


Survey of four precessing waveform models for binary black hole systems

Jake Mac Uilliam¹, Sarp Akçay¹, and Jonathan E. Thompson²

¹*University College Dublin, Belfield, D4, Dublin, Ireland*

²*Theoretical Astrophysics Group, California Institute of Technology, Pasadena, California 91125, USA*

 (Received 13 February 2024; accepted 20 March 2024; published 29 April 2024)

Angular momentum and spin precession are expected to be generic features of a significant fraction of binary black hole systems. As such, it is essential to have waveform models that faithfully incorporate the effects of precession. Here, we assess how well the current state-of-the-art models achieve this for waveform strains constructed only from the $\ell = 2$ multipoles. Specifically, we conduct a survey on the faithfulness of the waveform models SEOBNRV5PHM, TEOBRESUMS, IMRPHENOMTPHM, IMRPHENOMXPHM to the numerical relativity (NR) surrogate NRSUR7DQ4 and to NR waveforms from the SXS catalog. The former assessment involves systems with mass ratios up to 6 and dimensionless spins up to 0.8. The latter employs 317 short and 23 long SXS waveforms. For all cases, we use reference inclinations of zero and 90° . We find that all four models become more faithful as the mass ratio approaches unity and when the merger-ringdown portion of the waveforms are excluded. We also uncover a correlation between the coprecessing $(2, \pm 2)$ multipole mismatches and the overall strain mismatch. We additionally find that for high inclinations, precessing $(2, \pm 1)$ multipoles that are more faithful than their $(2, \pm 2)$ counterparts, and comparable in magnitude, improve waveform faithfulness. As a side note, we show that use of uniformly filled parameter spaces may lead to an overestimation of precessing model faithfulness. We conclude our survey with a parameter estimation study in which we inject two precessing SXS waveforms (at low and high masses) and recover the signal with SEOBNRV5PHM, IMRPHENOMTPHM, and IMRPHENOMXPHM. As a bonus, we present preliminary multidimensional fits to model unfaithfulness for Bayesian model selection in parameter estimation studies.

DOI: [10.1103/PhysRevD.109.084077](https://doi.org/10.1103/PhysRevD.109.084077)

I. INTRODUCTION

Over the course of three observing periods, the terrestrial network of gravitational wave interferometers has detected roughly 100 compact binary inspiral-merger events [1–29]. An overwhelming majority of these involved binary black hole systems where each compact body is expected to have non-negligible spin angular momentum, which has been confirmed by the analysis of the binary black hole population thus far detected [14,22,30]. Depending on the binary formation scenario, a significant subpopulation of these binaries can have spins misaligned with the orbital angular momentum [31–36].

The leading-order general relativistic effect of such misaligned spins is the precession of the orbital angular momentum vector around the total angular momentum of the system, which leaves a faint, but detectable imprint on the gravitational waveform as amplitude and phase modulations on a timescale 1.5 post-Newtonian (PN) orders longer than the orbital timescale [37,38], and at an order 1.5PN higher in the phase than the leading-order terms [39,40]. As such, spin effects are harder to infer, or even detect, than the chirp mass and the (symmetric) mass ratio of the binary system. Nonetheless, information

about black hole spins has been obtained successfully [6,13,14,16,17,19,20,22,23,25,26,28,30,41–44]. One way to convey this information is to provide distributions for the inferred spin magnitudes and tilts, the latter with respect to some reference frame. However, thus far, such posteriors have been mostly uninformative with a few exceptions as illustrated by Figs. 6, 10, and 11 of the three Gravitational Wave Transient Catalogs, respectively [14,22,30].

A more fruitful way of gleaning spin information has been via the construction of specific projections of the spins parallel and perpendicular to the orbital angular momentum at a reference frequency, usually 20 Hz. These projections reduce the seven-dimensional intrinsic parameter space (mass ratio and six spin components) to three dimensions, which was shown to essentially capture the phenomenology of precessing waveforms [45]. The parallel scalar has come to be known as the effective spin parameter χ_{eff} [46,47]. Though the term effective (or reduced or PN) spin has been used for similar scalars [48,49], we employ the phenomenological definition given in our Eq. (8) [45,50] which has become standard. It was shown in Ref. [47] that χ_{eff} is a conserved quantity up to 1.5PN, i.e., neglecting spin-spin and higher-order interactions. The magnitude of χ_{eff} changes very little even with the inclusion of these

interactions. As such, it has become a very useful quantity in parameter estimation, especially since it emerges in the PN series for the waveform phase at a more dominant order than all other parameters except for the chirp mass and the symmetric mass ratio [39,40,51]. In fact, in the strong field regime, the contribution of the χ_{eff} term is large enough to cause a well-known partial degeneracy between the symmetric mass ratio and the parallel component of the total spin [39,40,50–52].

χ_{eff} also provides a way to gather information about the properties of the binary black hole (BBH) population since isotropically distributed spins would result in a normal distribution for χ_{eff} centered on zero. Several studies have already looked at this over the ensemble of the detected BBHs and more or less agree that the distribution is somewhat asymmetric and peaks at values slightly above zero [53–59].

The perpendicular scalar is known as the effective precession parameter (or spin) χ_p [45,60] and is given below in Eq. (9). It was shown by Refs. [45,60] that the dominant effects of precession on the gravitational wave (GW) waveform can be characterized by mapping the four perpendicular components of the binary’s spin vectors to just one parameter, χ_p , assigned to be the sole perpendicular component of the spin of the larger (primary) component of the binary. Since χ_p provides information about the perpendicular component of the spins, specifically, the perpendicular component of the primary’s spin, it is taken as an indication of spin precession.

The inference of χ_p is more challenging than that of χ_{eff} because, for instance, in the limit that the orbital angular momentum is much larger than the total spin, the perpendicular components can be considered to be 0.5PN order higher than the parallel components [61]. This can also be understood in terms of the respective contributions of the parallel and the perpendicular components to the waveform phase. Comparing the contribution of the χ_{eff} -term to the overall phase with that of the accumulated precession phase [cf. Eq. (45) of Ref. [37]] for a canonical neutron star binary inspiraling from 10 Hz, we see that the former is roughly 2 orders of magnitude larger. Accordingly, at moderate signal-to-noise ratios (SNRs), it is not possible to obtain a clear measurement of the individual spins [60]. Moreover, the inclination of the orbit also matters; it is harder to detect precession in binaries where the line of sight is aligned with the total angular momentum as the precession-induced modulations are minimized [62]. Therefore, it is not surprising that there have thus far been only three GW events for which there is evidence for precession: GW190412 [20], GW190521 [21,26,63], and GW200129_065458 [22,64–66].

Discerning precession in compact binary inspirals (and mergers) is important for breaking parameter degeneracies and understanding binary formation scenarios. To this end, accurate modeling of spin effects has become crucial in

GW astronomy. Though nonprecessing, i.e., aligned-spin, quasicircular waveform models such as those given by Refs. [67–74] have matured to the faithfulness level of $\lesssim 10^{-3}$ with respect to numerical relativity waveforms,¹ the precessing models are approximately one half to an order of magnitude worse in faithfulness and much more sensitive to modeling systematics. As the detectors’ sensitivities improve steadily through O4, O5, and beyond, systematic errors in parameter estimation due to mismodeling the effects of precession will dominate over the statistical error, especially for signals with large SNR [39,75]. In fact, recent research has shown that even the current accuracy of the NR waveforms may not be enough for bias-free parameter estimation at the projected sensitivities of the third-generation ground-based interferometers [76]. For this reason, the waveform modeling community has been endeavoring toward building ever more faithful precessing models. There are now several precessing quasispherical waveform models borne out of various waveform “families” and all have achieved $\lesssim 10^{-2}$ faithfulness. The state of the art among these is NRSUR7DQ4 [77], the latest precessing member of the numerical relativity surrogate family [68,78–82] with an NR faithfulness of $\sim 10^{-4}$. It was recently employed in a reanalysis of most O3 events [83].

Another set of models is the NR-informed effective-one-body (EOB) family [46,84–88]. This contains two similar, but distinct subfamilies: SEOBNR [67,73,89–101] and TEOBRESUMS [74,102–116]. A third major family is the phenomenological inspiral-merger-ringdown, IMRPHEMOM waveforms [45,60,69–71,117–133], which have mostly been built in the frequency domain. This has enabled faster parameter estimation runs compared with the time-domain EOB models. However, there are now also a few time-domain IMRPHEMOM models [69,131].

Specific waveform models from each family are usually referred to as approximants. The state-of-the-art approximant from each of the families listed in the previous paragraph are, respectively, SEOBNRV5PHM [98,100,101,134] (recent upgrade from SEOBNRV4PHM [95–97]), TEOBRESUMS² [107,110], IMRPHEMOMTPHM [131] (time domain), and IMRPHEMOMXPHM [128] (frequency domain). Note, we do not employ here the more recent IMRPHEMOMXCODE [133] or IMRPHEMOMXO4A [135], or the upgraded IMRPHEMOMXPHM [136] though we do include a brief comparison involving these two models in Appendix C. Essentially, these models are “too new” as the bulk of our work was already completed by the time they appeared.

When a new waveform model is complete, it may undergo a review where it is extensively compared with numerical

¹We quote the approximate median value for a specific mismatch with respect to numerical relativity simulations over a large sample.

²Nonprecessing and precessing approximants in this family are all called by the same name.

relativity simulations and with various other approximants as well as undergoing parameter estimation tests. Such studies are also detailed in the articles introducing the specific models (see, e.g., the works cited above). Recently, the LIGO-Virgo-KAGRA Collaboration (LVK) also conducted a detailed study of the faithfulness of the waveforms generated by the models NRSUR7DQ4, SEOBNRV5PHM, TEOBRESUMS, IMRPHENOMTPHM, IMRPHENOMXPHM, and IMRPHENOMXO4A using a set of $\gtrsim 1500$ NR simulations, but this is not publicly available. However, some of this work has been documented in Ref. [98] (also see Ref. [135] for a similar study using BAM waveforms [137–139]). A separate study focusing on the parameter estimation performance of the precessing models was conducted in Ref. [140]. This work compared the system parameters inferred by NRSUR7DQ4, IMRPHENOMTPHM, IMRPHENOMXPHM, and SEOBNRV4PHM for an ensemble of nearly 60 “pure” BBH O3 events.

