A fully precessing higher-mode surrogate model of effective-one-body waveforms

Bhooshan Gadre,1,2 Michael Pürrer,1 Scott E. Field,3,4 Serguei Ossokine,1 and Vijay Varma1

1 Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, Potsdam 14476, Germany
2 Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University, Princetonplein 1, 3584 CC, Utrecht, The Netherlands
3 Department of Mathematics, University of Massachusetts, Dartmouth, MA 02747, USA
4 Center for Scientific Computing and Data Science Research, University of Massachusetts, Dartmouth, MA 02747, USA

(Dated: March 2, 2022)

We present a surrogate model of SEOBNRv4PHM, a fully precessing time-domain effective-one-body waveform model including subdominant modes. We follow an approach similar to that used to build recent numerical relativity surrogate models. Our surrogate is 5000M in duration, covers mass-ratios up to 1:20 and dimensionless spin magnitudes up to 0.8. Validating the surrogate against an independent test set we find that the bulk of the surrogate errors is less than ∼1% in mismatch, which is similar to the modelling error of SEOBNRv4PHM itself. At high total mass a few percent of configurations can exceed this threshold if they are highly precessing and they exceed a mass-ratio of 1:4. This surrogate is nearly two orders of magnitude faster than the underlying time-domain SEOBNRv4PHM model and can be evaluated in ∼50 ms. Bayesian inference analyses with SEOBNRv4PHM are typically very computationally demanding and can take from weeks to months to complete. The two order of magnitude speedup attained by our surrogate model enables practical parameter estimation analyses with this waveform family. This is crucial because Bayesian inference allows us to recover the masses and spins of binary black hole mergers given a model of the emitted gravitational waveform along with a description of the noise.

PACS numbers: 04.80.Nn, 04.25.dg, 95.85Sz, 97.80.-d 04.30.Db, 04.30.Tv

I. INTRODUCTION

The ground-breaking first detection of the merger of a binary black hole (BBH) system by Laser Interferometer Gravitational-Wave Observatory (LIGO) [1] in 2015 has opened a new window to the Universe and heralded the era of gravitational wave (GW) astronomy. Since then, the LIGO and Virgo detectors have observed an abundance of merging BBHs [2,3].

To detect and characterize BBH coalescences we rely on matched filtering and Bayesian inference [4,5] of the GW data recorded by the interferometer network. This technique requires accurate and fast models of the emitted gravitational waveform. While numerical relativity (NR) simulations [6,7] provide the most accurate waveforms available, they are computationally expensive, typically requiring anywhere between weeks to several months per simulation. In addition, NR simulations are limited in duration, and their coverage of the mass-ratio and spin space for precessing binaries is sparse [6].

As detector sensitivities improve and the number of observed BBHs increases, we are more likely to observe rare and unexpected events such as BBHs with asymmetric masses [8,9], or evidence for precession and merger kicks arising from spins misaligned with the orbital angular momentum [10,11,12], or intermediate mass black holes (BHs) [13]. The wealth of information provided by

precessing compact binary coalescences (CBCs) is invaluable to help us identify astrophysical formation channels [14,15]. Fast and accurate precessing waveform models with subdominant modes are crucial for extracting information about spin alignment [16,17].

Thus, to accurately infer the properties of CBCs with precessing spins, and binaries with unequal mass-ratios and/or high total mass where higher harmonics become important, we need reliable waveform models which can predict the GWs from such systems. The effective-one-body approach is a very successful and popular method for constructing such models. In particular, two different families of EOB models have emerged in recent years, SEOBNR [30,37] and TEOBResumS [38,42]. EOB models include calibration against NR have been shown to compare well to NR simulations not used in their construction [43]. These models also behave smoothly outside of their calibration range which makes them useful for the computation of template banks which need to span a large range of masses and spins. In this paper we focus on SEOBNRv4PHM, a recently constructed waveform model that includes precession and higher harmonics is SEOBNRv4PHM [37].

While the SEOBNRv4PHM model is faithful to NR [37], it is rather expensive to evaluate which limits its applicability in data analysis. For example, a single likelihood evaluation is tens of times slower than for waveforms with comparable physics from the Phenom family [44,45]. This means that it is challenging to use SEOBNRv4PHM for parameter estimation. In particular, standard samplers can have impractical runtimes even for relatively common events like GW150914: it would take several weeks to complete a run with the dynesty sampler [46] running
Another alternative approach is to use approximate grid-based rapid parameter estimation methods such as RIFT [48, 49] that can achieve high efficiency by evaluating the likelihood on a sparse grid and interpolating it across parameter space. While these methods have been very successfully used in production analyses [18, 23, 50, 51], they often require careful configuration and are hard to robustly scale to the wide variety of expected signals.

To maximize the utility of accurate-but-slow models, such as the SEOBNRv4PHM model, we require waveform-acceleration techniques that do not compromise waveform accuracy. The general solution to this problem is to construct interpolants or fits for waveform data after a dimensionality reduction step, where the data comes from an accurate underlying model that is typically found by solving differential equations. This approach has often been referred to as reduced-order or surrogate modeling. Terms that, while technically different, are by now used interchangeably within the gravitational-wave modelling community. Surrogate models are accurate in the region of parameter space over which they were trained as well as extremely fast to evaluate. Surrogate models have been developed to reproduce the radiation from complicated sources, including signals with $\sim 10^3$ waveform cycles [53, 56], arbitrarily many harmonic modes [52, 53, 57, 60], spinning binary systems [54, 56], precessing binary systems [58, 60], neutron star inspirals with tidal effects [61–63], large- to extreme- mass ratio systems [64, 65], and other diverse problems [64, 66, 69].

To date, surrogates have only been built for aligned-spin EOB models [54, 53, 54, 56, 61, 70]. In this paper, we build a surrogate for the precessing model SEOBNRv4PHM. Since there are, at present, no surrogate-modeling techniques for frequency-domain models of generically precessing systems we focus on time-domain techniques previously used for NR models [59, 60].

