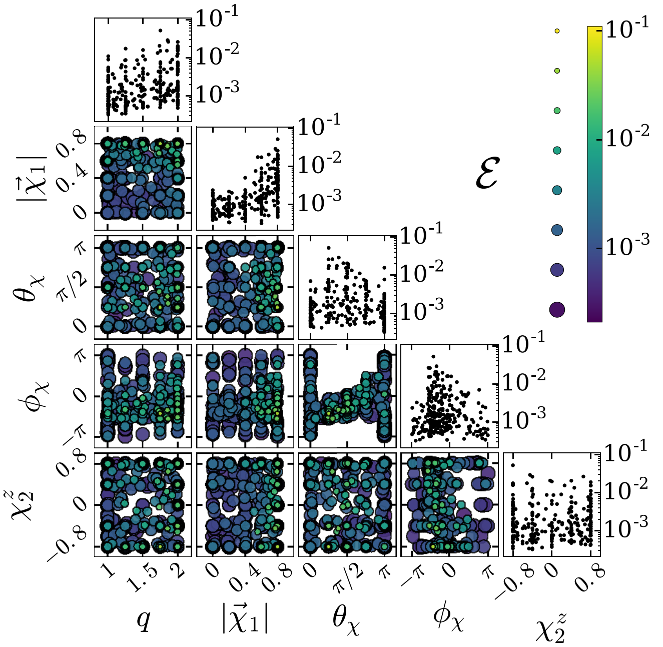

FIG. 14. Parameter dependence of the error $\mathcal{E}[h, h_S]$ when reproducing the set of NR waveforms with the surrogate. Diagonal: For each parameter plot, the black dots label the (parameter value, $\mathcal{E}[h, h_S]$) pairs. Off diagonal: For each pair of parameters, we show the 2D projection of parameters as in Fig. 4 while varying the color and size of the point based on the error $\mathcal{E}[h, h_S]$. Points are placed in order of increasing error, to ensure the small yellow points with large errors are visible. Larger spin magnitudes, especially for precessing spin configurations, correlate with larger errors.

waveform $h_\emptyset$ because there is error associated with the decomposition. This error, $\mathcal{E}_\emptyset \equiv \mathcal{E}[h_{\rm NR}, h_\emptyset]$, is also shown in Table III.

A first test is to verify that the NRSur4d2s surrogate can reproduce the set of NR waveforms from which it was built. The errors for those parameters are shown as the solid blue curve in Fig. 13. These errors are significantly larger than the NR resolution errors (cyan curve), which compare the highest and second highest NR resolutions. This indicates either that including additional NR waveforms when building the surrogate model would reduce the training error, or that the error is dominated by approximations made when building the model, such as the analytic treatment of ϕ_χ. The median training error is 0.00136, and in Sec. IV D we found that our approximation for the waveform's dependence on ϕ_χ resulted in errors up to 0.00684, indicating the model errors could be dominated by the error in this approximation. While the maximum training error is 0.05212, we only investigated the dependence on ϕ_χ for three cases and only for a few values of ϕ_χ. The parametric dependence of the training errors is illustrated in Fig. 14. Perhaps unsurprisingly, the largest errors occur at larger mass ratios and spin magnitudes, and for precessing spin directions.

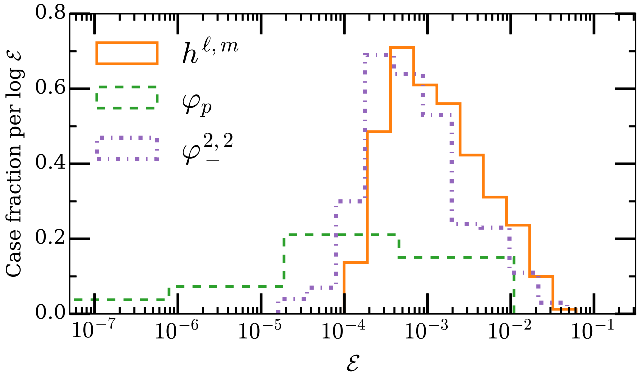

FIG. 15. Errors $\mathcal{E}_X$ showing the error contribution of a single surrogate component X.

To test the interpolation accuracy of the surrogate, we perform a cross-validation study. For each of ten trials, we randomly select $N_{\rm v} = 10$ waveforms which we call *validation waveforms*, and we build a trial surrogate using the remaining $N_{\rm t} = N - N_{\rm v}$ waveforms. The trial surrogate is evaluated at the $N_{\rm v}$ validation parameters, and the results are compared to the validation NR waveforms. These validation errors are shown as the purple dashed curve in Fig. 13. The validation errors are quite similar to the training errors, indicating we are not overfitting the data.

The maximum and median values of the training errors $\mathcal{E}_X$ are listed in Table III. The decomposition errors $\mathcal{E}_\emptyset$, also shown as the dotted brown curve in Fig. 13, are similar or smaller to the NR resolution errors and are therefore negligible. All component errors $\mathcal{E}_X$ include the decomposition errors by construction, and we see that $X = \tilde{h}^{\ell,m}$ leads to negligible errors except for the (2,2), (2,1) and (3,3) modes. The (2,2) mode is the dominant contribution to the error, and its error is dominated by the error in $\varphi_-^{2,2}$. The precession phase φ_p is the dominant precession error and is the next most significant contribution to the total error in h. Figure 15 shows histograms of the dominant sources of error, and Fig. 16 shows the time-dependent errors of these components for the case with the largest training error.

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

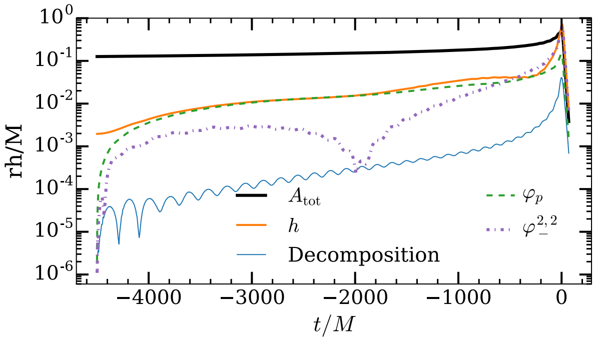

FIG. 16. Error contributions $\delta h(t)$ of those waveform data pieces X that have the largest error $\mathcal{E}[h, h_X]$ for a selected simulation: ID 79. To compute the error, the NR waveform is decomposed into the surrogate components, and component X is replaced with its surrogate evaluation. The waveform is then reconstructed, and $\delta h(t)$ is computed from Eq. (16). The solid black curve is given by Eq. (24). The dashed curve is the error in φ_p, which is the dominant error in modeling the precession, and the dominant error source during the inspiral. The dotted curve is the error in a quantity similar to twice the orbital phase and becomes the dominant error source during the merger and ringdown. The contribution from errors in the other waveform data pieces is smaller, as shown in Table III.

present in the NR waveforms used to build the surrogate are negligible compared to the NR resolution errors.

B. Frequency-domain comparisons

In this section we compute mismatches in the frequency domain between surrogate waveforms and NR waveforms. To ascertain the significance of these mismatches, we also compute mismatches between two NR waveforms with the same parameters but different resolutions. For comparison, we also compute mismatches between NR waveforms and the phenomenological inspiral-merger-ringdown waveform model IMRPhenomPv2 (which follows the procedure outlined in [19] with IMRPhenomD [20] as the aligned-spin model) and between the effective-one-body model SEOBNRv3 [23], both of which include the effects of precession.

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

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

We minimize the frequency domain mismatches over time and polarization angle shifts analytically as described in Appendix D and also minimize them over orbital phase shifts numerically. When we compare two waveforms, we choose one waveform as the *reference* waveform with fixed parameters and optimize over the parameters of the other waveform. When comparing two NR waveforms, the reference waveform is the one with the highest resolution; when comparing NR with some model waveform, the NR waveform is chosen as the reference.

The SEOBNRv3 and IMRPhenomPv2 waveforms are generated with the lalsimulation package [71]. Each SEOBNRv3 waveform is generated in the time domain; the spin directions are specified at the start of the waveform, which is determined by specifying a minimum frequency. We ensure the spin directions are consistent with those of the NR waveforms by varying the minimum frequency in order to obtain a waveform with a peak amplitude occurring $4500M$ after its initial time. The IMRPhenomPv2 waveforms are generated in the frequency domain, and the spin directions are specified at a reference frequency $f_{\rm ref}$ that can be freely chosen. For IMRPhenomPv2 it is not straightforward to determine $f_{\rm ref}$ such that the spin directions are specified at a time of $4500M$ before the peak amplitude. Therefore, we instead choose $f_{\rm ref}$ differently: we minimize the mismatches by varying $f_{\rm ref}$, with an initial guess of twice the initial orbital frequency of the NR waveform.