Here, we undertake a more systematic survey based on simulated data. First, in Sec. IV, using the NR surrogate NRSUR7DQ4 (henceforth NRSUR) as a proxy for NR, we compute the unfaithfulnesses of {SEOBNRV5PHM, TEOBRESUMS, IMRPHENOMTPHM, IMRPHENOMXPHM} to it for a discretely spaced and a random-uniformly-filled set of intrinsic parameters with the binary mass fixed to both a light and a heavy value, and inclination fixed to 0 and $\pi/2$. We focus solely on the $\ell = 2$ waveform strain faithfulness throughout our work. For the discrete parameter set, we document in detail the deterioration of faithfulness with increasing mass asymmetry, which is well known (see e.g., in Refs. [98,131,132]), but we also uncover multimodalities in the unfaithfulness distributions coming from strongly precessing cases. We further reveal a correlation involving the unfaithfulness of the coprecessing ($2, \pm 2$) multipoles. We additionally contrast the unfaithfulness results from the discrete set with those of a uniformly filled parameter set. Using this, we show that the latter type of parameter set may result in the overestimation of model faithfulness.

We then move on to direct comparisons with NR waveforms from the sxs catalog [141], choosing simulations that were not used in the calibration of the models of interest here. We employ a set of 317 sxs simulations with a mass ratio grouping similar to the discrete set above so that we can reapply the same analyses which result in our reaching very similar conclusions. We further employ a smaller set consisting of 23 sxs waveforms containing more than 100 GW cycles each in order to briefly assess the models’ faithfulness for longer inspirals.

Finally, to complete our survey, we perform zero-noise injections of a moderately and a strongly precessing numerical relativity waveform and recover their parameters with SEOBNRV5PHM, IMRPHENOMTPHM, and IMRPHENOMXPHM. As we explain in detail in Sec. VI, we are unable to present Parameter Estimation (PE) results for TEOBRESUMS for these injections.

Our faithfulness survey provides us with enough data to construct fits for unfaithfulness over the intrinsic parameter space. These fits can be used to generate a weighted categorical prior to inform model choice in a joint Bayesian analysis [142]. We reduce the dimensionality of the parameter space by employing the parallel and perpendicular effective spin projections for our fits. We then construct three-dimensional fits to the logarithm of the unfaithfulness as functions of the mass ratio and the spin projections. These fits can subsequently be employed to assign weights to each waveform model in various regions of the parameter space. Our results shown in Appendix A are preliminary as we intend to pursue this line of research elsewhere.

The remaining sections are organized as follows. In Sec. II, we briefly review precession dynamics and precessing waveform construction, and introduce the parallel and perpendicular spin projections. Section III introduces the various metrics which we employ to assess waveform faithfulness. We summarize our work in Sec. VII and conclude with a retrospective discussion in Sec. VIII. Appendices B and C contain additional model comparisons. Finally in Appendix D, we present timing benchmarks for the main models that we consider here. We provide our data and the notebooks/codes/scripts used to obtain it in a GIT repository [143].

We work with geometrized units where $G = c = 1$. m_1 and m_2 denote the masses of the primary and secondary components of the compact binary system, with $m_1 \geq m_2$. Accordingly, we define small and large mass ratios as $q = m_2/m_1 \leq 1$ and $Q = 1/q \geq 1$, the symmetric mass ratio $\eta = q/(1+q)^2$, and the total mass $M = m_1 + m_2$. \mathbf{S}_1 and \mathbf{S}_2 denote the spin vectors of each binary component with the respective dimensionless spin vectors given by $\chi_i = \mathbf{S}_i/m_i^2$ and $\chi_i := |\chi_i|$, $i = 1, 2$. Unless otherwise noted, we set $M = 1$. Overdots denote time derivatives whereas hats denote unit vectors. Throughout this article, we use the terms aligned and parallel interchangeably, as well as perpendicular/in plane/planar where the former direction is along the orbital angular momentum vector and the latter is in the orbital plane.

II. REVIEW OF PRECESSING WAVEFORM CONSTRUCTION

For this article, we focus only on quasispherical binary systems as most binaries are expected to have circularized by the time they enter the LIGO-Virgo-KAGRA band [144] which is supported by the GW data so far [145–152] (see Refs. [153,154] for a few exceptions). However, there are binary formation scenarios in which a moderate amount of eccentricity survives beyond the dechertz regime [155–168]. Therefore, it is important to have template banks of sufficiently faithful eccentric and precessing waveforms in the near future.

In general relativity, GWs have only two propagating degrees of freedom, + and \times polarizations, which can be obtained from the following multipole sum in terms of spin weight = -2 harmonics

$$h_+ - ih_\times = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{\ell} {}_{-2}Y^{\ell m}(\vartheta_{\text{LN},0}, \varphi_{\text{ref}}) h_{\ell m}(t), \quad (1)$$

where $\vartheta_{\text{LN},0} := \cos^{-1}(\hat{\mathbf{L}}_{\text{N},0} \cdot \hat{\mathbf{N}})$ is the orbital inclination with $\hat{\mathbf{N}}$ being the line-of-sight vector from the binary's center of mass to the observer. $\varphi_{\text{ref}} = \pi/2 - \varphi_c$ is a constant reference phase [169], where φ_c denotes the phase at coalescence.

The polarizations $h_{+,\times}$ couple with the detectors' antenna patterns on Earth. For an L-shaped interferometer, the GW strain in the time domain is given by

$$h(t) = F_+(\theta_s, \phi_s, \psi_s) h_+(t, \vartheta_{\text{LN},0}, \varphi_{\text{ref}}) + F_\times(\theta_s, \phi_s, \psi_s) h_\times(t, \vartheta_{\text{LN},0}, \varphi_{\text{ref}}), \quad (2)$$

where θ_s, ϕ_s are the source sky location angles, ψ_s is the source polarization with respect to the detector, and $F_{+,\times}$ are the detector antenna pattern functions that can be found in, e.g., Ref. [170]. The GW strain at the detector (henceforth just the strain) given by Eq. (2) is the quantity which we use in determining the faithfulness of the EOB and the phenomenological approximants via a certain sky-maximized mismatch defined in Sec. IV.

It was shown by Refs. [49,62,171] that precessing waveform multipoles, $h_{\ell m}$, can be built to a good approximation from the Euler rotation of aligned-spin (AS) multipoles via the following expression

$$h_{\ell m} = \sum_{m'=-\ell}^{\ell} D_{m',m}^{(\ell)*}(-\gamma, -\beta, -\alpha) h_{\ell m'}^{\text{AS}}, \quad (3)$$

where $D_{m',m}^{(\ell)}$ are Wigner's D matrices [172,173]. α, β , and γ are the Euler angles of the frame rotation with α, β being the spherical angles of the Newtonian orbital angular momentum vector \mathbf{L}_N with respect to a chosen frame, e.g., $\mathbf{L}_{N,0} := \mathbf{L}_N(f_0)$ where f_0 is a reference frequency. The third rotation by γ uniquely fixes the frame to the so-called minimal-rotation frame [171].

The Euler angles α, β are obtained from the time evolution of $\mathbf{L}_N = L_N \hat{\mathbf{L}}_N$ governed by the PN precession equations. At next-to-leading order (NLO), the orbit-averaged evolution equations can be written in the following form

$$\dot{\mathbf{S}}_i = \boldsymbol{\Omega}_i \times \mathbf{S}_i, \quad (4a)$$

$$\dot{\hat{\mathbf{L}}}_N = \boldsymbol{\Omega}_{\text{NLO}} \times \hat{\mathbf{L}}_N, \quad (4b)$$

with $i = 1, 2$ denoting the primary and the secondary, respectively. The precession frequencies are given by

$$\frac{\boldsymbol{\Omega}_i}{v^5} = \eta \left(2 + \frac{3}{2} q_i \right) \hat{\mathbf{L}}_N + \frac{v}{2} \{ \mathbf{S}_j - 3[(q_i \mathbf{S}_i + \mathbf{S}_j) \cdot \hat{\mathbf{L}}_N] \hat{\mathbf{L}}_N \}, \quad (5a)$$

$$\boldsymbol{\Omega}_{\text{NLO}} = -\frac{v}{\eta} (\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2), \quad (5b)$$

where v is the relative speed of the binary components in the center-of-mass frame, $q_1 = 1/q, q_2 = q$, and $j = 3 - i$. Radiation reaction is incorporated through the decay of the magnitude of the angular momentum, \dot{L}_N , while the direction of the total angular momentum vector $\mathbf{J} = \mathbf{L}_N + \mathbf{S}_1 + \mathbf{S}_2$ is kept fixed at a desirable PN order. It is standard to rewrite \dot{L}_N as $\dot{v}(v)$ or $\dot{\omega}(\omega)$ using $L_N = \eta/v = \eta\omega^{-1/3}$ where $\omega = v^3$ is the orbital frequency. Further details can be found in, e.g., Refs. [37,38,100,107,174,175].

The coupled system of ODEs consisting of the decay of L_N and Eqs. (4a) and (4b) can be numerically solved straightforwardly. Analytic solutions have also been found at this PN order via the multiscale approach [175–178]. PN information exists up to 3.5PN in the radiation reaction sector [179–181] and next⁵LO in the precession dynamics [100]. Once the solution $\mathbf{L}_N(t)$ is known, $\alpha(t)$ and $\beta(t)$ can be computed immediately, and, subsequently, γ from [171]

$$\dot{\gamma} = \pm \dot{\alpha} \cos \beta, \quad (6)$$

where there is a sign freedom in the right-hand side.

Note that Eq. (3) is an approximation for the true precessing multipoles, $h_{\ell m}^{\text{prec}}$, which can be extracted from NR simulations for example. One can then obtain the coprecessing multipoles via inverse of the transformation in Eq. (3)

$$h_{\ell m}^{\text{coprec}} = \sum_{m'=-\ell}^{\ell} D_{m',m}^{(\ell)}(\alpha, \beta, \gamma) h_{\ell m'}. \quad (7)$$

Though $h_{\ell m}^{\text{coprec}} \neq h_{\ell m}^{\text{AS}}$, the AS multipoles are, in general, a good approximation for the coprecessing multipoles [182] as we will show in Sec. IV A 3.