We have organized the paper as follows. In Sec. II we summarize the surrogate construction methodology and define the subdomains on which we construct the precessing surrogate. In the same section, we also describe the detailed choices we make for this particular surrogate construction. Results and accuracy of the new surrogate model, including its application for parameter estimation, are discussed in Sec. III. We devote Sec. IV to summarize our conclusions.

**II. METHODOLOGY**

**A. Waveform model**

In this section we briefly review the features of the time-domain model that is used in constructing the surrogate, SEOBNRv4PHM. This model is built using the effective-one-body (EOB) approach, where the the two-body dynamics is mapped into the motion of a reduced mass in an effective metric. Analytical information from several sources, such as PN theory, gravitational self-force, and BH perturbation theory is included in a resummed form. Results from NR simulations that accurately model the late-inspiral and highly dynamical merger regime are incorporated into the EOB framework via a calibration procedure.

SEOBNRv4PHM is a quasi-circular BBH model that includes precession and modes beyond the dominant quadrupole. This model is based on the aligned-spin multipolar model SEOBNRv4HM [44, 46] and is calibrated to 174 NR simulations in that regime. The model first computes the dynamics with full spin degrees of freedom. The dynamics is then used to construct a time-dependent co-precessing frame [71, 72] in which waveform modes are obtained by using aligned-spin expressions from SEOBNRv4HM with spins projected onto the orbital angular momentum at every point in time. The final precessing waveforms are obtained by rotating the co-precessing frame modes back to the inertial frame.

SEOBNRv4PHM has the $(2, \pm 2), (2, \pm 1), (3, \pm 3), (4, \pm 4), (5, \pm 5)$ modes in the co-precessing frame and enforces the conjugate symmetry $h_{t, m} = (-1)^{l} h_{t, -m}^{*}$ in this frame.

The waveform model takes as input the masses of the compact objects and Cartesian components of the spins which are defined in the source frame $\{e_x, e_y, e_z\}$ [73]. This frame is constructed at a particular time $t_{nf}$ (which is the start of the dynamics evolution) such that $e_z$ is along the *Newtonian* orbital angular momentum $\hat{L}_N$, $\hat{e}_z$ is along the $\hat{n}$ vector pointing from secondary to primary and $\hat{e}_y$ completes the triad. It is important to note that the same physical configuration would have different spin components in the source frame defined at two different reference times.

The waveform modes are constructed in several steps: i) the dynamics equations are integrated numerically to produce time-evolutions of the dynamical variables, ii) the dynamics is used in the construction of the EOB co-precessing frame, and the co-precessing frame inspiral modes, iii) the merger and ringdown are attached in the co-precessing frame iv) the waveform modes are rotated back to the inertial source frame.

**B. Surrogate construction method**

Our surrogate model is built using a combination of methodologies proposed in previous works [52, 53, 57].
The goal of a surrogate model is to take a pre-computed training set of waveform modes \( \{h^{l,m}(t; \tilde{\lambda}_i)\}_{i=1}^N \) at a fixed set of \( N \) points in parameter space \( \{\tilde{\lambda}_i\}_{i=1}^N \), and to produce waveform modes \( h^{l,m}(t; \tilde{\lambda}) \) at new parameter values. Here and throughout, the “S” subscript denotes a surrogate model for the true data. This modeling task is made easier by decomposing each waveform into many waveform data pieces. Our strategy is closely based on the one proposed in \[59\]. In particular, we model the waveform modes in a coorbital frame and parameterize each mode using the instantaneous value of the spin, \( \chi_{i}^{\text{coorb}}(t) \), in this frame:

\[
h^{l,m}(t; q, \tilde{x}_1, \tilde{x}_2) \leftrightarrow h^{l,m}_{\text{coorb}}(t; q, \chi_1^{\text{coorb}}(t), \chi_2^{\text{coorb}}(t)). \tag{1}
\]

Differing from \[59\], we directly model \( h^{l,m}_{\text{coorb}} \) instead of forming symmetric and anti-symmetric combinations; in the twist-up approximation used by the SEOBNRv4PHM waveform model, the asymmetric modes are identically zero.

The two-way arrow in Eq. (1) denotes a map between two descriptions of the waveform data. On the left-hand-side, we have the standard way of describing gravitational waveform data where the inertial-frame modes are parameterized by the mass ratio and spins measured at a reference time or frequency. On the right-hand-side, we have the waveform modes that are directly modeled. To go between these two representations, we also build a dynamics surrogate model that includes the following additional data pieces: the orbital phase, \( \phi(t) \), the unit quaternion that defines the coprecessing frame \( q(t) \), and the spins in the coorbital frame, \( \chi_{1,2}^{\text{coorb}}(t) \). This is done using the transformation \( T_C \) given by Eq. 25 of Ref. \[58\].

The individual spins of the two compact objects are naturally only available in the SEOBNRv4PHM model before merger. Therefore, to continue to have two spins present after the merger, we follow the approach used in NR surrogates \[60\] and artificially extend the component spins using 2 post-Newtonian (PN) spin evolution equations as described in \[74\]. We emphasize that these spins do not have a physical meaning, but are simply used as inputs to simplify the construction of fits.

For each waveform data piece, we construct a linear basis using singular value decomposition (SVD) with a prescribed tolerance. We then construct an empirical time interpolant \[78\] with the same number of empirical time nodes as basis functions for that data piece. The set of empirical time nodes \( \{T_j\} \) are chosen differently for each waveform data piece. Similar to classical interpolation (e.g., polynomial basis functions with Chebyshev nodes), using the SVD basis and set of empirical node times we are able to represent each data piece with an empirical interpolant. This interpolant is specific to the waveform data piece, which takes advantage of a nearly optimal compact representation of the parameterized time-series data \[52\].

FIG. 1. This figure shows the 14 subdomains in mass-ratio \((q)\) and effective spin parameter \((\chi_{\text{eff}})\) of the SEOBNRv4PHM surrogate. Different colors correspond to the division along \(q\) while shading types defines boundaries along \(\chi_{\text{eff}}\).