To transform the time domain waveforms into the frequency domain, we first taper them using Planck windows [69], rolling on for $t \in [t_0, t_0 + 1000M]$ and rolling off for $t \in [50M, 70M]$ where $t_0 = -4500M$ is the time at which the parameters are measured, and $t = 0$ is the time of peak waveform amplitude. We then pad them with zeros and compute the frequency domain waveforms via the fast Fourier transform (FFT). For the reference NR waveform, we obtain 30 random samples of the direction of gravitational wave propagation (θ, ϕ) from a distribution uniform in $\cos\theta$ and in ϕ, and we uniformly sample the polarization angle ψ between $[0, \pi]$ to obtain

$$h_\psi(t) = h_+(t)\cos(2\psi) + h_\times(t)\sin(2\psi). \tag{66}$$

For the nonreference waveform, we use the same parameters except we add an additional initial azimuthal rotation angle ϕ, a polarization angle ψ, and a time offset, and we optimize over these three new parameters to yield a minimum mismatch. Because the waveform models do not intrinsically depend on the total mass, we first use a flat noise curve to evaluate the overlap integrals; this provides a raw comparison between models. We evaluate Eq. (23) with $f_{\rm min}$ being twice the orbital frequency of the NR waveform at $t = -3500M$.

The mismatches using a flat noise curve are shown in the top panel of Fig. 17. We find that both the IMRPhenomPv2

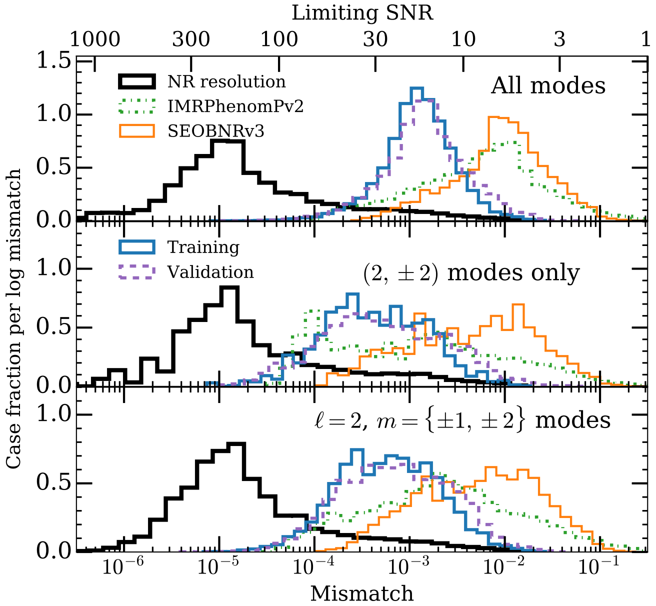

FIG. 17. Mismatches, computed using a flat noise curve, versus the highest resolution NR waveforms. Histograms are normalized to show the error fraction per log-mismatch, such that the area under each curve is the same. A sufficient but not necessary condition for a mismatch to have a negligible effect is that the signal-to-noise ratio (SNR) lies below the limiting SNR $\rho_* = 1/\sqrt{2\text{Mismatch}}$ given on the top axis [72]. Top: All modes available to each waveform model are included, and the NR waveforms use all $\ell \leq 5$ modes. Middle: All *coprecessing-frame* modes other than $(2, \pm 2)$ are set to zero in all waveforms. Bottom: All *coprecessing-frame* modes other than $(2, \pm 1)$ and $(2, \pm 2)$ are set to zero in all waveforms. These restricted mode studies are done to compare more directly with IMRPhenomPv2 and SEOBNRv3, which retain the coprecessing-frame modes of the middle and bottom panels respectively.

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

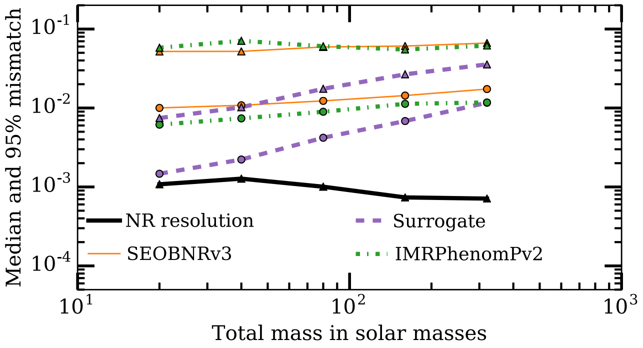

FIG. 18. Median (lower curves, circles) and 95th percentile (upper curves, triangles) mismatches for various total masses M using the advanced LIGO design sensitivity. The median NR resolution mismatches are all below 2×10^{-5}. The "surrogate" mismatches shown here are "validation" errors described in Sec. VI A.

Since a realistic noise curve will affect mismatches, we also compute mismatches for total masses M between 20 $M_\odot$ and 320 $M_\odot$ using the advanced LIGO design sensitivity [73]. In Fig. 18, the lower and upper curves for each waveform model denote the median mismatch and 95th percentile mismatch. We note that for $M < 114_\odot$, some NR and surrogate waveforms begin at $f_{\text{min}} > 10$ Hz and the noise-weighted inner products will not cover the whole advanced LIGO design sensitivity band. The surrogate model errors increase with total mass, indicating a larger amount of error in the merger phase and less error in the inspiral phase. Note that our largest systematic source of error, the approximate treatment of the waveform's dependence on the angle ϕ_χ, is much larger during the merger than during the inspiral, as discussed in Sec. IV D and plotted in Fig. 12. This error source arises from our attempt to model a 5D parameter space with a 4D surrogate model, so it will not be relevant for a full 7D surrogate model. Even with this error, our surrogate model performs better than the other waveform models up to 320 $M_\odot$ within the surrogate parameter space.

To determine if the discrepancy between the surrogate errors and NR resolution errors is due to an insufficient number of NR waveforms in the surrogate, we study how the errors depend on the number of waveforms used to build the surrogate. We construct trial surrogates using the first N_{train} NR waveforms for $N_{\text{train}} \in [30, 200]$; for validating the surrogate, we use the $N - 200$ waveforms that are not used to build any of these trial surrogates. By using the same $N - 200$ validation waveforms for all choices of N_{train}, we ensure that any changes in the error distribution resulting from changes in N_{train} are due to changes in the surrogate model and not in the set of validation waveforms.

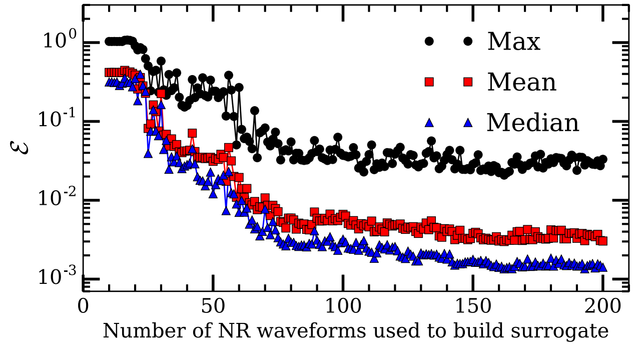

FIG. 19. Max, mean and medians of the distributions of $\mathcal{E}$ when building a surrogate using the first N waveforms and a validation set consisting the remaining $200 - N$ waveforms.

The validation errors, shown in Fig. 19, decrease quite slowly with additional waveforms when $N_{\rm train} > 100$, suggesting that the number of NR waveforms would have to increase dramatically to have a noticeable affect on the predictive ability of the surrogate.