The AS multipoles satisfy the relation $h_{\ell, -m}^{\text{AS}} = (-1)^{\ell} h_{\ell m}^{\text{AS}*}$, which does not hold for $h_{\ell m}^{\text{coprec}}$ due to the asymmetric emission of GWs above and below the orbital plane leading to the ‘‘bobbing’’ of the binary and eventually resulting in the well-known kick of the final black hole [182–185]. Although the difference in the $\pm m$ multipoles may be small, it seems to be non-negligible for unbiased parameter estimation of precessing binaries even at moderate SNRs [185], and most definitely so at high SNRs [186]. The only model used in this work that does not

neglect this multipole asymmetry is NRSUR7DQ4 [77], though we note that the newest phenomenological approximant PHENOMXO4a also contains this feature for the dominant, quadrupolar multipole [135,187].

Another approximation made above is the use of orbit-averaged ODEs to describe the precession dynamics which removes the nutation of the spins from the evolution of the binary.³ Though wide nutation angles may leave a unique signature on future GW events [188], the work of Ref. [189] has found no strong evidence of nutation in any detected event thus far.

Returning to the binary with the orbital and spin angular momenta vectors $\{\mathbf{L}_N, \mathbf{S}_1, \mathbf{S}_2\}$, let $\theta_i = \cos^{-1}(\hat{\mathbf{L}}_N \cdot \hat{\mathbf{S}}_i)$ denote the polar (tilt) and ϕ_i the azimuthal angle of each spin vector with $i = 1, 2$. The components of the spins parallel and perpendicular to $\hat{\mathbf{L}}_N$ are given by $\mathbf{S}_{i,\parallel} = (\mathbf{S}_i \cdot \hat{\mathbf{L}}_N)\hat{\mathbf{L}}_N = m_i^2 \chi_i \cos \theta_i \hat{\mathbf{L}}_N$ and $\mathbf{S}_{i,\perp} = \mathbf{S}_i - \mathbf{S}_{i,\parallel}$ with magnitude $m_i^2 \chi_i \sin \theta_i$.

We have already discussed the most commonly employed parallel and perpendicular scalars, χ_{eff} and χ_p , in Sec. I. χ_{eff} is given by [46]

$$\chi_{\text{eff}} = \frac{1}{1+q}(\chi_1 \cos \theta_1 + q\chi_2 \cos \theta_2) \quad (8)$$

and its perpendicular counterpart by [45,60]

$$\chi_p = \max \left(\chi_1 \sin \theta_1, q \frac{4q+3}{4+3q} \chi_2 \sin \theta_2 \right). \quad (9)$$

The Kerr spin limit $\chi_i \leq 1$ imposes the condition $0 \leq \chi_p \leq 1$ where $\chi_p = 0$ corresponds to a nonprecessing (aligned spin or spinless) and $\chi_p = 1$ to a maximally precessing binary. There is an elegant discussion in Ref. [190] on the physical meaning of this definition which leads to a generalized version of this parameter

$$\chi_p^{\text{Gen}} = \left[(\chi_1 \sin \theta_1)^2 + \left(q \frac{4q+3}{4+3q} \chi_2 \sin \theta_2 \right)^2 + 2q \frac{4q+3}{4+3q} \chi_1 \chi_2 \sin \theta_1 \sin \theta_2 \cos(\Delta\phi) \right]^{1/2}, \quad (10)$$

where

$$\cos(\Delta\phi) = (\hat{\mathbf{S}}_1 \times \hat{\mathbf{L}}_N) \cdot (\hat{\mathbf{S}}_2 \times \hat{\mathbf{L}}_N) \quad (11)$$

which is the span of the planar angle between \mathbf{S}_1 and \mathbf{S}_2 . χ_p^{Gen} is in fact equal to $|\hat{\mathbf{L}}_N|/\Omega_1$ [190] and a comparison of Eq. (10) with Eq. (9) reveals that χ_p^{Gen} can exceed 1 ($\chi_p^{\text{Gen}} \leq 2$) which can only be achieved by systems in which

³For non-orbit-averaged versions, see Eqs. (2.8) and (2.10) of Ref. [47].

both spins are large and mostly planar [190]. This fact has already been used for a simulated study of O4 events showing that for moderate to high SNRs, double-spin precession can be inferred [191].

There are several other recently proposed perpendicular scalars. For example, Ref. [107] introduced the following perpendicular scalar

$$\chi_{\perp,J} = \frac{|\mathbf{S}_{1,\perp} + \mathbf{S}_{2,\perp}|}{M^2}. \quad (12)$$

Reference [132] introduced a modification to χ_p for improved single-spin mapping in the merger-ringdown regime for binaries where $Q \lesssim 1.5$. Reference [192] replaced χ_p with a two-dimensional vector in the phenomenological waveform mapping to improve the faithfulness of the precessing $(2, \pm 1), (3, \pm 3), (4, \pm 4)$ multipoles. Here, we concern ourselves only with $\chi_p, \chi_p^{\text{Gen}}$ and $\chi_{\perp,J}$ all of which we collectively refer to as the χ_{\perp} 's.

As can be seen from the equations defining the χ_{\perp} 's, each one depends on quantities that evolve in time (in other words, in frequency). Traditionally, the values for these quantities are usually quoted at a reference frequency of 20 Hz, but this choice is somewhat arbitrary. Alternative reference times (or frequencies) have been suggested such as $t_{p-100} := t_{\text{peak}} - 100M$ and $t_{\text{ISCO}} := t(f = 6^{-3/2}/(M\pi))$, where t_{peak} denotes the time at which the coprecessing $(2, 2)$ multipole amplitude peaks, and t_{ISCO} is the time at which the GW frequency equals twice the Schwarzschild innermost stable circular orbit (ISCO) frequency. References [193,194] have recently adopted yet another alternative: the $t \rightarrow -\infty$ ($f \rightarrow 0$ Hz) limit where the spin tilt angles θ_1, θ_2 can be unambiguously obtained. However, the angle $\Delta\phi$ cannot be uniquely determined which means that we cannot compute χ_p^{Gen} or $\chi_{\perp,J}$ at past infinity. One can also work with averaged quantities instead. For example, Ref. [190] introduces a precession-cycle-averaged version of χ_p which they further extend to precession- and rms-averaged versions of χ_p^{Gen} in Ref. [175]. Unless otherwise noted, we quote the values of the various spin scalars at the initial time (frequency) $t_0(f_0)$.

III. QUANTIFYING WAVEFORM FAITHFULNESS

The faithfulness of a given template waveform strain $h'(t)$ to a target waveform $h_0(t)$ is measured in terms of the waveform match given by

$$\mathcal{M} = \max_{t_c, \varphi_c} \frac{\langle h_0 | h' \rangle}{\sqrt{\langle h_0 | h_0 \rangle \langle h' | h' \rangle}}. \quad (13)$$

The match is maximized over constant time and phase shifts which we take to be the time and phase shifts at coalescence t_c, φ_c . The angular brackets denote the noise-weighted inner product

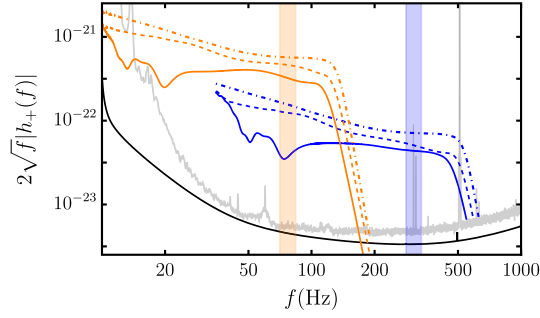


FIG. 1. Typical waveform amplitudes in the frequency domain from the sample of BBHs generated for the survey in Sec. IV A. In particular, we display in blue (orange) a trio of light (heavy) precessing binaries starting from $f_0 = 37.5$ Hz (12 Hz) with $Q = 1, \bar{1}, 2, 4$ (dot dashed, dotted, solid) at a fiducial luminosity distance of 500 Mpc and initial inclination of $\vartheta_{\text{LN},0} = 0$ shown against both Advanced LIGO design sensitivity [196] (black solid curve) and LIGO Livingston detector strain sensitivity during O3 [197] (faint gray curve). We used the approximant IMRPHENOMXPHEM to generate these waveforms. For each case, we have $\max(\chi_p) = 0.8$. The orange and blue vertical bands mark the regions of f_{peak} for each triplet of waveforms. Since f_{peak} varies for each set of parameters, we opted to show bands here instead of six additional vertical lines marking individual f_{peak} .

$$\langle h_0 | h' \rangle := 4\Re \int_{f_i}^{f_f} \frac{\tilde{h}_0(f) \tilde{h}'^*(f)}{S_n(f)} df, \quad (14)$$

where $\tilde{h}(f)$ are the Fourier transforms of the time-domain GW strains. The inner product is weighted by the one-sided power spectral density (PSD) $S_n(f)$ of the detector noise for which we use the Advanced LIGO [195] zero-detuned high-power design sensitivity [196]. The integration is performed from the initial frequency f_i to the final frequency f_f which we set equal to 1024 Hz. As can be seen from Fig. 1, this choice of final frequency is large enough to encompass the merger-ringdown portions of the waveforms for BBHs with $M = 37.5M_\odot, 150M_\odot$ which will be our chosen values for much of this article. To mediate the effects caused by finite signal duration, we choose f_i such that there are at least three waveform cycles between it and f_0 , the initial coprecessing (2, 2)-multipole frequency. We set the luminosity distance of the sources to the fiducial value of $d_L = 500$ Mpc.

If the system of interest is precessing then the detector antenna patterns of Eq. (2) become time dependent. Consequently, the waveform strain at the detector becomes dependent on constant and time-varying extrinsic parameters (and of course the intrinsic parameters) which we can write as

$$h(t) = F_+(\theta_s, \phi_s, \psi_s(t))h_+(\vartheta_{\text{LN},0}, \varphi_{\text{ref}}, t_{\text{ref}}) + F_\times(\theta_s, \phi_s, \psi_s(t))h_\times(\vartheta_{\text{LN},0}, \varphi_{\text{ref}}, t_{\text{ref}}). \quad (15)$$

By introducing an effective amplitude \mathcal{A} ,

$$\mathcal{A}(\theta_s, \phi_s) = \sqrt{F_+^2(\theta_s, \phi_s, \psi_s) + F_\times^2(\theta_s, \phi_s, \psi_s)} \quad (16)$$

and an effective polarizability κ via

$$e^{i\kappa(\theta_s, \phi_s, \psi_s)} = [F_+(\theta_s, \phi_s, \psi_s) + iF_\times(\theta_s, \phi_s, \psi_s)] / \mathcal{A}(\theta_s, \phi_s), \quad (17)$$

we can rewrite Eq. (15) as

$$h(t) = \mathcal{A}(\theta_s, \phi_s) \{ \cos[\kappa(\theta_s, \phi_s, \psi_s)]h_+(\vartheta_{\text{LN},0}, \varphi_{\text{ref}}, t_{\text{ref}}) + \sin[\kappa(\theta_s, \phi_s, \psi_s)]h_\times(\vartheta_{\text{LN},0}, \varphi_{\text{ref}}, t_{\text{ref}}) \}, \quad (18)$$

where we suppressed the t dependence in the right-hand side.