Finally, for each waveform data piece we construct a parametric fit for the data at time \(T_j\). As a concrete example, suppose we need to build a parametric fit of the co-orbital 22-mode data, \( h^{22}_{\text{coorb}} \), at time \(T_j\). This 7-dimensional fit is given by

\[
h^{22}_{\text{coorb}}(T_j; q, \chi_1^{\text{coorb}}(T_j), \chi_2^{\text{coorb}}(T_j)) \approx h^{22}_{\text{coorb}, \beta}(T_j; q, \chi_1^{\text{coorb}}(T_j), \chi_2^{\text{coorb}}(T_j)), \tag{2}
\]

where we have explicitly shown that the data is parameterized not by the spins at some reference time but rather the spins at time \(T_j\). The function \( h^{22}_{\text{coorb}, \beta} \) is given by a linear combination of basis functions. Similar to Ref. \[59\], we choose the basis functions to be a tensor product of 1D monomials in the spin components and \(x(q)\), where \(x\) arises from an affine mapping from the interval \([q_{\text{min}}, q_{\text{max}}]\) to the standard interval \([-1, 1]\). We consider up to cubic functions in \(x\) and up to quadratic functions in the spin components. Using all possible terms of this form would lead to overfitting and low-quality fits. We correct for this issue by using the forward-stepwise greedy fitting method described in Appendix A of Ref. \[58\], which selects at most 150 relevant terms in such a way that avoids overfitting. Increasing the polynomial order in spin components or the mass ratio leads to a very steep increase in the number of terms, making the fit computationally infeasible.

C. Domain decomposition

Our goal is to build an extensive time-domain surrogate of SEOBNRv4PHM that can be used to analyze most of the BBH signals we expect to see with current and near-future ground based detectors. To accomplish this, we build our model on a mass ratio interval of \(q := m_1/m_2 \in [1, 20]\), which encompasses all BBH signals detected to date. However, we restrict the spin magnitudes to satisfy \(|\chi_i| \leq 0.8\) and place no restrictions on the
spin surrogate models \cite{SEOBNRv4PHM} for which high spin simulations are rather sparse. For the SEOBNRv4PHM model, where we do not have this restriction, we have explored building surrogates for spin magnitudes up to 1. However, we found that surrogate errors increased rather steeply beyond a spin magnitude of 0.8 and therefore imposed this spin bound. We leave an extension of the surrogate up to extremal spins as future work. We construct the surrogate with a duration of 5000\(M\) prior to merger.

Despite significant experimentation, we were unable to build high-accuracy surrogate models when attempting a global parametric fit over the full parameter space mentioned above. This is a common issue for ambitious regression problems, and the general solution is domain decomposition. In the domain decomposition approach, we break up the large parameter space into smaller subdomains and build models on each patch. The general expectation is that high-accuracy parametric fits over a smaller subdomain will be easier to achieve, since variations in the waveform data pieces are reduced.

We divide our target parameter domain into 14 subdomains in the effective spin parameter, \(\chi\), and mass ratio directions. There are two reasons to choose subdomains in mass ratio and effective spin. First, the number of cycles in GW signals of fixed duration depend most strongly on \(q\) and \(\chi\). Second, throughout the inspiral \(q\) is constant while \(\chi\) is nearly so \cite{8}. For the mass ratio, we place our domain boundaries at \(q = 1, 2, 4, 7, 10, 15,\) and 20. For \(q > 4\), the mass-ratio domains are further divided into 3 subdomains each in effective spin with boundaries at \(\chi = (-0.8, -0.3, 0.3, 0.8)\). The resulting subdomains are shown in Fig. [1]. To have a smooth combined surrogate across the subdomain boundaries, we extend the subdomains so that they overlap by \(\pm 0.25\) in \(q\) and by \(\pm 0.05\) in \(\chi\).

**D. Choice of time nodes**

In this section, we discuss the time resolution used for the waveform and dynamics surrogates. To construct a surrogate for the waveform modes, we choose 2000 time nodes equally spaced in orbital phase for a representative waveform in the subdomain. We build a reduced order empirical interpolant representation for each of the waveform modes in the coorbital frame, as described in Sec. [II.B] with an SVD tolerance of \(10^{-5}\). To build a dynamics surrogate model, we choose 25 time nodes per orbit. We find that these particular choices sample the data pieces in time well enough so that we can interpolate this data to the desired sampling frequency with high accuracy.

**E. Training and validation set generation**

To construct an accurate surrogate, we first need to build a training set that is sufficiently dense and captures the variations of the data well at each time node. To achieve this, in each subdomain we proceed in three steps:

1. We start with 2000 samples in each subdomain uniform in \(q\) and for each component spin, uniform in a spherical volume with \(|\chi| \leq 0.85\). While our target problem is defined on \(|\chi| \leq 0.8\), we find that sampling up to 0.85 will result in models that retain high accuracy near \(|\chi| \approx 0.8\).

2. Next, we add 2000 sample points chosen uniformly in \(q\) and uniformly in \(\chi\). To sample uniformly in \(\chi\), we use rejection sampling on the spin component samples drawn uniformly in the spherical volume as above.

3. Finally, to sample adequately in the effective precession spin parameter \(\chi_p\), we start by drawing 8000 points uniform in \(q, \chi_p,\) and \(\chi\), again using rejection sampling, and then choose 2000 of the samples which maximize the Euclidean distance in the \(\chi_p - \chi\) plane.

With the above choice of sampling method, we net a total of 6000 training points per subdomain, and thus a grand total of 84,000 points for the entire surrogate. The procedure allows us not only to have an adequate coverage in \(\chi_p\) and \(\chi\) (which encode the dominant precessional and aligned-spin effects), but also provides sufficient samples in the remaining spin degrees of freedom in the waveform space to capture subdominant spin effects.