C. Representing arbitrary spin directions

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

$$f\colon\ (q,\vec{\chi}_1,\vec{\chi}_2) \to \vec{x}_{\rm model} \tag{67}$$

from the 7D space of binaries with arbitrary spin directions to a lower-dimensional parameter subspace [17,18,74]. The use of a model with such a parameter space mapping in gravitational wave source parameter estimation leads to equivalence classes

$$\{(q,\vec{\chi}_1,\vec{\chi}_2) : f(q,\vec{\chi}_1,\vec{\chi}_2) = \vec{x}_{\rm model}\}, \tag{68}$$

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

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

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

$$\vec{\chi}_{\rm eff}^{\rm Drop} = \vec{\chi}_1. \tag{69}$$

A second possibility would be to use a similar parameter mapping as is used in IMRPhenomP [19] with an effective precessing spin $\chi_{\rm p}$ [74] and take

$$B_1 = \left(2 + \frac{3}{2q}\right)\left(\frac{q}{1+q}\right)^2, \tag{70}$$

$$B_2 = \left(2 + \frac{3q}{2}\right)\left(\frac{1}{1+q}\right)^2, \tag{71}$$

$$i^* = \underset{i=1,2}{\operatorname{argmax}} B_i \|\vec{\chi}_i^\perp\|, \tag{72}$$

$$\vec{\chi}_{\rm eff}^{\chi_{\rm p}} = \frac{B_{i^*}}{B_1}\vec{\chi}_{i^*}^\perp + \chi_1^z\hat{z}, \tag{73}$$

where $\vec{\chi}_i^\perp$ is the part of $\vec{\chi}_i$ orthogonal to the Newtonian orbital angular momentum, which is $(\chi_i^x, \chi_i^y, 0)$ at $t = t_0$. This mapping uses the in-plane spin components of whichever spin contributes the most to precession at leading PN order, scaled appropriately and placed on the heavier black hole. This mapping is particularly effective when the in-plane spins of the smaller BH are negligible, i.e., for high mass ratios, and for long duration GWs. However, it has also been shown to prove sufficient for binaries similar to GW150914 [3,25].

In our case, we have a couple precession cycles at most, and we might consider adding the effects of the in-plane components of the two spins. A further motivation to add the spins is that for nearly equal masses, the precession rates of the two spins will be nearly equal [51,75]. When adding the dimensionless spins, we can either do so directly

$$\vec{\chi}_{\rm eff}^{\rm Add} = \vec{\chi}_1 + \frac{1}{q^2}\vec{\chi}_2^\perp \tag{74}$$

or again using the leading order PN contribution to precession

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

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

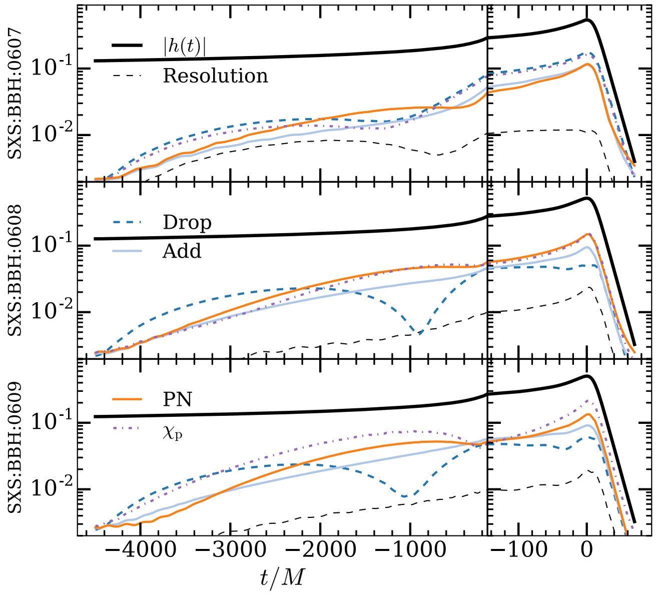

FIG. 20. Comparing surrogate evaluations to three NR waveforms (top, middle and bottom plots) with spins outside the 5D parameter subspace. For each case and each of the four parameter mappings, the surrogate model waveform error is shown. In all cases, the "PN" mapping performs well at the very start of the waveform and the "drop" mapping performs poorly, but there is no clear overall best mapping. Surrogate modeling errors contribute to the difficulty in assessing the quality of the mappings.

$$\vec{\chi}_{\mathrm{eff}}^{\mathrm{PN}} = \vec{\chi}_1 + \frac{B_2}{B_1}\vec{\chi}_2^{\perp}. \tag{75}$$

We do a brief investigation of the quality of these parameter space mappings using three additional SpEC simulations. The waveforms are aligned as described in Sec. III D, and their parameters at $t = t_0$ are measured and listed in Table V. For each case and each parameter space mapping, we compute the mapped parameters and compare the surrogate evaluation with the mapped parameters to the NR waveform. The time-dependent waveform errors are shown in Fig. 20 and $\mathcal{E}$ values as well as mismatches are given in Table VI. ∼0.01, which is larger than the median surrogate errors but well within the possible range of surrogate errors, so we cannot rule out that these errors are dominated by surrogate error. The "drop" parameter space mapping performs reasonably well since the cases investigated are far enough away from equal mass that the spin of the smaller black hole has a small effect on the waveform.

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

	$\mathcal{E}$			Median mismatch		
Map	0607	0608	0609	0607	0608	0609
Drop [Eq. (69)]	**0.016**	0.007	0.008	**0.0054**	0.0026	0.0031
Add [Eq. (74)]	0.007	0.009	**0.013**	0.0046	0.0051	**0.0076**
PN [Eq. (75)]	0.008	0.019	**0.021**	0.0041	0.0075	**0.0109**
χ_p [Eq. (73)]	0.014	0.018	**0.044**	0.0050	0.0074	**0.0161**

VII. BUILDING THE FREQUENCY DOMAIN SURROGATE

Evaluating the NRSur4d2s surrogate takes ∼$1s$ on a single modern processor. Evaluating all coprecessing modes takes ∼$0.21s$, evaluating the frame quaternions $q(t)$ takes ∼$0.38s$ and is dominated by evaluating Eq. (38) sequentially for all times, and rotating the modes into the inertial frame with the transformation T_Q takes ∼$0.41s$. Gravitational wave parameter estimation is typically done using Markov-chain Monte Carlo simulations [5] and can require $\mathcal{O}(10^8)$ waveform evaluations; this motivates us to build a faster surrogate model. We also wish the faster surrogate model to be in the frequency domain, where most parameter estimation is currently done. Accelerated frequency-domain surrogates have been built in 3D [29,36] using cubic tensor-spline interpolation of the waveform amplitudes and phases at some sparsely sampled frequency points.

To build the frequency-domain NRSur4d2s_FDROM surrogate, we first choose a uniformly spaced grid of $N = N_q \times \ldots \times N_{\chi_2^z}$ points in our 5D parameter space and evaluate the NRSur4d2s surrogate model at each point on the grid. We taper the waveforms with Planck windows [69], rolling on for $t \in [-4500M, -3500M]$ and rolling off for $t \in [50M, 70M]$. We then pad the waveform modes with zeros and perform a fast Fourier transform to obtain the frequency domain modes $\tilde{h}^{\ell,m}(f)$. We then downsample the frequency domain waveforms to a nonuniformly spaced set of frequencies, which are chosen to be the same for all waveforms and to be uniformly spaced in gravitational-wave phase for an equal-mass zero-spin binary. This significantly reduces the cost of evaluating the model, with a negligible loss in accuracy. For each mode $\tilde{h}^{\ell,m}(f)$, we build an empirical interpolant in frequency using all N waveforms, and we keep the first 100 basis vectors. At each empirical frequency node, we fit the real and imaginary parts of each mode across parameter space using a cubic tensor-product spline; we use "not-a-knot" boundary conditions that have a constant third derivative across the first and last knots [76]. Finding the spline coefficients involves solving a sparse linear system of size $(N_q + 2) \times \ldots \times (N_{\chi_2^z} + 2)$, for which we used Suitesparse [77,78] and/or SuperLU DIST [79,80]. The advantage of using a spline is that the evaluation cost is nearly independent of the grid size N, and requires only $4^{d=5}$ coefficients and basis functions to be evaluated.

TABLE VII. Grid sizes for tensor-spline interpolation in the frequency-domain surrogate. The size in each dimension is chosen such that surrogates for 1D slices in all dimensions have comparable interpolation errors.

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

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

To ensure that the empirical interpolants and parameter space splines are sufficiently accurate, we construct many frequency-domain surrogates for increasingly large parameter space grids. We monitor the differences between the frequency domain surrogate waveforms and the FFT of the tapered NRSur4d2s waveforms, and we demand that these differences decrease with increasing grid size. We use a different number of grid points in each parameter-space dimension, since the waveforms vary more in some dimensions than others. To determine the number of grid points to use, we construct frequency-domain surrogates for 1D slices of the parameter space, where the other parameters are fixed at a single intermediate value. We then arbitrarily choose a value of N_q, the number of grid points covering the dimension of mass ratio, and we determine the maximum error of the 1D surrogate in which only the mass ratio q is varied. Call this error E_q. Then we find the number of points $N_{|\chi_1|}$ for which the 1D surrogate for $|\chi_1|$ has an error of approximately E_q and similarly for the other parameters. The resulting grid sizes are listed in Table VII. In Fig. 21, we see that the errors converge as the grid size increases.