We can now define the sky-maximized (optimized) faithfulness (match) between the target strain and the waveform template as

$$\mathcal{M}_{\text{opt}} = \max_{t'_{\text{ref}}, \varphi'_{\text{ref}}, \kappa', \phi'} \frac{\langle h_0 | h' \rangle}{\sqrt{\langle h_0 | h_0 \rangle \langle h' | h' \rangle}}, \quad (19)$$

where $t'_{\text{ref}}, \varphi'_{\text{ref}}, \kappa'$ are template parameters to be optimized over. The details of the optimization can be found in Refs. [110,198]. In particular, the κ' optimization is performed analytically, while t'_{ref} is maximized via the inverse fast Fourier transform. The maximization over φ'_{ref} is performed numerically using a dual annealing algorithm [110,128]. ϕ' represents the final degree of freedom to be maximized over, i.e., the freedom to shift ϕ_1, ϕ_2 by a constant amount which leaves $\Delta\phi$ unchanged. We perform this maximization using another dual annealing algorithm [110]. It is evident from Eq. (18) that the match is also a function of the inclination angle $\vartheta_{\text{LN},0}$, but since we fix the inclinations in our comparisons at a given reference frequency, we do not need to optimize over them.

Note that \mathcal{M}_{opt} is a function of φ_{ref} and κ so, as a final step, we compute its average over an evenly spaced grid for $\{\varphi_{\text{ref}}, \kappa\} \in [0, 2\pi) \times [0, \pi/2)$ with $6 \times 7 = 42$ elements to obtain

$$\mathcal{M}_{\text{opt,av}} := \frac{1}{42} \sum_{i=1}^{42} \mathcal{M}_{\text{opt}}(\varphi_{\text{ref},i}, \kappa_i). \quad (20)$$

This is done to marginalize over any dependence of the match on the sky position and obtain values which depend exclusively on the intrinsic parameters of the source. This quantity is similar to the sky-and-polarization-averaged faithfulness given by Eq. (35) of Ref. [98], but we do not average over the inclination.

For the remainder of this article, we employ the sky-averaged, optimized waveform mismatch

$$\bar{\mathcal{M}}_{\text{opt}} := 1 - \mathcal{M}_{\text{opt,av}} \quad (21)$$

as our faithfulness gauge. The optimized mismatches that we quote henceforth will always be this average which we may occasionally refer to as the ‘‘full mismatch.’’ As a check, we also store the minimum and the maximum values of $\bar{\mathcal{M}}_{\text{opt}}$ over the grid as well as the standard deviation all which we present in the data provided in our GIT repository.

To further disentangle possible causes of waveform mismodeling, we additionally compute mismatches where we truncate the mismatch integral before the transition to plunge, specifically at $f_f = 0.6f_{\text{peak}}$ with $f_{\text{peak}} := f(t = t_{\text{peak}})$ denoting the GW frequency at the maximum (peak) amplitude of the coprocessing $\ell = 2$ strain of NRSUR7DQ4, i.e., the maximum of

$$\mathcal{A}_{\text{coprec}}(t) := \sqrt{\sum_{m=-2}^{m=2} h_{2m}^{\text{coprec}}(t)} \quad (22)$$

with $t_0 < t_{\text{peak}}$ [169]. We have checked that $0.6f_{\text{peak}} \geq f_{\text{MECO}}$, the minimum energy circular orbit (MECO) frequency [70]. Accordingly, we present results for not only $\bar{\mathcal{M}}_{\text{opt}}$, but also for the merger-ringdown excluded (inspiral-only) mismatch

$$\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}} := \bar{\mathcal{M}}_{\text{opt}}(f_f = 0.6f_{\text{peak}}). \quad (23)$$

IV. FAITHFULNESS SURVEY I: COMPARISONS WITH NRSUR7DQ4

We begin our survey with Sec. IV A where we assess the faithfulness of SEOB, TEOB, TPHM, and XPHM to NRSUR7DQ4 using a discrete grid for the intrinsic parameters at zero inclination, by which we mean $\vartheta_{\text{LN},0} := \vartheta_{\text{LN}}(f_0) = 0$. Since the direction of $\hat{\mathbf{L}}_{\text{N}}$ evolves in time, so does ϑ_{LN} . Thus, the best we can do is specify its value at some reference frequency. In principle, we could have employed the (approximately) fixed inclination with respect to the total angular momentum, θ_{JN} . However, the ϕ' optimization of Eq. (19) becomes nontrivial in this case.

We also include in Sec. IV A a comparison in the extrapolation region of NRSUR7DQ4, where the model is stated to be robust [77]. In Sec. IV B, we consider the same discrete set at $\vartheta_{\text{LN},0} = \pi/2$ thus changing the multipole content of the strain in Eq. (1). We perform a similar analysis in Sec. IV C over a random-uniformly-filled intrinsic parameter space and compare our findings with those of Secs. IV A and IV B. We should add that the most recent version of SEOB (v5) was released near the completion of this work so we had initially conducted our

TABLE I. Relevant parameters for the binary systems which we use in our faithfulness survey in Sec. IV A. f_0 denotes the initial frequency from which we start the binary evolution. We fix $\phi_1 = 0$ at f_0 without loss of generality.

Total mass	$M \in \{37.5M_{\odot}, 150M_{\odot}\}$
Mass ratio	$Q \in \{1.\bar{1}, 2, 4, 6\}$
Initial frequency	$f_0 = \begin{cases} \{30, 31, 35, 38.5\} \text{ Hz} & \text{for } M = 37.5M_{\odot}, \\ \{7.5, 8, 9, 10\} \text{ Hz} & \text{for } M = 150M_{\odot} \end{cases}$
Spin magnitudes	$\chi_1 = \chi_2 = 0.8$
Tilt angles	$\cos \theta_{1,2} = \{-\frac{\sqrt{3}}{2} + k\frac{\sqrt{3}}{4}\} \cup \{-1\}, k = 0, \dots, 4$ if $\theta_1 = \pi, \theta_2 \neq \pi$
Azimuthal angles	$\phi_1 = 0, \phi_2 = k\pi/4, k = 1, \dots, 8$

survey on SEOBNRV4PHM. As a useful sidenote, we present a brief comparison of v5 and v4 in Appendix B.

A. Discrete parameter set with zero inclination

As both LIGO and Virgo noise curves are frequency, and therefore mass, dependent, we consider a set of light and heavy binaries for our survey with $M = 37.5M_{\odot}$ and $150M_{\odot}$. We divide each set into four equal-size subsets separated by mass ratio values of $Q = 1.\bar{1}, 2, 4, 6$. We have chosen $Q = 1.\bar{1}$ ($q = 0.9$) to break the $Q = 1$ symmetry where the nonprecessing (2, 1) multipole equals zero, i.e., $h_{21}^{\text{AS}} = 0$. The subset with $Q = 6$ is in the so-called extrapolation region of NRSUR7DQ4 so we delegate the comparison pertaining to it to Sec. IV A 4.

A lighter binary mostly accumulates SNR during its inspiral whereas a heavy enough binary could have as much SNR accumulated during the merger-ringdown stage as the inspiral assuming signals enter the detector band at $\gtrsim 20$ Hz. We illustrate this in Fig. 1 with three light and three heavy BBH waveforms with $Q = 1.\bar{1}, 2, 4$ and the same spin vectors. The merger-ringdown SNR for the light binaries ranges from a quarter to \lesssim half of its inspiral counterpart (computed from 38 Hz; see f_0 's in Table I) while the same SNR for the heavy systems ranges from 70% to 90% of its inspiral counterpart (from 12 Hz). Therefore, by considering light and heavy systems, we are effectively dividing our survey sample into two halves whereby merger-ringdown modeling is much more important for waveform faithfulness in one half (heavy BBHs) than the other (light BBHs).

Any comparison with NRSUR7DQ4 is ultimately limited by the fact that the surrogate waveforms have a maximum inspiral time length of $4300M$. For this reason, we have chosen $f_0 = \{30, 31, 35\}$ Hz for the $Q = 1.\bar{1}, 2, 4$ subsets of the $M = 37.5M_{\odot}$ set and $f_0 = \{7.5, 8, 9\}$ Hz for the same subsets of the $M = 150M_{\odot}$ set. These yield between 27 and 47 GW cycles for light BBHs and between 25 and 45 GW cycles for the heavy BBHs.