In the construction of the training set mentioned above, \(\chi\) and \(\chi_p\) are measured at the reference dimensionless frequency \(M \omega = 0.014\). As the constructed surrogate has a duration of 5000\(M\) until merger, our spin distribution evolves from the one sampled at the reference frequency to the one measured at the reference time for the surrogate of \(-5000M\). The choice of the surrogate duration roughly corresponds to the waveform’s \((2, 2)\) mode starting at 20 Hz when both the component masses are \(25M_\odot\).

We verify the sufficiency of the chosen training set by validating the surrogate model against an independent set of validation waveforms sampled using the same procedure as used for the training set. We reach the final training set by iteratively enriching it with configurations which have mismatches greater than 1.5% in the validation set. We find that this enrichment improves the surrogate model only for some of the subdomains. In particular, the surrogate accuracy can improve if the enriched training set resolves under-represented features and improves overall goodness of fit. We empirically found that after two such iterations the model’s accuracy no longer improves.
III. RESULTS

A. Accuracy

In this section we define error measures for surrogate components and study their impact on the inertial waveform modes generated by the surrogate. We also compute the noise-weighted Fourier domain matches typically used in gravitational-wave data analysis.

1. Error measures for surrogate data pieces

The approximation quality of each surrogate data piece is assessed by computing a root-mean-square (RMS) error

$$\mathcal{E}_{\text{rms}}[X] := \sqrt{\frac{\int_{t_{\text{min}}}^{t_{\text{max}}} (\dot{X} - \dot{X}_{\text{EOB}})^2 dt}{T}},$$

where $X_{\text{EOB}}(t)$ is any time dependent waveform data piece taken from the SEOBNRv4PHM model, $\dot{X}(t)$ is the surrogate for this data piece, and $T = t_{\text{max}} - t_{\text{min}}$ is the total time duration.

In addition, we consider a relative $L_2$-type error as defined by [58] (which can be motivated from a white-noise time-domain mismatch)

$$\mathcal{E}_X := \frac{1}{2} \int_{t_{\text{min}}}^{t_{\text{max}}} \sum_{\ell,m} |h_{\ell m}^X - h_{\ell m}^\text{EOB}|^2 dt,$$

where $h_{\text{EOB}}(t)$ is the true inertial frame SEOBNRv4PHM waveform for a particular BBH configuration. In contrast, $h_X(t)$ is computed using the surrogate waveform decomposition described in Sec. IIB and the following prescription: all waveform data pieces are given by the true SEOBNRv4PHM data except for one particular data piece, $X$, which is taken from the surrogate for this quantity. This construction allows us to study how the surrogate modeling errors in a particular data piece (e.g. the orbital phase, the quaternions, or the coorbital modes) affect the overall inertial modes output by the surrogate. Furthermore, we also compute $\mathcal{E}_X$ for the complete surrogate i.e. using only surrogate data pieces without any true SEOBNRv4PHM data, as this is the way the surrogate will be used in practical applications and it contains the total error budget.

Fig. 2 shows error histograms for individual data pieces used to construct the surrogate on the training set and the overall surrogate error. We can see from both panels that the orbital phase (and hence the orbital frequency) is the source of the dominant error as compared to the inertial frame waveform errors in the right panel. This shows that the phase (or frequency) model largely limits the accuracy of the surrogate. On the validation set which is comprised by more than 84000 independent test cases we can see similar behavior as on the training set, as shown in Fig. [3].

In addition, we can also observe from the $\mathcal{E}_X$-histograms that on both data sets the coorbital modes and quaternion fits do not limit the surrogate accuracy. The RMS error histograms suggest that the tails showing large errors in the spin fits do not affect the coorbital mode and quaternion fits. To complement the histograms, we note that the surrogate fits worsen as we approach the merger. This affects the accuracy of the surrogate model for BBH systems with high total masses.

2. Match computations

We further check the accuracy of the surrogate model by computing matches against the original SEOBNRv4PHM waveform model. We use the “min-max match” (denoted by $M$) as defined by Eq. B10 in Appendix B of [53] along with the advanced LIGO design power spectral density (PSD) [54]. We perform the match computation for two values of the total mass of the binary, either $50M_\odot$ or $200M_\odot$. Furthermore, we fix the inclination angle of the binary to 60° and use a frequency range between $f_{\text{min}} = 25 \text{ Hz}$ and $f_{\text{max}} = 2048 \text{ Hz}$. We specify the initial configuration for SEOBNRv4PHM waveform generation in the $\hat{L}_\text{N}$-frame.

To generate the corresponding surrogate waveform and data pieces, we transform the spin vectors from the $\hat{L}_\text{N}$-frame to the coprecessing frame which is the one used for the surrogate model.

In the following we discuss match results for the training and validation sets at a total of $50M_\odot$ and $200M_\odot$, as shown in Figs. [4][6][5] and [7]. In each of these figures, we show scatter plots in mass-ratio, effective aligned and precession spin parameters to better understand the distribution of matches over the parameter space and thus learn about the accuracy of the surrogate over parameter space.

We start with Fig. 4 which shows the matches between the surrogate and the training set of SEOBNRv4PHM waveforms at a total masses of $50M_\odot$. We see that the surrogate is in good agreement with the parent model, with the bulk of the matches being better than 0.99 (or, equivalently, a mismatch of less than 1%). We find that only 68 configurations ($\lesssim 0.1\%$) in the training set exceeds this threshold. We observe that configurations with high mismatches tend to have large effective precessional spin parameter $\chi_{\text{P}}$, slightly anti-aligned $\chi_{\text{eff}}$, and cluster towards more unequal mass-ratios $q$. This indicates that the strongly precessing configurations are the most difficult to accurately model, and we discuss possible explanations for this in the conclusion. Matches calculated on the independent validation set shown in Fig. [6] closely parallel the result we observed on the training set. We find 53 configurations which exceed 1% mismatch and overall the worst mismatch lies around 2% as shown in Fig. 8.
FIG. 2. Histograms of errors due to the errors in respective surrogate data pieces on the training set. \textit{Left:} RMS errors due to dynamics quantities: spin vectors (solid violet and dash-dotted brown), unit quaternions (red dashed), and orbital phase (yellow long dash-dotted); \textit{Right:} $L^2$-type errors for the inertial modes due to immediate dependent quantities: inertial modes (blue dash-dotted; as as the full surrogate), orbital phase (orange long dash-dotted), quaternions (green dashed), and coorbital modes (pink).