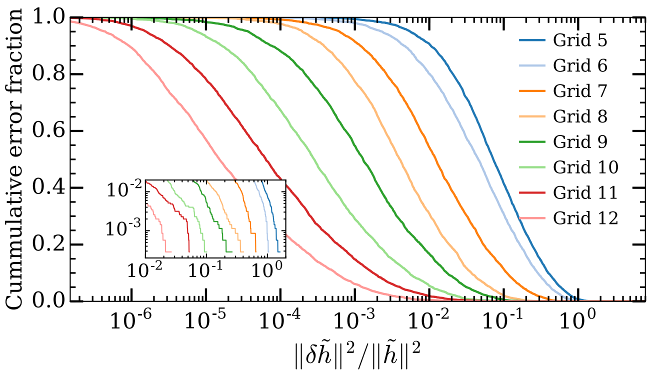

FIG. 21. Cumulative error distributions of the frequency domain NRSur4d2s_FDROM surrogate waveforms compared to the time domain NRSur4d2s surrogate waveforms transformed to the frequency domain, evaluated for randomly chosen uniformly distributed parameters. The curves indicate the fraction of errors at least as large as the indicated error. The NRSur4d2s_FDROM output converges to the FFT of the NRSur4d2s output as the grid size is increased.

VIII. DISCUSSION

We have built the first NR surrogate model of BBH waveforms that covers a multidimensional portion of the BBH parameter space. This extends the work in [34], where a one-dimensional (i.e. zero spin) NR surrogate served as a proof of principle that surrogate models of NR waveforms can be made highly accurate. The nonspinning surrogate model is inappropriate for use in GW parameter estimation, as neglecting all spin effects could lead to large parameter biases. Extending the parameter space to include both aligned spin components and one precessing component makes the new model presented here the first NR surrogate suitable for gravitational wave parameter estimation. While two of the in-plane spin components are still neglected by the NRSur4d2s surrogate model, IMRPhenomPv2 neglects similar information but obtains parameters for GW150914 that are compatible with those obtained using SEOBNRv3, which includes all spin components [25]. We note, however, that for edge-on systems otherwise similar to GW150914 IMRPhenomPv2 can obtain biased parameter estimates [33].

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

Use of the NRSur4d2s surrogate is also limited by the length (i.e. number of orbits) of the waveforms used to build it. GW151226 enters the sensitive LIGO band approximately 55 cycles before merger [2], while the NRSur4d2s surrogate produces waveforms with between 30 and 40 cycles before merger. Since these waveforms are tapered before building the faster NRSur4d2s_FDROM surrogate, the latter includes only 25 to 35 cycles before merger. There are a few ways to build an NR surrogate with longer waveforms, so that the surrogate is applicable to GW events of lower total mass. First, one could build a surrogate model using longer NR waveforms. A less computational expensive option would be to hybridize [81–84] the NR waveforms with PN or EOB waveforms before building a surrogate model. A final option would be to use a time domain surrogate which produces waveforms of moderate length as done here, to hybridize the surrogate output with PN or EOB waveforms before transforming them into the frequency domain, and finally to build a frequency domain surrogate for the hybrid waveforms.

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

Since we have not performed Cauchy characteristic extraction [85–88], but instead have extracted waveforms from the simulations at a series of finite radii and then extrapolated them to infinite radius [66], the (2,0) modes of the numerical waveforms in the coprecessing frame may not be accurate [89]. In particular, we do not see the expected gravitational wave memory in the real part of the (2,0) mode [90,91]. This should lead to negligible errors for most LIGO purposes, since the memory signal is low frequency and has very little contribution within the LIGO band. However, NRSur4d2s would not be suitable to detect a memory signal with a method requiring templates that include memory. A direct measurement of the memory signal using the method proposed in [92], however, could make use of waveforms from NRSur4d2s, as they have the $(2,\pm1)$ and $(3,\pm3)$ modes in the coprecessing frame necessary to determine the sign of the memory.

The errors in the NRSur4d2s surrogate are significantly larger than the resolution of the NR waveforms used in its construction. An incomplete treatment of the spin angle φ_χ (see Fig. 3) is one large source of error, and a complete 7D NR surrogate model would not suffer from this issue. Aligning the rotation of the waveforms (see Sec. III D) closer to merger might reduce the errors, since $\varphi_-^{2,2}$ at the empirical nodes would have less variation across parameter space. Since the parameters of the NR simulations were chosen such that $\vec{\chi}_2$ is aligned with the orbital angular momentum $4500M$ before merger, it would be nontrivial to build a surrogate model from these NR waveforms if the rotation alignment were performed at some other time. This is another issue which will be resolved by including all seven dimensions of parameter space.

Incorporating additional NR waveforms into the NRSur4d2s surrogate should also reduce the surrogate errors, although Fig. 19 indicates that with the current surrogate choices a very large number of additional NR waveforms would be needed for a significant reduction. Alternative methods of fitting empirical nodes could also help. The training and validation errors in Figs. 17 and 13 are nearly identical, while in [34] the validation errors were roughly a factor of 2 larger than the training errors. This suggests we may be under fitting the data and could use tighter parameter space fit tolerances.

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

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

(ii) *Decay of the parametric fitting error*. It is known that expanding (with orthogonal projection) a smooth function with polynomials results in an exponentially decaying approximation error. We believe the waveform data pieces evaluated at empirical nodes can be described by a smooth function plus (relatively small) noise. Thus, just as in the case of the basis projection error, the fitting error is expected to decay exponentially before the noise sources dominate the approximation. This can be seen in Fig. 22, where the exponential decay only lasts for approximately ten coefficients before noise sources cause the validation errors to flatten and then slowly rise.

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

NR waveforms of different resolutions should yield similar differences to comparing the NR waveforms themselves.

(iv) *Residual structure*. We could examine the parametric fit residuals and cross-validation residuals as a function of parameters. If the surrogate model captures the dominant features of NR waveforms then these residuals should appear random. From Fig. 14 we see that the largest errors occur at large values of $|\vec{\chi}_1|$ and for intermediate values of θ_χ, where precession has the largest effect. This indicates additional highly precessing NR simulations may help significantly in reducing the surrogate errors.

ACKNOWLEDGMENTS

We thank Michael Boyle, Alessandra Buonanno, Kipp Cannon, Maria Okounkova, Richard O'Shaughnessy, Christian Ott, Harald Pfeiffer, Michael Pürrer, and Saul Teukolsky for many useful discussions throughout this project. We also thank Andy Bohn, Nick Demos, Alyssa Garcia, Matt Giesler, Maria Okounkova, and Vijay Varma for helping to carry out the SpEC simulations used in this work. This work was supported in part by the Sherman Fairchild Foundation and NSF Grant No. PHY-1404569 at Caltech. J. B. gratefully acknowledges support from NSERC of Canada. Computations were performed on NSF/NCSA Blue Waters under allocation PRAC ACI-1440083; on the NSF XSEDE network under allocation TG-PHY990007; on the Zwicky cluster at Caltech, which is supported by the Sherman Fairchild Foundation and by NSF Grant No. PHY-0960291; and on the ORCA cluster at California State University at Fullerton, which is supported by NSF Grant No. PHY-1429873, the Research Corporation for Science Advancement, and California State University at Fullerton.