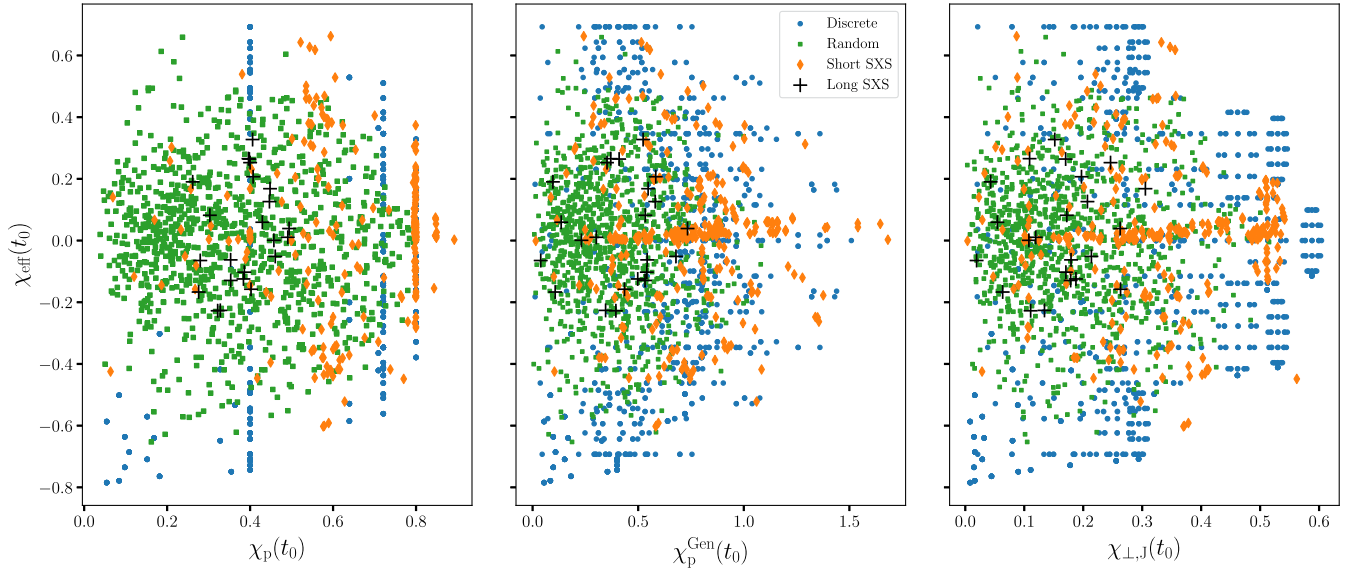


FIG. 2. The coverage of the spin space for all our parameter sets shown using the parallel and perpendicular effective spin scalars computed at the initial time t_0 . In each panel, $\chi_{\text{eff}}(t_0)$ of Eq. (8) is plotted in the vertical axis. From left to right, $\chi_p(t_0)$ [Eq. (9)], $\chi_p^{\text{Gen}}(t_0)$ [Eq. (10)], and $\chi_{\perp,J}(t_0)$ [Eq. (12)] are plotted in the horizontal axes, respectively. The blue disks correspond to the values of these quantities coming from our discrete parameter set of Sec. IV A (see Table I). The green squares, orange diamonds, and the black crosses represent the same quantities obtained from our random-uniformly-filled set (Sec. IV C), the short and the long SXS sets (Secs. V A and V B), respectively. As χ_p has no dependence on the azimuthal components of the spins, the parameters of the discrete and the short SXS sets yield degenerate values clustered at $\chi_p \approx 0.4, 0.72, 0.8$, whereas both χ_p^{Gen} and $\chi_{\perp,J}$ cover their respective ranges better since they are ϕ_i dependent.

Each Q subset consists of a grid of spin angles $\{\theta_1, \theta_2, \Delta\phi := \phi_2 - \phi_1\}$ where the tilt angles are evenly spaced in $\cos\theta_i$ from $-\sqrt{3}/2$ to $\sqrt{3}/2$ in steps of $\sqrt{3}/4$. We also included the grid points with $\cos\theta_i = -1$ ($i = 1$ or 2) with the intention to create a small subset of cases with near-transitional precession [37] to test the models' robustness. Specifically, the eight cases with $Q = 4, \theta_1 = \pi, \theta_2 = \pi/6$ yield $\hat{\mathbf{L}}_{\text{N}}(t_0) \cdot \mathbf{J}_{\text{N},0} \approx -0.02$ and $|\mathbf{J}_{\text{N},0}| \approx 0.025 \approx 0.033 |\mathbf{L}_{\text{N},0}|$ (in units of M^2). The addition of these extra grid points creates a somewhat skewed coverage of the parameter space as can be seen by the blue dots in Fig. 2.

$\Delta\phi$ runs from $\pi/4$ to 2π in steps of $\pi/4$. We fix the dimensionless Kerr spin parameter χ to 0.8 for all cases resulting in $\chi_p \leq 0.8, \chi_{\perp,J} < 0.55, \chi_p^{\text{Gen}} < 1.45$, and $|\chi_{\text{eff}}| < 0.7$. We summarize these parameter choices in Table I, and show the coverage of the dimensionless intrinsic parameter space in terms of the parallel and perpendicular spin projections in Fig. 2 as the blue disks. The figure also shows the same quantities plotted for the uniformly filled parameter set of Sec. IV C and the SXS sets of Sec. V.

The three-angle $\{\theta_1, \theta_2, \Delta\phi\}$ grid has in total $(6^2 - 1) \times 8 = 280$ elements. This then yields $280 \times 3 = 840$ separate binaries to consider per total mass separated by $Q = 1, \bar{1}, 2, 4$. Overall, we have $840 \times 2 = 1680$ systems for each of which we compare the waveforms generated by {SEOB, TEOB, TPHM, XPHM} to NRSUR via the averaged,

optimized mismatch $\bar{\mathcal{M}}_{\text{opt}}$ given by Eqs. (19) and (21) for which the last quantity to consider is the lower limit of the match integral, i.e., f_i . For the $M = 37.5M_{\odot}$ set, we pick $f_i = f_0 + 3$ Hz which captures approximately 25 to 42 GW cycles depending on the intrinsic parameters. This allows for plenty of SNR accumulation in the inspiral stage before transitioning to plunge. On the other hand, for the $M = 150M_{\odot}$ set, we deliberately fix $f_i = 11$ Hz leading to fewer inspiral cycles captured in the match integration, thus increasing the relative contribution of the merger-ringdown stage to the overall SNR.

In Fig. 3, we show the main result of this section where we plot the distributions of $\bar{\mathcal{M}}_{\text{opt}}$ between NRSUR7DQ4 and SEOB, TEOB, TPHM, XPHM. In the left (right) panel, we present the mismatches for the light (heavy), i.e., $M = 37.5M_{\odot}$ ($150M_{\odot}$), systems for all $Q = 1, \bar{1}, 2, 4$. An apparent feature in the figure is the superior faithfulness of SEOB exhibiting no cases of $\bar{\mathcal{M}}_{\text{opt}} > 0.035$ and having $\approx 22\%$ of the heavy-binary mismatches below 10^{-3} . We also note comparable performances among TEOB, TPHM, and XPHM. For the heavy systems, the distributions of $\bar{\mathcal{M}}_{\text{opt}}$ for these approximants become wider with more cases of $\bar{\mathcal{M}}_{\text{opt}} > 0.035$.

Perhaps the most curious feature in Fig. 3 is the little ‘‘bump’’ in the heavy XPHM histogram at $\bar{\mathcal{M}}_{\text{opt}} \sim 0.3$ with a similar smaller ‘‘island’’ for the light-binary XPHM

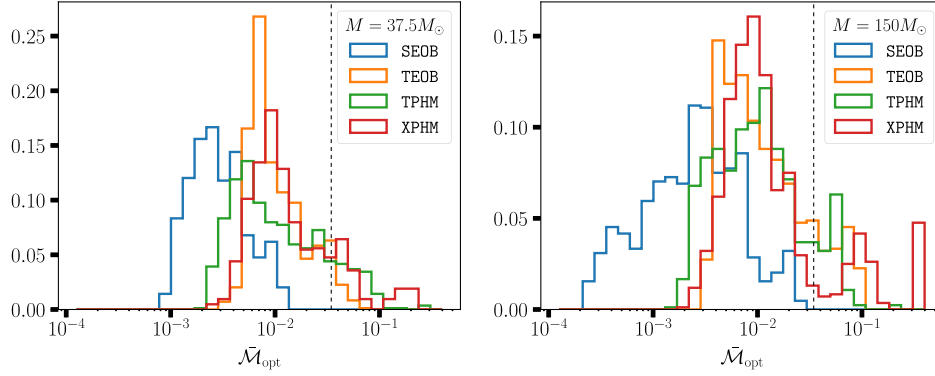


FIG. 3. The main result of this section: the distributions of the sky-optimized mismatch, $\bar{\mathcal{M}}_{\text{opt}}$, given by Eq. (21) between the models listed in the legend and NRSUR7DQ4 which we take as a proxy for numerical relativity. The parameters for the waveforms are given in Table I with the $Q = 6$ subset excluded from the figure and we consider only the $\ell = 2$ strain for the mismatches. The left (right) panels display the results for low (high) mass, $M = 37.5M_{\odot}$ ($M = 150M_{\odot}$), systems, respectively. SEOB (v5), being the most recent of the four models, shows improvement, especially for heavier systems. The dashed vertical line in both panels marks mismatch of $1 - 0.9^{1/3} \approx 0.035$ translating to an event loss rate of 10% [199,200].

mismatches at $0.1 \leq \bar{\mathcal{M}}_{\text{opt}} \leq 0.3$. For both light and heavy systems, we find that these are the same 40 cases which have $Q = 4, \theta_1 = \pi$ with the worst mismatches coming from the $\theta_2 > \pi/2$ subset. For these configurations, which have the most negative z projections of spins, i.e., $\chi_{\text{eff}} \lesssim -0.5$, the default multiple scale analysis (MSA) prescription for the precession dynamics of XPHM breaks down [178] and the model defaults to the single-spin, next-to-next-to-leading-order (NNLO) prescription for the spin dynamics with a 3PN approximation to $|\mathbf{L}|$. This is a known shortcoming of XPHM and is expected to be fixed in the next version of the model [136].

1. Effects of the merger-ringdown portion of GWs on faithfulness

The increased amplitude/phase disagreement between NR and waveform models in the plunge-merger-ringdown stages of the binary evolution is a well-known shortcoming, usually illustrated in terms of time-domain plots of the waveforms. For this reason, we investigated how the mismatches plotted in Fig. 3 change when neglecting the merger-ringdown portions of the waveforms. We quantify this in terms of the merger-ringdown truncated (i.e., inspiral only) mismatch given in Eq. (23). There are two reasons for which we *a priori* expect the merger-ringdown- (MR) truncated waveforms to agree better with NRSUR; i.e., we expect the $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ distributions to occupy lower values than their $\bar{\mathcal{M}}_{\text{opt}}$ counterparts. The first reason pertains to signal morphology: The inspiral is much smoother than the MR and therefore easier to optimize over in the match computation. The second reason has to do with the modeling: The MR stages are more difficult to model analytically, especially without input from NR simulations. As such, the waveform models employ NR-informed fits for the MR regime.

Our second expectation is a larger shift between the $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and $\bar{\mathcal{M}}_{\text{opt}}$ distributions for the $M = 150M_{\odot}$ sample since the MR part of the signal is much more important for these heavier binaries.

These expectations are confirmed in Fig. 4 where we present the distributions of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and $\bar{\mathcal{M}}_{\text{opt}}$ as (half) violin plots for the four approximants for both the light and the heavy BBH samples. We can clearly see that the distributions of the MR-excluded mismatches ($\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$), i.e., the darker-shaded half violins, are all shifted to lower values than the lighter-shaded half violins representing the distributions of $\bar{\mathcal{M}}_{\text{opt}}$. This downward shift is more pronounced for the heavier cases ($M = 150M_{\odot}$) with the median of the $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ distributions roughly half an order of magnitude lower than the $\bar{\mathcal{M}}_{\text{opt}}$ distributions (less so for TEOB). The shift is less prominent for the light BBHs ($M = 37.5M_{\odot}$) where the MR portion of the signal is much less important than the inspiral.