FIG. 3. Histograms of errors due to respective surrogate data pieces on the validation set. Description of panels and plotted quantities as in Fig. 2.

At higher total mass the match computation focuses on higher frequency content in the surrogate. We have already remarked above, that surrogate fits tend to have lower accuracy for time nodes which are close to the merger and thus surrogate accuracy worsens towards higher frequencies. Therefore, we expect that matches will overall worsen for higher mass systems. We can see in Fig. 5 that this is clearly visible at a total mass of 200$M_\odot$. For this total mass, very few GW cycles are in LIGO band and they are concentrated in the merger-ring-down region. We see that the distribution pattern of the matches is similar to the 50$M_\odot$ result, but there is longer tail of low matches: we find that 4\% of the cases exceed 1\% mismatch and 70 cases have more than 10\% mismatch. Again, the behavior on the validation set shown in Fig. 7 follows the training set very closely, finding 4\% cases with a mismatch exceeding 1\% and 64 cases with a mismatch greater than 10\% as can be seen in Fig. 8.

We further test the validity of the surrogate model for extrapolated spin magnitudes exceeding the limit of 0.8 we used in the surrogate construction. To do this, we compare the surrogate to the SEOBNRv4PHM waveforms with at least one the component spin magnitudes in the range (0.8, 0.99) for a total mass of 50$M_\odot$. For spin magnitudes up to 0.9, we obtain mismatches up to 5\%. If we further increase the spin on the larger BH beyond 0.9 the mismatches can reach values as high as 20\% if the mass ratio is larger than 5. For mass ratios less than 5, mismatches do not exceed 10\%. In addition, mismatches do not exceed 10\% for any value of the secondary spin magnitude if the spin magnitude of the primary is below 0.9 – indeed we expect that the spin of the secondary BH should only have a very small effect on the waveform if the mass ratio is very unequal.

B. Speed

Our main purpose of building a surrogate for SEOBNRv4PHM is to improve the speed of waveform generation while preserving accuracy, so that it can be used for characterizing BBH mergers. In this section we focus on the achieved speedup between the SEOBNRv4PHM
model and our surrogate, and compare to one of the efficient phenomenological waveform models which describes waveforms with similar physical assumptions.

We compare the surrogate waveform generation time against that of \texttt{SEOBNRv4PHM} and the \texttt{IMRPhenomXPHM} \cite{45} waveform model for different mass ratios given a fiducial spin configuration. We choose a starting frequency of 25 Hz. As our surrogate is only 5000 M long, it may not be able to start from 25 Hz for a few low mass and asymmetric mass configurations. In that case we use the lowest allowed surrogate frequency to generate waveforms and compare the costs. The total cost we consider includes the computational cost tapering and Fourier transform (FFT) for the surrogate and \texttt{SEOBNRv4PHM} as both of these models are time domain.

In Fig. 4 we plot the ratio of the waveform generation times between either \texttt{SEOBNRv4PHM} or \texttt{IMRPhenomXPHM} and the surrogate as a function of total mass and mass ratio. We can see that the surrogate is about 40-70 times faster than the original \texttt{SEOBNRv4PHM} model and just 5-8 times slower than the frequency domain phenomenological \texttt{IMRPhenomXPHM} waveform. This makes it feasible to do Bayesian parameter estimation studies which we describe in the next section.

C. Application: parameter estimation

As a practical application of the new surrogate, we perform Bayesian inference on GW signals. We consider two cases, (a) recovering a software injection of an \texttt{SEOBNRv4PHM} signal assuming an average (zero) noise realization, and (b) performing inference on real data for GW150914.

For case (a) we prepared the software injection using \texttt{pycbc} \cite{85}. The signal has a detector-frame total mass $70 M_\odot$, (inverse) mass-ratio $q := 1/q = 1/5$, dimensionless component spins of $\vec{\chi}_1 = (0.5, 0.05, 0.3)$ and $\vec{\chi}_2 = (0.3, -0.2, 0.3)$ at a reference frequency of 20 Hz. The inclination angle of the signal is $\theta_{JN} = 2.58$. The
FIG. 6. Scatter plots of matches between the surrogate and underlying precessing EOB waveform model for the validation set and a total mass of 50 $M_\odot$. The setup of the panels and the shown matches follows Fig. 4.

FIG. 7. Scatter plots of matches between the surrogate and underlying precessing EOB waveform model for the validation set and a total mass of 200 $M_\odot$. The setup of the panels and the shown matches follows Fig. 4.

FIG. 8. This figure shows cumulative histograms of matches ($M$) between the surrogate and the SEOBNRv4PHM waveform model, for training (TS) and validation (VS) sets and two different total masses 50 and 200 $M_\odot$.

chosen signal lies in the vicinity of subdomain boundaries ($q = 4$, $\chi_{\text{eff}} = 0.3$) of the surrogate model. Therefore, it constitutes a test of the smoothness of the surrogate at these boundaries where we switch between independently trained surrogates defined on adjacent subdomains (see Fig. 1). The signal has a network SNR of $\sim 27$ in the three detector advanced LIGO - Virgo HLV network. We use the bilby Bayesian inference code [86] with the dynesty [87] nested sampler [88, 89].