APPENDIX A: FORWARD-STEPWISE GREEDY FIT ALGORITHM

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

$$X_{mS}(\boldsymbol{\lambda}) = \sum_{i=1}^{M} c_i B^i(\boldsymbol{\lambda}) \tag{A1}$$

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

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

$$B^{\vec{\alpha}}(\boldsymbol{\lambda}) = \prod_{l=1}^{d} B_l^{\alpha^l}(\lambda^l). \tag{A2}$$

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

(i) $B_1^k(q) = T_k(2q - 3)$

(ii) $B_2^k(|\chi_1|) = (\frac{|\chi_1|}{0.8})^k$

(iii) $B_3^k(\theta_\chi) = \cos(k\theta_\chi)$

(iv) $B_4^k(\chi_2^z) = T_k(\frac{\chi_2^z}{0.8})$,

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

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

$$B_3^k(\theta_\chi) = \sin((k+1)\theta_\chi), \tag{A3}$$

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

The above choices yield 1008 possible basis functions (1512 for φ_p), which is more than $N \leq 300$, so we will use only a subset of the possible basis functions. We determine elements $B^i \in \{B^{\vec{\alpha}}\}$ of this subset in a greedy manner with a forward-stepwise least-squares fit [70]. We proceed by iteratively updating two quantities: r_j^n, which is the jth fit residual at the nth iteration, and $b_j^{\vec{\alpha},n}$, which is the orthogonal component of the basis function $B^{\vec{\alpha}}$ at the nth iteration evaluated at parameters $\boldsymbol{\lambda}_j$. For the zeroth iteration we begin with

$$r_j^0 = X_m(\boldsymbol{\lambda}_j) \tag{A4}$$

$$b_j^{\vec{\alpha},0} = B^{\vec{\alpha}}(\boldsymbol{\lambda}_j). \tag{A5}$$

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

$$d_n^{\vec{\alpha}} = \sum_j r_j^n b_j^{\vec{\alpha},n}. \tag{A6}$$

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

$$\vec{\alpha}_n^* = \underset{\vec{\alpha}}{\operatorname{argmax}} |d_n^{\vec{\alpha}}| \tag{A7}$$

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

$$r_j^{n+1} = r_j^n - d_n^{\vec{\alpha}_n^*} b_j^{\vec{\alpha}_n^*,n} \tag{A8}$$

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

$$b_j^{\vec{\alpha},n+1} = b_j^{\vec{\alpha},n} - e^{\vec{\alpha},n} b_j^{\vec{\alpha}_n^*,n} \tag{A9}$$

$$e^{\vec{\alpha},n} = \sum_j b_j^{\vec{\alpha},n} b_j^{\vec{\alpha}_n^*,n}. \tag{A10}$$

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

This procedure does not indicate which value of m (the number of fit coefficients) to use. Using N fit coefficients would be overfitting the data and setting individual fit tolerances by hand for each empirical node of each data component would be time consuming and error prone. So instead, we repeat the above procedure for different values of m, we perform cross-validation studies on the resulting fits, we find the value of m that leads to the smallest validation errors (call this value m^*), and we choose $m = m^*$. For each trial $k = 1, \ldots, K = 50$ of this cross-validation procedure, we randomly divide the N data points into $N_v = 5$ validation points and $N_t = N - N_v$ training points. Using only the training data, we perform the above greedy forward-stepwise fitting procedure. For values of $m \in [0, N_t]$, we obtain a least-squares fit with m coefficients using the training data and evaluate the fit residuals $r_j^{m,k}$ for the validation data. We choose

$$m^* = \underset{m}{\operatorname{argmin}} \sum_{k=1}^{K} \max_{j=1}^{N_v} (r_j^{m,k})^2. \tag{A11}$$

We use the maximum over j because we seek to minimize the largest fit residuals, and we sum in quadrature over k rather than maximize to account for cases where data points with large errors or corner cases are selected as validation points, which can lead to large fit residuals. The dependence of the residuals on m for one case is shown in Fig. 22.

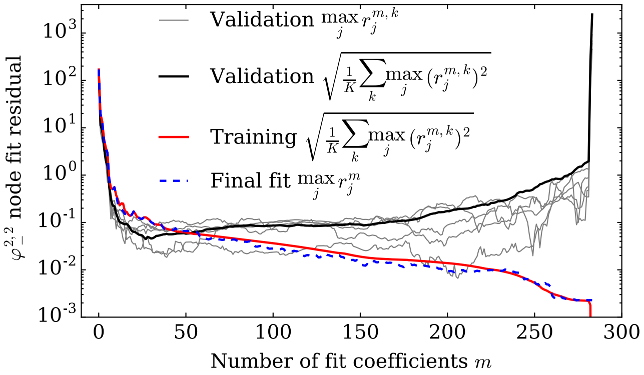

FIG. 22. Fit residuals for the second empirical node of $\varphi_-^{2,2}$ at $t = -806.5M$. Blue dashed: The maximum fit residual using all data. Thin grey lines: Maximum validation residual for individual trials. Thick black line: The root mean square (rms) of the validation residuals for $K = 50$ trials. It takes its minimum value at $m = 30$, which determines the number of fit coefficients to use for this node in the model. Red: The rms of the training residuals for $K = 50$ trials.

APPENDIX B: COMPARING REDUCED BASIS CONSTRUCTIONS

We compare two commonly used methods to generate a reduced basis in gravitational waveform reduced-order modeling. The first uses a singular value decomposition (SVD) of a data set whose output consists of a set of basis vectors ranked by their “singular values”, which are eigenvalues when the input data is square. The SVD reduced basis follows by truncating the output basis beyond a selected singular value. The resulting basis is accurate up to that singular value as measured in a root-mean-square norm. The second method uses a greedy algorithm, which is iterative and nested, to expose the most relevant elements of the input (or *training*) data set [93,94]. The greedy algorithm selects the element with the largest current projection error (as measured by a specified norm), orthonormalizes the selected element with respect to the current basis, and adds this orthonormalized element to the set of basis vectors. In practice, one uses an iterated, modified Gram-Schmidt process [95] for orthonormalization, which is robust to the accumulation of numerical round-off effects from subtraction until very large basis sizes. The algorithm ends when the largest projection error is below a specified tolerance; it also ends if a previously selected training data element is selected again, which, if it were allowed to occur, would introduce a linearly dependent element to the basis. The output includes a (greedy) reduced basis and a set of parameters or labels that indicate the most relevant elements of the training data from which the basis is built.

Both SVD and greedy methods output a reduced basis that accurately represents the training data to the requested singular value or tolerance. The output of the SVD algorithm depends only on the training data. The greedy algorithm, on the other hand, begins by choosing one of the training data elements as the first basis vector, so its output depends also on that choice. How that choice is made is often arbitrary and may depend on the application. For example, one may seed the greedy algorithm with an arbitrary element from the training set or choose the element that has the largest absolute value or norm. However, it has been shown that the choice of seed is largely irrelevant as the greedy algorithm seeks to minimize the maximum projection error across the entire training set, no matter what the seed. The resulting variations in the size of the greedy reduced basis due to arbitrary seed choices are marginal and typically span a few percent about the mean size [55,56,96].

Practical implementations of the SVD algorithm can be found rather easily because of its broad use across many disciplines. Therefore, building an SVD reduced basis for a training set of waveforms is as straightforward as calling the appropriate programmed function. However, if the training data contains N waveforms with L time or frequency samples then the SVD algorithm is $\mathcal{O}(N^2L)$, which can be intensive in both time and physical memory. For this reason, the authors in [97] divide the full training space into narrow strips in one direction of the parameter space. Dividing the training space into smaller subsets results in a direct product of reduced bases, one basis for each subset. Unfortunately, the total number of the basis elements tends to be larger than if one had performed a SVD on the full training data (if it can be done). Consequently, the reduction of the data is not maximized.

One often has considerable flexibility in designing a greedy algorithm for a specific application. If the training set remains fixed throughout the course of the greedy algorithm (see [59] for an example where this is not the case) then each iteration step can be performed in constant time so that the totality scales as $\mathcal{O}(nN)$ if n is the number of reduced basis elements needed to reach the specified tolerance. Typically, $n \ll N$ so that greedy algorithms tend to terminate more quickly than an application of SVD on the same training data, though there is some additional influence from implementation details. The greedy algorithm can be parallelized to break up the computation of expensive integrals across different processes [98]. In addition, the size and memory requirements of a very large training set pose little problem for greedy algorithms. The training space can be divided into subsets so that a reduced basis is built for each with a tolerance up to numerical round-off as measured in the L_∞ norm (to have pointwise accuracy for the data). Then, one may apply a second greedy algorithm on the full training data by using instead the basis data on each subset to represent the original data of each subset. In this way, one can generate a reduced basis that spans all the subsets and maximizes the reduction of the full training set [99]. Combining this two-step greedy algorithm with the parallelization of the projection integrals discussed above provides a viable and practical strategy for building a reduced basis for training sets of virtually any size. Another strategy is to randomly repopulate the training set at each iteration of the greedy algorithm [59,100]. This approach requires that the training data can be generated at will for any parameter values but also avoids storing prohibitively large amounts of data at any step in the greedy algorithm.

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

Let us next investigate a toy problem to facilitate a comparison of the outputs of a basic greedy algorithm and SVD. We consider a function

$$X(t;\lambda) = \sin(\lambda t) + 10^{-5}\sin(10\lambda t) + 10^{-10}\xi(t), \qquad \text{(B1)}$$

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

Our training set will consist of $N = 1000$ uniformly spaced values of λ. Figure 23 shows training data for the smallest and largest parameter values considered here. We sample the function in (B1) at 10,000 uniformly spaced times.