One can also discern a smaller secondary peak in most of the $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ distributions which comes from the mismatches of the $Q = 4$ subset. This peak is somewhat obscured in Fig. 4 where we plot the entire ensemble of $Q = 1, \bar{1}, 2, 4$ subsets. We investigate this in more detail in the next section where we separate the mismatch data by mass ratio.

One can further ask whether or not there are regions in the spin space where there is a larger gap between $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and $\bar{\mathcal{M}}_{\text{opt}}$ than other regions. To this end, we define the following ratio $\mathcal{R} := \bar{\mathcal{M}}_{\text{opt}} / \bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and look at how it is distributed in the $\{\chi_{\perp}\text{'s}, \chi_{\text{eff}}, Q\}$ space, where χ_{\perp} 's represents one element of the set $\{\chi_p, \chi_p^{\text{Gen}}, \chi_{\perp, j}\}$ at a time.

For the light BBHs, we observe $\mathcal{R} \lesssim \{2, 3.5, 6\}$ for $\{\text{SEOB}, \text{TEOB}, \text{TPHM}\}$ regardless of Q . For XPHM, $\mathcal{R} \lesssim \{4.6, 6, 9\}$ for $Q = \{1, \bar{1}, 2, 4\}$. Moreover, \mathcal{R} seems to increase monotonically for TEOB and TPHM as the spin

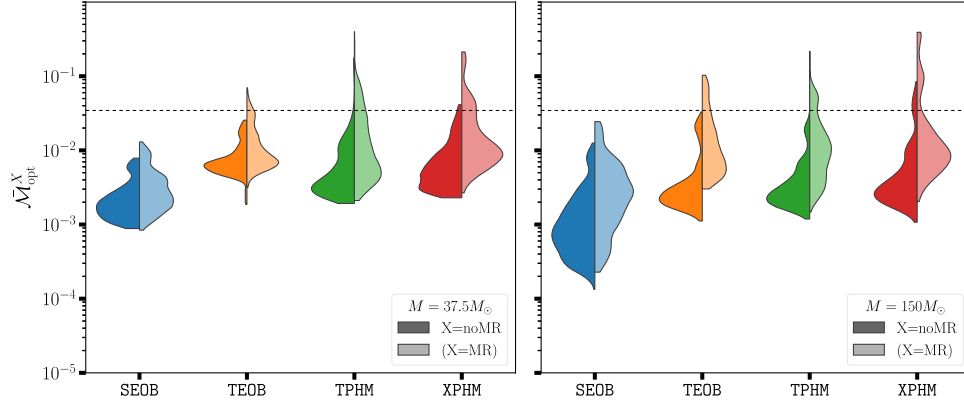


FIG. 4. Merger-ringdown truncated mismatches versus full-length waveform mismatches. The dark-shaded half violins are the distributions of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ [Eq. (23)] between the numerical relativity surrogate NRSUR7DQ4 and the four approximants listed along the bottom horizontal axis of each panel. Similarly, the light-shaded half violins are the distributions of $\bar{\mathcal{M}}_{\text{opt}}$ [Eq. (21)], which are shown in the legend as $\bar{\mathcal{M}}_{\text{opt}}^{X=\text{MR}}$. These are the same quantities plotted in Fig. 3 as histograms. The left (right) panel contains the data from the light (heavy), $M = 37.5M_{\odot}$ ($150M_{\odot}$), binaries. We used the default Gaussian kernel density estimator of the SEABORN library [201] to smooth the histograms for the violin plots. The horizontal dashed line marks the mismatch value of $1 - 0.9^{1/3} \approx 0.035$.

vectors become more planar, i.e., as $\theta_{1,2} \rightarrow \pi/2$, whereas for XPHM, the large- \mathcal{R} region is mostly situated in the $\theta_{1,2} \rightarrow 5\pi/6$ region. As for SEOB, since $1 \leq \mathcal{R} \lesssim 2$, no specific region stands out.

For the heavy BBHs, \mathcal{R} increases significantly for all models. SEOB still yields the smallest values with $\mathcal{R} \lesssim 6$ regardless of the mass ratio with a small region of $\mathcal{R} \approx 8$ in the $Q = 4$ subset. The values of \mathcal{R} output by the other three models show more variation with mass ratio as $\mathcal{R} \lesssim \{12, 18, 14\}$ for TEOB, $\mathcal{R} \lesssim \{6, 10, 12\}$ for TPHM, and $\mathcal{R} \lesssim \{6, 10, 17\}$ for XPHM corresponding to $Q = \{1, \bar{1}, 2, 4\}$. Unlike the light set, we now observe larger- \mathcal{R} regions in the $\chi_{\text{eff}} > 0$ half of the parameter space for SEOB, TEOB, and TPHM with the opposite relation applying to XPHM.

In order to better determine how much of the increased mismatches are genuinely due to mismodeling of the merger-ringdown regimes, one could construct hybrid waveforms where the inspiral-only part generated by the four models is attached to the plunge-merger-ringdown part of the corresponding NRSUR7DQ4 waveform. One could then compute $\bar{\mathcal{M}}_{\text{opt}}$ between the hybrid waveforms and the full NRSUR7DQ4 waveforms, and compare these values to the ones obtained here. Though this strategy is, in principle, straightforward, its implementation is highly nontrivial with precession. We leave this for future work.

2. The dependence of waveform unfaithfulness on the mass ratio

Another known degradation of waveform faithfulness occurs for systems with more component mass asymmetry, i.e., large (small) values of $Q(q)$. In terms of our sample, this should translate to increasing mismatches with increasing Q (or decreasing q). In order to better reduce the potential “contamination” of the mismatches due to

merger-ringdown mismodeling, we present only $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ separated by Q values here. Accordingly, we should expect to see $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(Q = 1, \bar{1}) \leq \bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(Q = 2) \leq \bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(Q = 4) \leq \bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(Q = 6)$, where we are appropriating the \leq symbol to mean that the distributions have higher occupancy per bin at lower mismatch values. The results of this breakdown by mass ratio are shown in Fig. 5, where we plot the distributions of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ as separate half violins for the $Q = 1, \bar{1}, 2, 4, 6$ subsets, but delegate our discussion of the last subset ($Q = 6$) to Sec. IVA 4.

Let us recall that increasing Q increases the number of cycles in the range $[f_i, f_f]$, so part of the increased mismatch may be simply due to having longer waveforms. However, this is not the only factor. Given a value of Q , the spin configurations that yield the longest waveforms are those with $\theta_i = \pi/6$ and the shortest waveforms are the ones with $\theta_1 = 5\pi/6$ (150°), $\theta_2 = \pi$. It is true that the cases with $\{Q, \theta_1, \theta_2\} = \{6, \pi/6, \pi/6\}$ have higher mismatches than, e.g., the cases with $\{1, \bar{1}, \pi/6, \pi/6\}$, where the former cases have nearly twice as many cycles in the $[f_i, f_f]$ interval. On the other hand, the mismatches for $\{Q, \theta_1, \theta_2\} = \{6, \pi/6, \pi/6\}$ are lower than the cases with $\{6, \pi/2 \lesssim \theta_1 \lesssim 5\pi/6, \pi/2 \lesssim \theta_2 \lesssim 5\pi/6\}$, where the latter cases have 10 to 20 fewer cycles in band. Therefore, waveform length alone cannot explain the increased mismatches.

Figure 5 essentially corroborates our expectations. We find that for both light and heavy systems, the best agreement with NRSUR for each approximant is always by the $Q = 1, \bar{1}$ subset with the $Q = 2$ subset yielding slightly worse mismatches. What is additionally apparent from the figure is the relative upward shift of the mismatch distributions for the $Q = 4, 6$ subsets, i.e., a clear deterioration of waveform faithfulness for more mass asymmetric

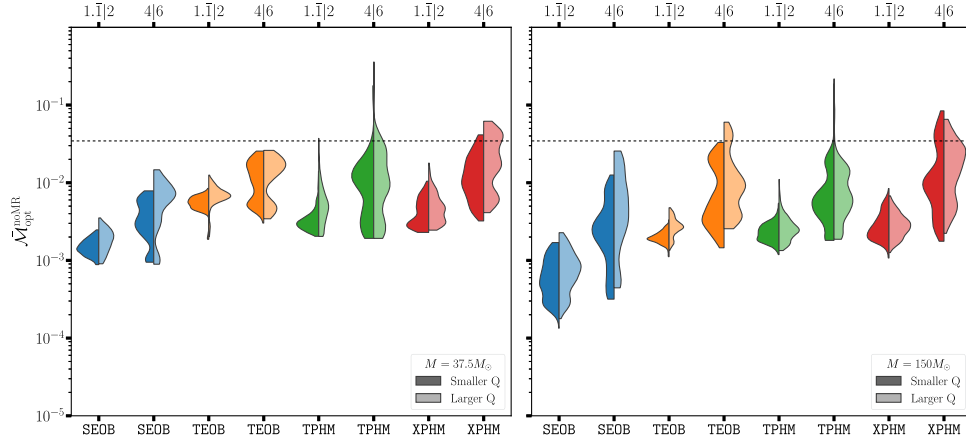


FIG. 5. Mass ratio separated, merger-ringdown truncated mismatches, $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$, between NRSUR7DQ4 and the approximants {SEOB, TEOB, TPHM, XPHM} labeled along the bottom axes and plotted in blue, orange, green, red, respectively. For each approximant, we present four half violins representing, from left to right, the $\vartheta_{\text{LN},0} = 0$ mismatches for the subsets with mass ratios $Q = 1.\bar{1}, 2, 4, 6$ which are shown along the top axes. The $Q = 6$ subset is discussed separately in Sec. IV A 4 as it involves comparisons with NRSUR7DQ4 in its extrapolation region. The horizontal dashed line marks the mismatch value of $1 - 0.9^{1/3} \approx 0.035$.

systems. We also observe an emergence of multimodalities in the distributions of the $Q = 4, 6$ mismatches.