Marginal posterior PDFs are shown in Fig. 10 where the signal is recovered with the surrogate and, as a comparison, with IMRPhenomXPHM [15]. Since we are using a zero-noise injection we expect that the likelihood peaks close to the signal parameters, and since the signal is loud, this should also be the case for the posterior distribution. We can verify that indeed the contours indicating the 90% credible regions of the posterior PDFs are close to centered around the true signal parameters indicated by red asterisks for all parameters shown. We do not see any indication that the posterior would be non-smooth at subdomain boundaries which is expected given the demonstrated accuracy of the surrogate com-
FIG. 9. This figure shows the speed-up factor between the precessing SEOBNRv4PHM model and its surrogate (squares), and between the IMRPhenomXPHM waveform model and the SEOBNRv4PHM surrogate (triangles). The fiducial starting frequency is 25 Hz (see text for the details) with fixed spin vectors \( \vec{\chi}_1 = (0.6, 0.1, 0.2) \) and \( \vec{\chi}_2 = (0.1, 0.1, 0.6) \).

Compared to the SEOBNRv4PHM model. The IMRPhenomXPHM posteriors show a noticeable bias in chirp mass and effective precession spin, with the true parameter values close to 90% credible region boundary. This is not too surprising, since SEOBNRv4PHM and IMRPhenomXPHM are expected to differ at the level of up to a few percent in mismatch (see Fig. 9 in Ref. [90]). Smaller discrepancies are seen in distance and inclination. We can also see that the shape and extent of the 90% credible regions is overall similar, but, for instance the mass-ratio, is estimated more precisely for the surrogate compared to IMRPhenomXPHM.

In case (b) we analyzed GW150914 with the surrogate model and compared to the latest published posteriors available from the LVC, which are from the GWTC-1 catalog [3]. Our analysis of GW150914 used the parallel version of the bilby code [47] and open data from GWOSC [61]. The parameter estimation run with the surrogate model took 18 hours on 144 cores. If the dominant cost is due to waveform evaluations, we estimate the corresponding run without a surrogate model would have taken about 5 weeks. Furthermore, in realistic production settings, the sampler is typically run multiple times with independent seeds and different settings such as the choice of prior or trigger time.

When interpreting our results, we note that settings between our analysis and the LVC analysis are different, and the LVC posteriors are combined from the IMRPhenomPv2 [32] and SEOBNRv3 [32] waveform models, neither of which contains higher harmonics beyond \( \ell = 2 \). While SEOBNRv3 models generic precessing binaries, IMRPhenomPv2 uses an effective precession model. In contrast, the surrogate includes higher harmonics up to \( \ell = 5 \). Given these caveats, we see in Fig. 11 that the agreement between the two posteriors is overall good, but the 90% credible regions tend to be somewhat tighter for the surrogate. To give a quantitative comparison, we find that the Jensen-Shannon divergence (JSD) between the LVC and surrogate marginal posteriors have JSD values smaller than 0.04 bits with the largest value occurring for the effective precession spin. When comparing different waveform models such JSD values have been shown to occur in LIGO O3a analyses [4]. Even though higher harmonics were not found to be important [24] for this close to equal-mass signal, it is plausible that the surrogate which includes these higher harmonics would have somewhat tighter posteriors. The fact that IMRPhenomPv2 does not model generic precession could also play a role. Furthermore, the analysis of GW150914 performed in Ref. [94] with various Phenom models, and in particular, IMRPhenomXPHM, corroborates that more physically complete waveform models which include higher harmonics find tighter posteriors for this event (see. Table III in Ref. [94] for credible intervals). Comparing further, we find that both the SEOBNRv4PHM surrogate and IMRPhenomXPHM infer noticeably higher medians for the effective precession spin parameter than the GWTC-1 analysis: SEOBNRv4PHM surrogate finds a median \( \chi_P \) of 0.31 and IMRPhenomXPHM 0.44, compared to a median of 0.23 in the GWTC-1 posterior. In contrast, the medians of the mass-ratio and the effective-aligned spin are very close to the values found for the GWTC-1 analysis. In summary, we have demonstrated that the surrogate posterior is broadly consistent with, and slightly more constrained than the published LVC results, and in good agreement with posteriors obtained with IMRPhenomXPHM.

IV. CONCLUSION

In this paper we have presented a surrogate for the precessing SEOBNRv4PHM waveform model based on techniques developed for NR surrogates with a few modifications. In particular, we have used domain decomposition (see Sec. II C) to deal with the large range the surrogate covers in mass-ratio while keeping the polynomial fits below cubic order. In Sec. III E we discussed that a larger training set along with an enrichment strategy is beneficial to achieve good accuracy for at least some of the subdomains. Over all domains we used a dataset of about 170,000 waveforms in total when combining the training and validation sets which is about two orders of magnitude more waveform data than used for NR surrogates so far. This surrogate is built up to mass ratio \( q = 20 \) and spin magnitudes of 0.8. When we attempted to build a surrogate for higher spin magnitudes we found that the fit accuracy degraded substantially and we leave

---

1 We have used the following dynasty sampler settings: \( \text{nlive}=2000 \), \( \text{nact}=20 \), and a \( d\log z \) stopping criteria of 0.1.
such an extension for future work.

As has been observed previously for NR surrogates \cite{60}, we find that the accuracy of our surrogate is limited by the modeling of the orbital phase (or frequency) as shown in Figs. 2 and 3. We observe that the accuracy of the fits decreases as we approach the merger, presumably because the waveforms are less smooth in this regime. As shown in Figs. 4 and 6, we find that the bulk of the mismatches between the surrogate and SEOBNRv4PHM is smaller than 1\% with the worst mismatch around 2\% for a total mass of 50M⊙. Since the surrogate fits worsen as we approach the merger, the surrogate accuracy degrades for systems of high total masses. At a total mass of 200M⊙ we find that 4\% of the configurations exceed a mismatch of 1\%. The worst mismatches are found at large values of the effective precession spin parameter \( \chi_p \), slightly negative effective aligned spin \( \chi_{eff} \) and mass ratios greater than 4. The surrogate’s reduced performance in this region of parameter space could be due to any combination of the modeling approximations or the underlying SEOBNRv4PHM waveform model itself. To check the former, we carried out extensive experimentation with parameters defining the surrogate model (fit tolerances, the data pieces’ parameterizations, number of basis elements, number of dynamics nodes, etc). To check the SEOBNRv4PHM waveform data we experimented with turning off the non-quasi-circular orbit corrections and checking non-smoothness diagnostics. Despite these efforts, we have been unable to determine the underlying cause of this problem.