We construct three reduced bases. The first is built from an SVD on the training data. The second uses a greedy algorithm to generate a reduced basis and a corresponding set of parameters; here we use the L_2 norm to measure the difference between each training set element and its projection onto the basis. The third is built in the same

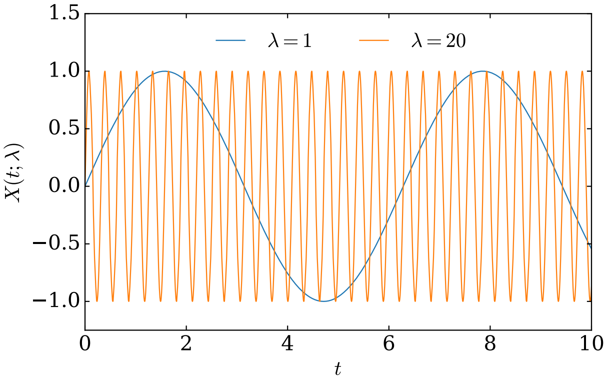

FIG. 23. Plots of $X(t;\lambda)$ for our toy problem evaluated at the smallest and largest values of λ in the training set.

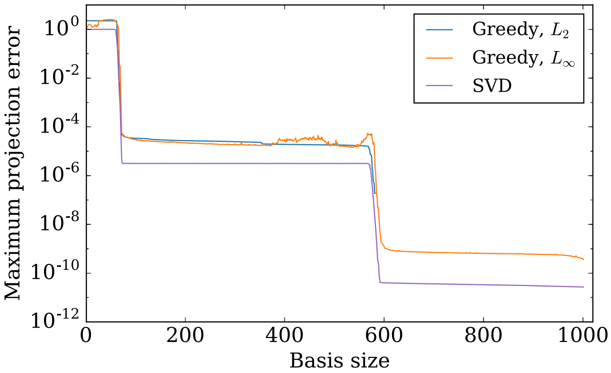

FIG. 24. Maximum projection errors of all three reduced bases (see text for a description) versus the size of the basis.

greedy manner as the second but uses the L_∞ norm to measure the projection error. Recall that the L_2 error constitutes a kind of average as it involves an integration in time whereas the L_∞ error measures the largest, pointwise, absolute difference and is thus more stringent. Figure 24 shows the maximum projection errors, as measured with their respective norms, associated with these three methods as a function of the size of the basis. The absolute tolerance on the greedy algorithm bases is 10^{-14} while the smallest singular value kept is 10^{-14} relative to the largest. We observe three plateaus for each of the cases, which can be attributed to each algorithm trying to resolve the features at the $\mathcal{O}(1)$, $\mathcal{O}(10^{-5})$, and $\mathcal{O}(10^{-10})$ scales in the data; see (B1). In fact, none of the algorithms are able to completely resolve the very low-amplitude stochastic features until the training set has been exhausted and all data has been used to build the reduced bases. Notice that the error curve is somewhat noisy for the L_∞ case while the other two are smooth. Also, the maximum projection error for the L_2 case ends at about 10^{-7} due to a parameter being selected a second time.

Figure 25 shows the projection errors (as measured in the L_2 norm) onto each of the three reduced bases for test data generated by randomly selecting 1000 values of λ in the training interval [1, 20]. The errors for "Greedy, L_∞" and "SVD" lie nearly on top of each other while those for "Greedy, L_2" are relatively large because the effective greedy algorithm tolerance for this basis is only 10^{-7} as discussed above. In all cases, the small-amplitude stochastic noise in the data prevents the projection errors of the test data from being less than a few times 10^{-10}; see (B1).

Finally, the SVD method is able to produce a reduced basis with elements that smooth many uncorrelated features manifest in the training data. Such smoothing is useful for surrogate model building because the resulting basis elements tend to exhibit smoother variation in time or frequency; this translates into smoother variations across parameters, thereby yielding more accurate fits for the parametric variation at the empirical interpolation nodes.

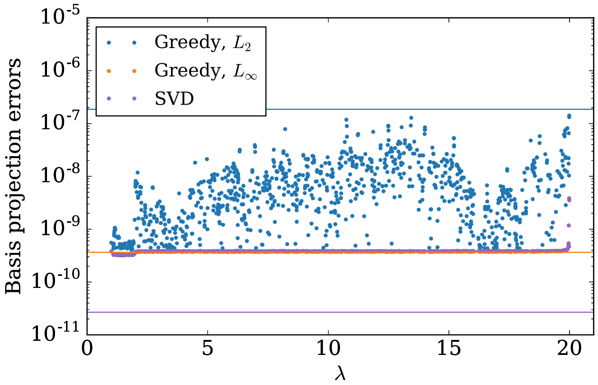

FIG. 25. Projection errors, measured in the L_2 norm for the three reduced bases described in the text, computed for test data generated from 1000 randomly selected parameters λ in [1, 20]. The corresponding colored lines indicate the smallest projection errors on the training sets shown in Fig. 24. The errors for "Greedy, L_∞" and "SVD" lie nearly on top of each other. However, the maximum projection error implied by SVD (purple line) underestimates the true errors (dots) by an order of magnitude.

The reduced bases produced by greedy methods tend to not to share this smoothing ability of the SVD method.

To demonstrate SVD's smoothing abilities, we replace the function in (B1) with a smooth oscillating term plus a stochastic term with amplitude of 10% of the first so that the noise is visible to the naked eye,

$$X(t;\lambda) = \sin(\lambda t) + 0.1\xi(t). \tag{B2}$$

We build three reduced bases on the corresponding training sets (with the same t and λ intervals and samples) using the same methods as before. Figure 26 shows the tenth basis element as a function of t for each of the three reduced basis building strategies. The two bases built from a greedy method exhibit the noise found in the training data. However, the SVD basis element in the bottom panel reveals a smooth function with very low amplitude noise, much lower than appears in the training data amplitudes.

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

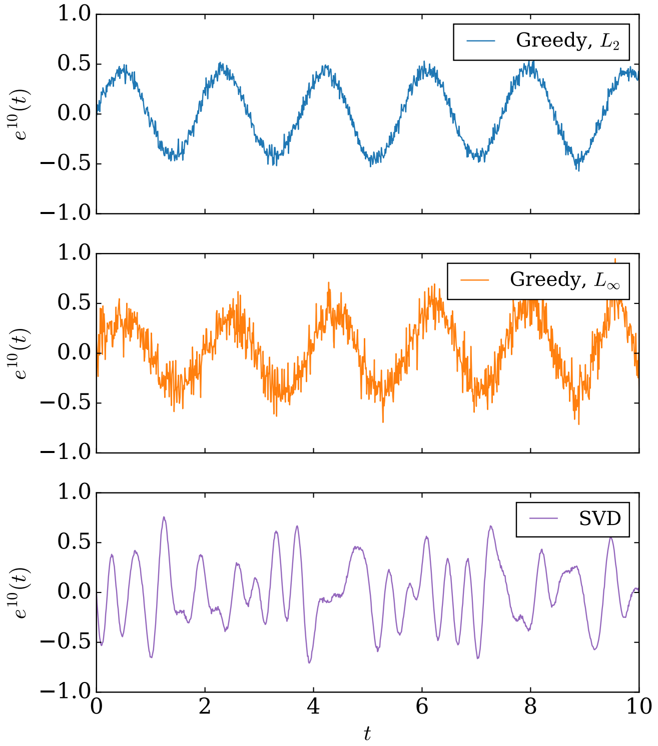

FIG. 26. The tenth basis element as a function of t from the three reduced bases elements described in the text. The training data used is given by the parametrized function in (B2) and exhibits relatively large amplitude fluctuations. Whereas the top two plots show significant noise in the basis element, the SVD method smooths away, almost completely, the uncorrelated stochastic features to generate a basis element that is smooth in t.

benefits of both the greedy and SVD methods in building NRSur4d2s.

APPENDIX C: MOTIVATING THE USE OF $\mathcal{E}$

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

$$1 - \mathcal{O} = 1 - \frac{\langle h_1, h_2 \rangle}{\sqrt{\langle h_1, h_1 \rangle \langle h_2, h_2 \rangle}}, \tag{C1}$$

where $\langle \cdot, \cdot \rangle$ is often chosen to be the frequency domain noise-weighted inner product [64]

$$\langle a, b \rangle_f = 4\mathrm{Re} \int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df. \tag{C2}$$

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

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

$$\langle a, b \rangle_t = \mathrm{Re} \int_{t_{\min}}^{t_{\max}} a(t) b^*(t) dt \tag{C3}$$

to obtain the same overlap error. While a completely flat power spectral density is unphysical, the design sensitivity of aLIGO [101] varies only by a factor of ∼2 between 50 Hz and 1000 Hz. Putting rigorous limits on weighted frequency domain errors based on unweighted time domain errors is not straightforward [72,102], but the time domain errors are computationally cheap to compute, useful for quantifying time domain waveform models, and (like NR waveforms and our surrogate model NRSur4d2s) independent of the total binary mass M.