In more detail, we observe that the worst mismatches from the $Q = 4$ subset predominantly come from the cases with $\chi_p \gtrsim 0.7$ ($0.6 \lesssim \chi_p^{\text{Gen}} \leq 1$ and $\chi_{\perp,J} \gtrsim 0.4$), where we quote the values at f_0 . This is somewhat expected as these cases precess more strongly. However, for the heavy-mass subset, some of the worst mismatches show up by $\chi_p \gtrsim 0.4$ indicating additional mismatch dependence on the parallel projection of the spins. This might be connected with some of the observed multimodalities. For example, we discovered that the two distinct peaks in TEOB's $Q = 4$ mismatch distributions can be mapped to two separate regions in the parallel-perpendicular spin projection space. Specifically, we find that for $M = 150M_\odot$, the mismatches of the upper peak ($\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}} > 10^{-2}$) solely come from cases with $\chi_{\text{eff}} < 0$. Light SEOB exhibits a less prominent bimodality separated at $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}} \approx 0.004$ with the mismatches coming from the upper peak corresponding to cases with $\chi_p \gtrsim 0.7$ and mostly $\chi_{\text{eff}} < 0$. Light TPHM also exhibits a bimodality separated at $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}} \approx 0.004$ corresponding to $\chi_p \gtrsim 0.7$ regardless of the value of χ_{eff} . Heavy XPHM distribution separates at $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}} \approx 0.02$ with mismatches of the upper peak all coming from $\chi_{\text{eff}} < 0$ cases.

We show how the multimodalities of these four cases map to $\{\chi_{\perp,J}, \chi_{\text{eff}}\}$ space in Fig. 6 for the $Q = 4$ subset. For this figure, we opted to use $\chi_{\perp,J}$ instead of χ_p (or χ_p^{Gen}) for visual clarity. From the figure, we observe that for the $\chi_{\perp,J} \gtrsim 0.4$ ($\chi_p \gtrsim 0.5$) cases, the $\chi_{\text{eff}} < 0$ corner of the parameter space seems to be more challenging than the $\chi_{\text{eff}} > 0$ corner for all approximants except for TPHM. The aforementioned breakdown of XPHM's MSA prescription is also exhibited in the $\chi_{\text{eff}} \lesssim -0.5, \chi_{\perp,J} \lesssim 0.05$ ($\chi_p \lesssim 0.2$) corner of the parameter space.

Finally, let us finish this section with a brief breakdown of the full mismatches ($\bar{\mathcal{M}}_{\text{opt}}$) in terms of mass ratio so as to complement Fig. 5. For SEOB, we find no cases of

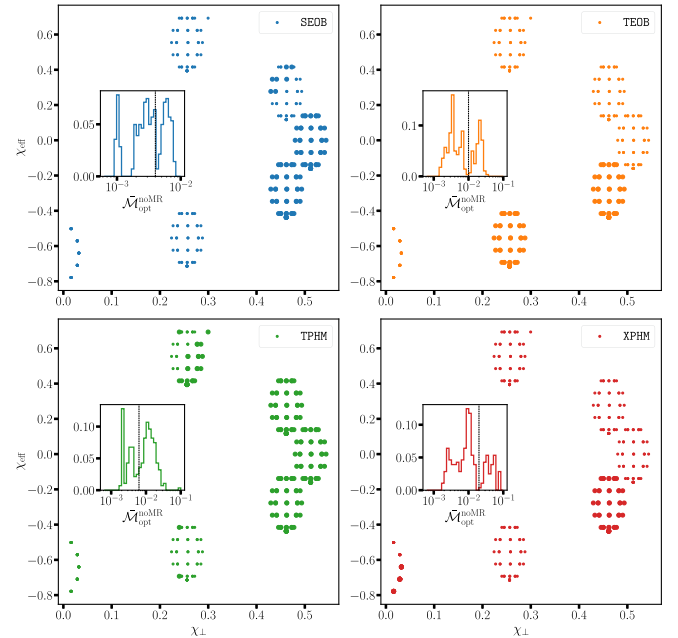


FIG. 6. The association of certain peaks in the mismatch distributions with specific regions in the spin parameter space. In each of the four panels, we scatter plot the $\{\chi_{\perp,J}, \chi_{\text{eff}}\}$ values of the $Q = 4$ subset of the discrete parameter space. The larger disks correspond to points which yield values of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ [Eq. (23)] in the rightmost peaks of the histograms plotted in the insets where the thick, vertical dashed lines separate the multiple modes (peaks) of each histogram. The smaller squares mark the complementary $Q = 4$ cases. The SEOB, TPHM histograms in the first and third insets correspond to the $Q = 4$ half violins from the left panel of Fig. 5, while the rest from the right panel.

$\bar{\mathcal{M}}_{\text{opt}} > 0.035$ for $Q \leq 4$ as is clear from Figs. 3 and 4. {TEOB, TPHM, XPHM} are also robust for $Q \leq 2$ with very few cases of $\bar{\mathcal{M}}_{\text{opt}} > 0.035$ for either light or heavy BBHs, worst being 3% of the $M = 150M_{\odot}(37.5M_{\odot})$ cases for TEOB (TPHM), but their faithfulness degrades for higher values of Q . For example, at $Q = 4$, {14%, 40%, 50%} of the light-mass cases yield $\bar{\mathcal{M}}_{\text{opt}} > 0.035$ which becomes {40%, 36%, 49%} for the heavy-mass set for {TEOB, TPHM, XPHM}, respectively. We can also contrast these percentages with their $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}} > 0.035$ counterparts which are {0%, 2%, 4%} for the light and {0%, 1%, 2%} for the heavy sets. In short, there is a noticeable degradation of model performance in going from $Q \lesssim 2$ to $Q \gtrsim 4$ for all models including SEOB though overall it is always better than 0.965 faithful for $Q \leq 4$.

3. The effect of AS/coprecessing multipoles on faithfulness

Since each precessing multipole of the four models is constructed using Eq. (3) or some variant of it (e.g., using coprecessing multipoles instead of AS multipoles), there are two main modeling systematics at interplay here: (i) systematics coming from the modeling of precession dynamics that are manifest in the Euler angles used in the frame rotation, and (ii) systematics in the AS (or coprecessing) multipoles that are being Euler rotated. We focus on the latter in this section.

Even if the Euler angles of {SEOB, TEOB, TPHM, XPHM} equal those of NRSUR (not actually the case), strain mismatches can still be large if the Euler-rotated multipoles poorly match NRSUR’s coprecessing multipoles. Moreover, in the case of unequal Euler angles, the AS/coprecessing multipole mismatch may still end up being the dominant systematic. We investigate this here by computing AS/coprecessing multipole mismatches between the four models and NRSUR’s coprecessing multipoles.

First, let us add a few details. To our knowledge, among the four models, TEOB is the only one that rotates actual AS multipoles. SEOB twists AS multipoles with the constant spin parameters χ_i replaced by the time-varying $\chi_i(t) \cdot \hat{\mathbf{L}}_N(t)$ evolved via SEOB-specific dynamics. TPHM and XPHM rotate AS multipoles with modified remnant properties [128,131]. Though these multipoles may be referred to as coprecessing, they are all approximations and thus, not exactly equal to $h_{\ell m}^{\text{coprec}}$ of Eq. (7). Therefore, let us denote these approximate coprecessing multipoles by $h_{\ell m}^{\text{AS}}$ as well. These still satisfy the $m \leftrightarrow -m$ multipole symmetry. Though exact for nonprecessing systems, this is an approximation for the true coprecessing multipoles which are known to violate this symmetry (see, e.g., Fig. 2 of Ref. [182]).

We can now ask: How faithful are AS multipoles to the true coprecessing multipoles? And what is the penalty in using AS multipoles to construct precessing waveforms?

Reference [182] provides the first detailed answer to this question using comparisons to 72 NR simulations though only six cases have $Q \geq 4$. They find that the AS $(2, \pm 2)$ multipoles are faithful representations of their coprecessing counterparts with only one (five) out of the 72 simulations resulting in mismatches larger than 0.03 (0.01) (see their Fig. 2 and Table III), but that the AS $(2, \pm 1)$ multipoles may not be considered to be so (*ibid.*). Especially relevant here is their specific comparison with a short BAM [138,139] simulation (ID 28) with $Q = 3, \chi_p = 0.8$, where they identify the m -multipole asymmetry as the cause of the high mismatch.

In the following, we conduct a study similar to Ref. [182]’s. Specifically, we compute mismatches between the coprecessing $(2, \pm 2)$ multipoles of NRSUR7DQ4 and those of {SEOB, TEOB, TPHM, XPHM}. We do this by modifying the source code for the GWSURROGATE package. We focus on h_{22}^{AS} which is dominant in Eq. (3), which we also treat as a proxy for results pertaining to $h_{2,-2}^{\text{AS}}$. We proceed by first computing a new quantity: $\bar{\mathcal{M}}_{22,\text{AS}}$, the standard $\{t_c, \varphi_c\}$ -optimized mismatch between NRSUR’s coprecessing $(2,2)$ multipole and each AS $(2,2)$ multipole of the four models. We then look for correlations between these mismatches and the strain mismatches $\bar{\mathcal{M}}(\vartheta_{\text{LN},0} = 0)$, $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)$. Indeed, as Fig. 7 highlights, there is a very prominent correlation between $\log \bar{\mathcal{M}}_{22,\text{AS}}$ and $\log \bar{\mathcal{M}}$ for TPHM and TEOB, with XPHM showing a weaker trend and SEOB hardly showing any. The linear correlation of the logs implies a power law relation between these mismatches. Additionally, we observe that

- (i) the correlations are stronger between $\bar{\mathcal{M}}_{22,\text{AS}}$ and $\bar{\mathcal{M}}$ than $\bar{\mathcal{M}}_{\text{opt}}$. This means that the optimization of the mismatch over extrinsic parameters partially “smears out” this relation;
- (ii) the correlations are much more prominent for the light BBHs, where more inspiral cycles are included in the match integrals;
- (iii) the correlations persist in all mass ratio subsets, but are stronger for $Q = 1.1$ and 2.