While it is possible to evaluate the surrogate outside of its training domain, such as extrapolating the surrogate to higher spin magnitudes than the training bound of 0.8, this incurs an increase of mismatches up to 5\% (a < 0.9) and 20\% (a < 0.99 and q > 5) for q total mass of 50M⊙.

In Fig. 9 we demonstrate that our surrogate is about 50 times faster than SEOBNRv4PHM and about 5 times slower than the IMRPhenomXPHM waveform model. This is a significant speedup compared to the original EOB model and will enable the use of standard sampling methods for Bayesian inference with this waveform family.

As an application of our surrogate to data analysis studies we showed in Fig. 10 that it allows for unbiased estimation of binary parameters when the precessing GW signal is modeled by SEOBNRv4PHM and we are using an average “zero” noise realization. The surrogate also leads to posterior distributions which are consistent with, albeit slightly tighter, than those found in past LVC anal-
FIG. 11. Marginal posterior PDFs for \textit{SEOBNRv4PHMSur} and LVC GWTC-1 [3] overall posteriors for the binary black hole signal GW150914. The panels show 90% credible regions for the following quantities: top left: source-frame component masses, top right: the source-frame chirp mass and (inverse) mass-ratio, bottom left: the effective aligned and effective precessing spin parameters, and bottom right: the luminosity distance and inclination angle. The surrogate recovers posteriors which agree well with published results by the LVC [3], but are overall slightly tighter.

yses of GW150914 as demonstrated in Fig. 11.

The surrogate we present here has a length of 5000 M and therefore cannot model the inspiral for low mass binaries. It is therefore of interest to build surrogates for significantly longer waveforms so as to enable accelerated analysis of costly low mass GW events. A first step is to investigate the limitations of the current surrogate construction method. To this end we also constructed a longer surrogate (covering 80000 M), which corresponds to a starting frequency of 20 Hz for an equal mass binary with a total mass of 20M⊙. This surrogate only includes two subdomains in mass ratio with the boundaries at \( q = 1, 2, \) and 4. We found a surrogate accuracy similar to the shorter surrogate we have discussed in-depth in this paper when restricting to the overlapping region of parameter space. This longer surrogate turns out to be more than order of magnitude faster computationally than \textit{SEOBNRv4PHM}. We found that the construction of this longer surrogate for each of its subdomains is computationally prohibitive when using 25 points per orbit to construct the dynamical surrogate which would result in 20000 time points. Apart from this computational issue, we did not find any hard limitations in the current method, in particular its reliance on the solution of an ODE system in time and the fitting errors. Yet, even a duration of 80000 M is not long enough in practice as the duration increases very rapidly as the total mass and/or the starting frequency decreases, and for low-mass binaries detectable by ground-based detectors, we would need to cover durations of millions of M, which does not seem feasible with the current methodology. Therefore, it will be very interesting to study alternative surrogate construction methods or hybridization of the surrogate with post-Newtonian or post-adiabatic EOB waveforms in the inspiral.

ACKNOWLEDGEMENTS

We thank Jonathan Blackman, Alessandra Buonanno, Steffen Grunewald, Sylvain Marsat, Ajit Mehta, and Harald Pfeiffer for helpful discussions throughout the project. We like to thank Marta Colleoni for carefully going through the manuscript and providing useful comments. This research has made use of data, software and/or web tools obtained from the Gravitational Wave
Open Science Center (https://www.gw-openscience.org/), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. S.E.F. acknowledges support from NSF Grants Nos. PHY-2110496 and PHY-1806665. This project has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 896869. The significant part of the computational work for the project is done on Hypatia, a general purpose gravitational wave data analysis cluster and a couple large workstations named Saga and Odin at AEI Potsdam-Golm. A portion of the computational work of this project was performed on the CARNIe cluster at UMassD, which is supported by the ONR/DURIP Grant No. N00014181255. LIGO Laboratory and Advanced LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain.