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

$$\int_{S^2} a(\theta, \phi) b^*(\theta, \phi) d\Omega = \sum_{\ell, m} a^{\ell, m} b^{\ell, m*} \tag{C4}$$

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

$$\delta h^2 = \frac{1}{T} \sum_{\ell, m} \|\delta h^{\ell, m}\|_t^2 \tag{C5}$$

$$= \frac{1}{T} \int_{S^2} \|h_1(t, \theta, \phi; \boldsymbol{\lambda}_1) - h_2(t, \theta, \phi; \boldsymbol{\lambda}_2)\|_t^2 d\Omega \tag{C6}$$

$$= \frac{1}{T} \int_{S^2} (\|h_1\|_t^2 + \|h_2\|_t^2 - 2\langle h_1 h_2 \rangle_t) d\Omega, \tag{C7}$$

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

$$\frac{\delta h^2}{\sum_{\ell, m} \|h_1^{\ell, m}\|_t^2} = \frac{2 \int_{S^2} w(\theta, \phi)(1 - \mathcal{O}(\theta, \phi)) d\Omega}{\int_{S^2} w(\theta, \phi) d\Omega}, \tag{C8}$$

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

$$\mathcal{E} \equiv \frac{1}{2} \frac{\delta h^2}{\|h_1\|^2} \tag{C9}$$

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

APPENDIX D: MISMATCHES OPTIMIZED OVER TIME AND POLARIZATION SHIFTS

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

$$\mathcal{O} = \frac{\langle h_+, h^m_+\rangle + \langle h_\times, h^m_\times\rangle}{\sqrt{(\langle h_+, h_+\rangle + \langle h_\times, h_\times\rangle)(\langle h^m_+, h^m_+\rangle + \langle h^m_\times, h^m_\times\rangle)}}$$

with a real inner product given by

$$\langle a, b\rangle = \mathrm{Re}[\langle a, b\rangle_C] \tag{D1}$$

$$\langle a, b\rangle_C = \int \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(|f|)} df. \tag{D2}$$

As in Eq. (23), a tilde denotes a frequency domain signal, which is computed by using an FFT after tapering the ends of the time domain signal. In this case, the complex inner product $\langle\cdot,\cdot\rangle_C$ is integrated over the negative and positive frequency intervals $[-f_{\mathrm{max}}, -f_{\mathrm{min}}]$ and $[f_{\mathrm{min}}, f_{\mathrm{max}}]$ for some positive f_{min} and f_{max}. Note that for any two real functions $a(t)$ and $b(t)$, we have

$$\tilde{a}(-f)\tilde{b}^*(-f) = (\tilde{a}(f)\tilde{b}^*(f))^* \tag{D3}$$

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

Defining complex gravitational wave signals

$$h(t) = h_+(t) - ih_\times(t) \tag{D4}$$

$$h^m(t) = h^m_+(t) - ih^m_\times(t), \tag{D5}$$

we can compute a complex overlap

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

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

$$\mathcal{O} = \mathrm{Re}[\mathcal{O}_C]. \tag{D6}$$

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

$$h^m(t) \to h^m_T(t) = h^m(t + \delta t)e^{2i\psi}, \tag{D7}$$

$$\tilde{h}^m(f) \to \tilde{h}^m_T(f) = \tilde{h}^m(f)e^{2i\psi}e^{2\pi i\delta t}, \tag{D8}$$

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

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

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

$$\mathrm{mismatch} = 1 - \max_{\psi,\delta t}\mathcal{O}(\psi, \delta t) \tag{D9}$$

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

APPENDIX E: POST-NEWTONIAN SURROGATE WAVEFORM DECOMPOSITION

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

PN waveforms do not contain a merger phase, so we cannot use the peak amplitude to align the waveforms in time. We instead choose $t = 0$ to correspond to an orbital angular frequency of 0.09. This frequency is computed from the waveform [68]. We choose $t_{\mathrm{min}} = -5000M$, $t_0 = -4500M$, $t_f = -100M$, and $t_{\mathrm{max}} = 0$. The PN waveforms used to build the PN surrogate then have domain

$t \in [-5000M, 0]$, and the PN surrogate waveforms have domain $t \in [-4500M, -100M]$. The parameters of the PN waveforms are given at $t = t_0$. The rotation alignment at $t = t_0$ is the same as for the NR waveforms, described in Sec. III D.

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

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

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

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

[1] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 061102 (2016).

[2] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 241103 (2016).

[3] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 241102 (2016).

[4] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 221101 (2016).

[5] J. Veitch *et al.*, Phys. Rev. D **91**, 042003 (2015).

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

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

[8] http://www.black-holes.org/SpEC.html.

[9] Einstein Toolkit home page: http://einsteintoolkit.org.

[10] S. Husa, J. A. González, M. Hannam, B. Brügmann, and U. Sperhake, Classical Quantum Gravity **25**, 105006 (2008).

[11] B. Brügmann, J. A. González, M. Hannam, S. Husa, U. Sperhake, and W. Tichy, Phys. Rev. D **77**, 024027 (2008).

[12] F. Herrmann, I. Hinder, D. Shoemaker, and P. Laguna, Classical Quantum Gravity **24**, S33 (2007).

[13] M. A. Scheel, M. Giesler, D. A. Hemberger, G. Lovelace, K. Kuper, M. Boyle, B. Szilágyi, and L. E. Kidder, Classical Quantum Gravity **32**, 105009 (2015).

[14] I. Ruchlin, J. Healy, C. O. Lousto, and Y. Zlochower, Phys. Rev. D **95**, 024033 (2017).

[15] C. O. Lousto and Y. Zlochower, Phys. Rev. Lett. **106**, 041101 (2011).

[16] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, arXiv:1508.07250.

[17] P. Ajith, M. Hannam, S. Husa, Y. Chen, B. Bruegmann, N. Dorband, D. Mueller, F. Ohme, D. Pollney, C. Reisswig, L. Santamaria, and J. Seiler, Phys. Rev. Lett. **106**, 241101 (2011).

[18] L. Santamaría, F. Ohme, P. Ajith, B. Brügmann, N. Dorband, M. Hannam, S. Husa, P. Mösta, D. Pollney, C. Reisswig, E. L. Robinson, J. Seiler, and B. Krishnan, Phys. Rev. D **82**, 064016 (2010).

[19] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Phys. Rev. Lett. **113**, 151101 (2014).

[20] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Prrer, X. Jimnez Forteza, and A. Boh, Phys. Rev. D **93**, 044007 (2016).

[21] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Phys. Rev. D **93**, 044006 (2016).

[22] A. Taracchini, A. Buonanno, Y. Pan, T. Hinderer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, A. H. Mroue, H. P. Pfeiffer, M. A. Scheel, B. Szilágyi, N. W. Taylor, and A. Zenginoglu, Phys. Rev. D **89**, 061502 (2014).

[23] Y. Pan, A. Buonanno, A. Taracchini, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **89**, 084006 (2014).

[24] A. Boh *et al.*, Phys. Rev. D **95**, 044028 (2017).

[25] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. X **6**, 041014 (2016).

[26] J. Aasi *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **32**, 115012 (2015).

[27] T. Damour, P. Jaranowski, and G. Schaefer, Phys. Rev. D **78**, 024009 (2008).

[28] T. Damour and A. Nagar, Phys. Rev. D **79**, 081503 (2009).

[29] M. Pürrer, Phys. Rev. D **93**, 064041 (2016).

[30] K. Jani, J. Healy, J. A. Clark, L. London, P. Laguna, and D. Shoemaker, Classical Quantum Gravity **33**, 204001 (2016).

[31] P. Kumar, K. Barkett, S. Bhagwat, N. Afshari, D. A. Brown, G. Lovelace, M. A. Scheel, and B. Szilgyi, Phys. Rev. D **92**, 102001 (2015).

[32] S. Babak, A. Taracchini, and A. Buonanno, Phys. Rev. D **95**, 024010 (2017).

[33] B. P. Abbott *et al.* (Virgo, LIGO Scientific Collaboration), arXiv:1611.07531.

[34] J. Blackman, S. E. Field, C. R. Galley, B. Szilágyi, M. A. Scheel, M. Tiglio, and D. A. Hemberger, Phys. Rev. Lett. **115**, 121102 (2015).