[1] B. P. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. Lett. 116, 061102 (2016) [arXiv:1602.03837 [gr-qc]]
[2] B. P. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. X 6, 041015 (2016) [Erratum: Phys. Rev. X 8, 039903 (2018)], arXiv:1606.04856 [gr-qc]
[3] B. P. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. X 9, 031040 (2019) [arXiv:1811.12907 [astro-ph.HE]].
[4] R. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. X 11, 021053 (2021) [arXiv:2010.14527 [gr-qc]].
[5] R. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. X 11, 021053 (2021) [arXiv:2111.03606 [gr-qc]].
[6] B. S. Sathyaprakash and S. V. Dhurandhar, Phys. Rev. D 44, 3819 (1991).
[7] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, Phys. Rev. D 85, 122006 (2012) [arXiv:gr-qc/0509116].
[8] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlan, et al., Physical Review D 91, 042003 (2015).
[9] M. Boyle et al., Class. Quant. Grav. 36, 195006 (2019) [arXiv:1904.04831 [gr-qc]].
[10] P. Ajith et al., Gravitational waves. Numerical relativity - data analysis. Proceedings, 9th Edordo Amaldi Conference, Amaldi 9, and meeting, NRDA 2011, Cardiff, UK, July 10-15, 2011, Class. Quant. Grav. 29, 124001 (2012) [Erratum: Class. Quant. Grav. 30, 199401 (2013)], arXiv:1201.5319 [gr-qc].
[11] J. Asa et al. (LIGO Scientific, Virgo, NINJA-2), Class. Quant. Grav. 31, 115004 (2014) [arXiv:1401.0929 [gr-qc]].
[12] http://www.einstein.gatech.edu/catalog/
[13] K. Jani, J. Healy, J. A. Clark, L. London, P. Laguna, and D. Shoemaker, Class. Quant. Grav. 33, 204001 (2016) [arXiv:1605.03204 [gr-qc]].
[14] http://ccrg.rit.edu/~RITCatalog/
[15] J. Healy, C. O. Lousto, Y. Zlochower, and M. Campanelli, Class. Quant. Grav. 34, 224001 (2017) [arXiv:1703.03423 [gr-qc]].
[16] J. Healy and C. O. Lousto, Physical Review D 102, 104018 (2020).
[17] J. Healy and C. O. Lousto, “The fourth rth binary black hole simulations catalog: Extension to eccentric orbits,” (2022) [arXiv:2202.00018 [gr-qc]].
[18] R. Abbott et al. (LIGO Scientific, Virgo), Astrophys. J. Lett. 896, L44 (2020) [arXiv:2006.12611 [astro-ph.HE]].
[19] R. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. D 102, 043015 (2020) [arXiv:2004.08342 [astro-ph.HE]].
[20] M. Hannam, C. Hoy, J. E. Thompson, S. Fairhurst, V. Raymond, and m. o. t. LIGO (VIRGO), (2021), arXiv:2112.11300 [gr-qc].
[21] C. Hoy, C. Mills, and S. Fairhurst, (2021), arXiv:2111.10455 [gr-qc].
[22] V. Varma, S. Biscoveanu, T. Islam, F. H. Shaik, C.-J. Haster, M. Isi, W. M. Farr, S. E. Field, and S. Vitale, (2022), arXiv:2201.01302 [astro-ph.HE].
[23] R. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. Lett. 125, 101102 (2020) [arXiv:2009.01075 [gr-qc]].
[24] B. Farr, D. E. Holz, and W. M. Farr, Astrophys. J. Lett. 854, L9 (2018) [arXiv:1709.07896 [astro-ph.HE]].
[25] C. Kimball, C. Talbot, C. P. L. Berry, M. Carney, M. Zevin, E. Thrane, and V. Kalogera [Astrophys. J. 900, 177 (2020)] arXiv:2005.00023 [astro-ph.HE].
[26] J. Calderón Bustillo, P. Laguna, and D. Shoemaker, Phys. Rev. D 95, 104038 (2017) [arXiv:1612.02340 [gr-qc]].
[27] R. Green, C. Hoy, S. Fairhurst, M. Hannam, F. Pannarale, and C. Thomas, Phys. Rev. D 103, 124023 (2021) [arXiv:2010.04131 [gr-qc]].
[28] O. Pratten, P. Schmidt, R. Buscichio, and L. M. Thomas, Phys. Rev. Research 2, 043096 (2020).
[29] V. Varma, M. Isi, S. Biscoveanu, W. M. Farr, and S. Vitale, Phys. Rev. D 105, 024045 (2022) [arXiv:2107.09692 [astro-ph.HE]].
[30] A. Buonanno and T. Damour, Phys. Rev. D 59, 084006 (1999) [arXiv:gr-qc/9811091].
[31] A. Buonanno and T. Damour, Phys. Rev. D 62, 064015 (2000) [arXiv:gr-qc/0001013].
[32] 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) [arXiv:1307.6232 [gr-qc]].
[33] S. Babak, A. Taracchini, and A. Buonanno, Phys. Rev. D 95, 024010 (2017) [arXiv:1607.05661 [gr-qc]].
[34] A. Bohé, L. Shao, A. Taracchini, A. Buonanno, S. Babak, I. W. Harry, I. Hinder, S. Ossokine, M. Pürrer, V. Raymond, T. Chu, H. Fong, P. Kumar, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace,
dfinstad, F. Pannarale, T. Massinger, S. Kumar, M. Tápai, L. Singer, S. Khan, S. Fairhurst, A. Nielsen, S. Singh, K. Chandra, shasvath, and B. U. V. Gadre, "gwastro/pycbc:," (2021).

[86] G. Ashton, M. Hübner, P. D. Lasky, C. Talbot, K. Ackley, S. Biscoveanu, Q. Chu, A. Divakarla, P. J. Easter, B. Goncharov, F. Hernandez Vivanco, J. Harms, M. E. Lower, G. D. Meadors, D. Melchor, E. Payne, M. D. Pitkin, J. Powell, N. Sarin, R. J. E. Smith, and E. Thrane, Astrophys. J. Suppl. 241, 27 (2019), arXiv:1811.02042 [astro-ph.IM].

[87] J. S. Speagle, Monthly Notices of the Royal Astronomical Society 493, 3132 (2020) arXiv:1904.02180 [astro-ph.IM].

[88] J. Skilling, in Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering American Institute of Physics Conference Series, Vol. 735, edited by R. Fischer, R. Preuss, and U. V. Toussaint (2004) pp. 395–405.

[89] J. Skilling, Bayesian Analysis 1, 833 (2006).

[90] M. Colleoni, M. Mateu-Lucena, H. Estellés, C. García-Quirós, D. Keitel, G. Pratten, A. Ramos-Buades, and S. Husa, Phys. Rev. D 103, 024029 (2021) arXiv:2010.05830 [gr-qc].

[91] R. Abbott et al. (LIGO Scientific, Virgo), SoftwareX 13, 100658 (2021) arXiv:1912.11716 [gr-qc].

[92] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Phys. Rev. Lett. 113, 151101 (2014) arXiv:1308.3271 [gr-qc].

[93] B. P. Abbott et al. (LIGO Scientific, Virgo), Class. Quant. Grav. 34, 104002 (2017) arXiv:1611.07531 [gr-qc].

[94] M. Mateu-Lucena, S. Husa, M. Colleoni, H. Estellés, C. García-Quirós, D. Keitel, M. d. L. Planas, and A. Ramos-Buades, (2021), arXiv:2105.05960 [gr-qc].