[35] S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio, Phys. Rev. X **4**, 031006 (2014).

[36] M. Pürrer, Classical Quantum Gravity **31**, 195010 (2014).

[37] H. P. Pfeiffer, L. E. Kidder, M. A. Scheel, and S. A. Teukolsky, Comput. Phys. Commun. **152**, 253 (2003).

[38] G. Lovelace, R. Owen, H. P. Pfeiffer, and T. Chu, Phys. Rev. D **78**, 084017 (2008).

[39] L. Lindblom, M. A. Scheel, L. E. Kidder, R. Owen, and O. Rinne, Classical Quantum Gravity **23**, S447 (2006).

[40] B. Szilágyi, L. Lindblom, and M. A. Scheel, Phys. Rev. D **80**, 124010 (2009).

[41] M. A. Scheel, M. Boyle, T. Chu, L. E. Kidder, K. D. Matthews, and H. P. Pfeiffer, Phys. Rev. D **79**, 024003 (2009).

[42] B. Szilágyi, Int. J. Mod. Phys. D **23**, 1430014 (2014).

[43] P. Schmidt, M. Hannam, and S. Husa, Phys. Rev. D **86**, 104063 (2012).

[44] L. Pekowsky, R. O'Shaughnessy, J. Healy, and D. Shoemaker, Phys. Rev. D **88**, 024040 (2013).

[45] P. Schmidt, M. Hannam, S. Husa, and P. Ajith, Phys. Rev. D **84**, 024046 (2011).

[46] R. O'Shaughnessy, B. Vaishnav, J. Healy, Z. Meeks, and D. Shoemaker, Phys. Rev. D **84**, 124002 (2011).

[47] M. Boyle, R. Owen, and H. P. Pfeiffer, Phys. Rev. D **84**, 124011 (2011).

[48] K. Cannon, J. Emberson, C. Hanna, D. Keppel, and H. Pfeiffer, Phys. Rev. D **87**, 044008 (2013).

[49] K. Cannon, C. Hanna, and D. Keppel, Phys. Rev. D **85**, 081504 (2012).

[50] J. Kaye, Ph. D. thesis, Brown University, 2012.

[51] T. A. Apostolatos, C. Cutler, G. J. Sussman, and K. S. Thorne, Phys. Rev. D **49**, 6274 (1994).

[52] L. E. Kidder, Phys. Rev. D **52**, 821 (1995).

[53] D. Christodoulou and R. Ruffini, Phys. Rev. D **4**, 3552 (1971).

[54] B. Aylott, J. G. Baker, W. D. Boggs, M. Boyle, P. R. Brady *et al.*, Classical Quantum Gravity **26**, 165008 (2009).

[55] S. E. Field, C. R. Galley, F. Herrmann, J. S. Hesthaven, E. Ochsner, and M. Tiglio, Phys. Rev. Lett. **106**, 221102 (2011).

[56] S. Caudill, S. E. Field, C. R. Galley, F. Herrmann, and M. Tiglio, Classical Quantum Gravity **29**, 095016 (2012).

[57] M. Boyle, Phys. Rev. D **87**, 104006 (2013).

[58] Gwframes, https://github.com/moble/GWFrames.

[59] J. Blackman, B. Szilágyi, C. R. Galley, and M. Tiglio, Phys. Rev. Lett. **113**, 021101 (2014).

[60] F. Herrmann, S. E. Field, C. R. Galley, E. Ochsner, and M. Tiglio, Phys. Rev. D **86**, 084046 (2012).

[61] D. A. Hemberger, M. A. Scheel, J. Blackman, S. E. Field, C. R. Galley, P. Schmidt, R. Smith, S. Ossokine *et al.* (to be published).

[62] M. Boyle, L. E. Kidder, S. Ossokine, and H. P. Pfeiffer, arXiv:1409.4431.

[63] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics (Springer New York Inc., New York, NY, USA, 2001).

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

[65] D. Foreman-Mackey, JOSS **1**, 24 (2016).

[66] M. Boyle and A. H. Mroué, Phys. Rev. D **80**, 124045 (2009).

[67] S. Ossokine, M. Boyle, L. E. Kidder, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **92**, 104028 (2015).

[68] M. Boyle, Phys. Rev. D **87**, 104006 (2013).

[69] D. McKechan, C. Robinson, and B. Sathyaprakash, Classical Quantum Gravity **27**, 084020 (2010).

[70] R. R. Hocking, Biometrics **32**, 1 (1976).

[71] L. S. Collaboration, LSC Algorithm Library software packages LAL, LALWRAPPER, and LALAPPS, https://github.com/lscsoft/lalsuite.

[72] L. Lindblom, B. J. Owen, and D. A. Brown, Phys. Rev. D **78**, 124020 (2008).

[73] D. Shoemaker (LIGO Collaboration), LIGO Document No. T0900288-v3, 2010.

[74] P. Schmidt, F. Ohme, and M. Hannam, Phys. Rev. D **91**, 024043 (2015).

[75] A. Buonanno, Y. Chen, Y. Pan, and M. Vallisneri, Phys. Rev. D **70**, 104003 (2004).

[76] C. De Boor, *A Practical Guide to Splines*, Applied Mathematical Sciences Vol. 27 (Springer-Verlag, Berlin, 1978).

[77] T. A. Davis, ACM Trans. Math. Softw. **30**, 196 (2004).

[78] T. A. Davis, ACM Trans. Math. Softw. **30**, 165 (2004).

[79] X. S. Li and J. W. Demmel, ACM Trans. Math. Softw. **29**, 110 (2003).

[80] X. Li, J. Demmel, J. Gilbert, iL. Grigori, M. Shao, and I. Yamazaki, SuperLU Users' Guide, Technical Report No. LBNL-44289, Lawrence Berkeley National Laboratory, 1999, http://crd-legacy.lbl.gov/~xiaoye/SuperLU/. Last update: August 2011.

[81] J. C. Bustillo, A. Boh, S. Husa, A. M. Sintes, M. Hannam *et al.*, arXiv:1501.00918.

[82] M. Boyle, Phys. Rev. D **84**, 064013 (2011).

[83] F. Ohme, M. Hannam, and S. Husa, Phys. Rev. D **84**, 064029 (2011).

[84] I. MacDonald, S. Nissanke, and H. P. Pfeiffer, Classical Quantum Gravity **28**, 134002 (2011).

[85] C. J. Handmer, B. Szilgyi, and J. Winicour, Classical Quantum Gravity **33**, 225007 (2016).

[86] J. Winicour, Living Rev. Relativ. **1**, 5 (1998).

[87] M. C. Babiuc, B. Szilágyi, J. Winicour, and Y. Zlochower, Phys. Rev. D **84**, 044057 (2011).

[88] N. T. Bishop, R. Gomez, L. Lehner, M. Maharaj, and J. Winicour, Phys. Rev. D **56**, 6298 (1997).

[89] N. W. Taylor, M. Boyle, C. Reisswig, M. A. Scheel, T. Chu, L. E. Kidder, and B. Szilágyi, Phys. Rev. D **88**, 124010 (2013).

[90] D. Pollney and C. Reisswig, Astrophys. J. Lett. **732**, L13 (2011).

[91] M. Favata, Classical Quantum Gravity **27**, 084036 (2010).

[92] P. D. Lasky, E. Thrane, Y. Levin, J. Blackman, and Y. Chen, Phys. Rev. Lett. **117**, 061102 (2016).

[93] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms* (The MIT Press, Cambridge, MA, 2009).

[94] P. Binev, A. Cohen, W. Dahmen, R. A. DeVore, G. Petrova, and P. Wojtaszczyk, SIAM J. Math. Anal. **43**, 1457 (2011).

[95] W. Hoffmann, Computing **41**, 335 (1989).

[96] C. R. Galley and P. Schmidt, arXiv:1611.07529.

[97] K. Cannon, A. Chapman, C. Hanna, D. Keppel, A. C. Searle, and A. J. Weinstein, Phys. Rev. D **82**, 044025 (2010).

[98] S. Field, GreedyCPP home page: https://bitbucket.org/sfield83/greedycpp.

[99] C. R. Galley (unpublished).

[100] J. S. Hesthaven, B. Stamm, and S. Zhang, ESAIM: M2AN **48**, 259 (2014).

[101] LIGO Scientific Collaboration, Advanced ligo anticipated sensitivity curves, (2011), https://dcc.ligo.org/LIGO-T0900288/public.

[102] L. Lindblom, Phys. Rev. D **80**, 064019 (2009).