For TPHM and TEOB, the $Q = \{1.1, 2\}$ subset correlation coefficients are larger than 0.9. Moreover, we observe that for the cases that yield $\bar{\mathcal{M}}_{22,\text{AS}} \gtrsim 0.05$, both TEOB’s and TPHM’s $\bar{\mathcal{M}}$ can be fit by lines parametrized by $\bar{\mathcal{M}}_{22,\text{AS}}$ with slopes $\gtrsim 1$, not just their logarithms. This is an indication that the AS multipole unfaithfulness, when large enough, becomes the dominant systematic in waveform faithfulness for these two models. This linearity is less obvious for SEOB and XPHM. The former model produces the most faithful AS multipoles, $\bar{\mathcal{M}}_{22,\text{AS}} < 0.01$ for almost every case, whereas the latter model has the most outliers indicating other systematics contaminating this relation such as the already mentioned MSA-related breakdown. Despite these exceptions, when $\bar{\mathcal{M}}_{22,\text{AS}}$ is

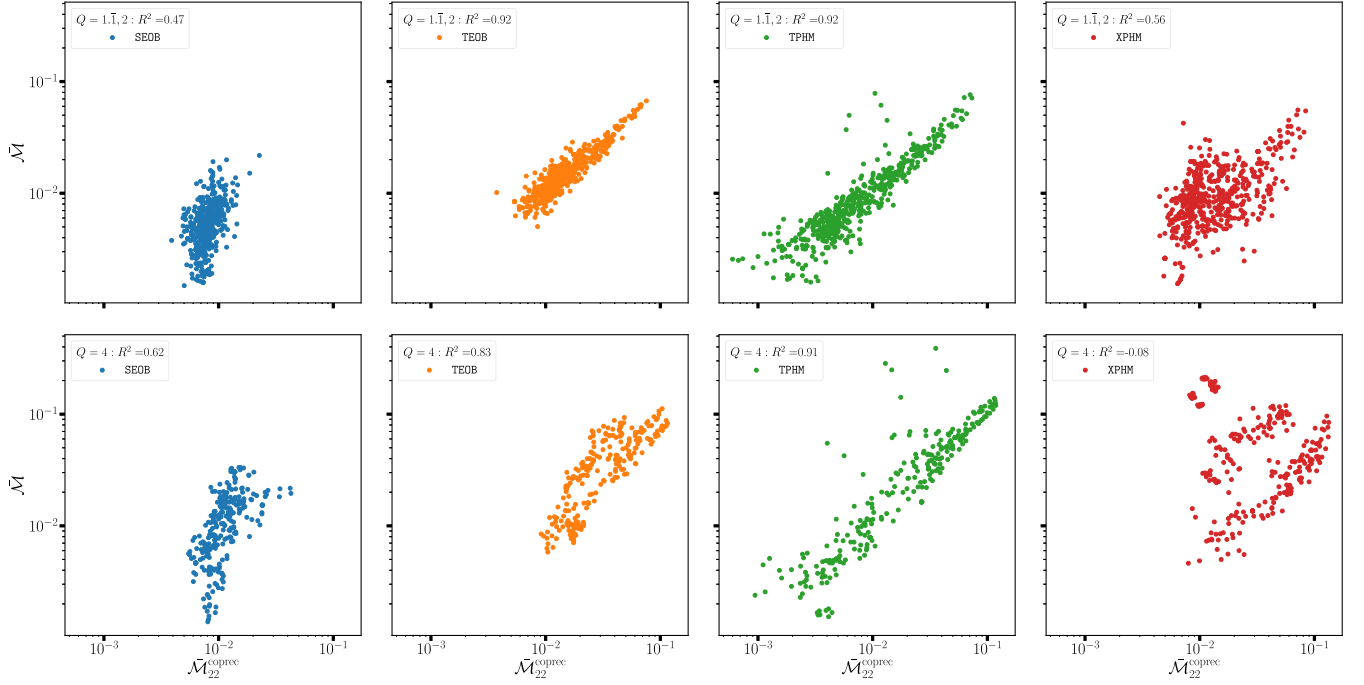


FIG. 7. The correlation between $\bar{\mathcal{M}}_{22,AS}$ and $\bar{\mathcal{M}}(\vartheta_{LN,0} = 0)$ for the $M = 37.5M_{\odot}$ cases, where $\bar{\mathcal{M}}_{22,AS}$ is the mismatch [via Eq. (13)] between NRSUR’s coprecessing (2,2) multipole and each AS (2,2) multipole of the four models labeled in the legends where we also show the square of the Pearson correlation coefficient. The upper panels show the mismatches from the combined $Q = 1.\bar{1}$ and $Q = 2$ subsets, while the lower panels show the results from the $Q = 4$ subsets. The cluster of red dots in the upper left corner of the $Q = 4$ XPHM panel are the cases for which the aforementioned MSA prescription of XPHM’s precession dynamics breaks down.

large enough, the observations above also hold for these two models.

This systematic becomes more severe with increasing mass ratio as we determined by comparing the slopes of the linear fits to the mismatch data for TEOB and TPHM between the $Q = 1.\bar{1}, 2$, $Q = 4$ and the $Q = 6$ subsets. In each comparison, the slope of the larger Q subset was greater than the slope of the smaller Q subset, though we observed this to be more severe for TEOB than for TPHM.

When we looked at the intrinsic parameters of the cases that yield the highest values for $\bar{\mathcal{M}}_{22,AS}$ for each approximant, we found that the worst mismatches come from the $65^{\circ} \lesssim \theta_{1,2} \lesssim 115^{\circ}$ region, resulting in the largest planar spin projections. We have already identified this region as the most challenging one consistent with the literature. The fact that the AS multipole mismatches are also the highest in this region tells us that the true coprecessing multipoles differ much more than the AS multipoles for cases with mostly planar spins. We believe this to be mostly due to the imposed $m \leftrightarrow -m$ symmetry on the AS multipoles, which we confirmed to be more violated for the strongly precessing cases. We did this by comparing $\bar{\mathcal{M}}_{22,AS}$ with $\bar{\mathcal{M}}_{22,AS}^{\text{sym}}$, i.e., the AS multipole mismatches computed with respect to the m -symmetrized coprecessing NRSUR7DQ4 multipoles, and found that the quantity $\bar{\mathcal{M}}_{22,AS} - \bar{\mathcal{M}}_{22,AS}^{\text{sym}}$ peaks for

$65^{\circ} \lesssim \theta_{1,2} \lesssim 115^{\circ}$, i.e., for the cases with mostly in-plane spins.

When we extended the above analysis of the AS (2,2) multipole to the AS (2,1) multipole, we encountered some unexpectedly high values of $\bar{\mathcal{M}}_{21,AS}$ for $Q = 1.\bar{1}$ for all four models. This contradicts the thus-far observed trend of increasing mismatches with increasing Q . Upon closer examination, we discovered that the problem was due to NRSUR7DQ4: Since for $Q = 1$, $h_{21}^{\text{AS}} = h_{21}^{\text{coprec}} = 0$ exactly, the model generates a coprecessing (2,1) multipole that is barely above numerical error for $Q = 1.\bar{1}$. For this reason, we choose to disregard all $Q = 1.\bar{1}$ values of $\bar{\mathcal{M}}_{21,AS}$.

Focusing on the $Q \geq 2$ part of the discrete set, we observe that the previously observed tight correlation for the AS (2,2) multipole becomes much less strong for the AS (2,1) mode. In fact, the linear trend between $\log \bar{\mathcal{M}}_{21,AS}$ and $\log \bar{\mathcal{M}}$ is only discernible for TPHM, with SEOB and TEOB showing some positive correlation at the high-mismatch end as well. As was the case with the AS (2,2) multipole, the trends are more visible for the $M = 37.5M_{\odot}$ BBH set. As for the cases that yield the worst AS (2,1) mismatches, we find more of a spread in the $\{\theta_1, \theta_2\}$ space, but with the worst common mismatches coming from the $\theta_1 \leq \pi/2, \theta_2 \geq \pi/2$ corner, while the cases with maximum or near-maximum $|\chi_{\text{eff}}|$ values consistently yield the

lowest mismatches. We should keep in mind that even for the worst cases where this correlation is tight, it need not necessarily affect the overall unfaithfulness $\bar{\mathcal{M}}_{\text{opt}}$ significantly as the AS (2,1) multipole amplitude is roughly an order of magnitude smaller than its (2,2) counterpart. And as we discuss further below, the AS (2, ± 1) multipoles have a rather weak effect on the overall strain faithfulness.

The AS (2, ± 1) multipoles' importance for faithful precessing waveform construction was investigated by Ref. [182] where they showed that while the inclusion of the AS (2, ± 1) multipoles degrades the precessing (2, ± 1) multipole faithfulness (see their Table III and Fig. 4), it marginally improves the precessing (2, ± 2) multipoles. How these changes affect the faithfulness of the waveform strain is illustrated for a $Q = 5$ case in their work, where it is shown that the inclusion of the AS (2, ± 1) multipoles lowers the $\ell = 2$ strain unfaithfulness by 0.01 in the highest mismatch regions as can be gathered by comparing the middle left with the middle right panel in their Fig. 5. In the words of Ref. [182], this indicates that the improvement in the precessing (2, ± 2) multipoles, due to the inclusion of the (2, ± 1) AS multipoles, compensates for the degradation of the precessing (2, ± 1) multipoles.

We extend their work here by turning “off” $h_{2\pm 1}^{\text{coprec}}$ for NRSUR7DQ4 and looking at the unfaithfulness of the resulting strain with respect to the full-multipole content NRSUR7DQ4 strain. Specifically, we compute $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and $\bar{\mathcal{M}}_{\text{opt}}$ for the $\ell = 2$ strain at inclinations of zero and $\pi/2$. Though we use the same multipole content as Ref. [182], they considered all inclinations in the one example that they present whereas we suffice with two inclinations here, but repeat the computation for all 1120 cases of the discrete set.

For $\vartheta_{\text{LN},0} = 0$, we find that $\bar{\mathcal{M}}_{\text{opt}}$ ranges from $\sim 10^{-8}$ to 10^{-2} with mismatches increasing with Q and the highest values coming from the $\theta_{1,2} = \pi/2$ cases of the $Q = 4, 6$ subsets. For $\vartheta_{\text{LN},0} = \pi/2$, we observe $10^{-5} \lesssim \bar{\mathcal{M}}_{\text{opt}} < 0.035$, with the dependence on Q becoming less important. In other words, when the coprecessing (2, ± 1) multipoles are omitted, we observe increased strain unfaithfulness at higher inclinations. Accordingly, the coprecessing (2, ± 1) multipoles improve the strain faithfulness more for higher inclinations, consistent with the findings of Ref. [182]. Interestingly, the largest values of $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2)$ come from cases with $\theta_{1,2} \geq 5\pi/6$, not $\pi/2$ as was the case for $\vartheta_{\text{LN},0} = 0$.

4. Mismatches in the extrapolation region of NRSUR7DQ4

Since the approximants exhibit a degradation in faithfulness as Q increases, we extended our investigation into the extrapolation regime of NRSUR7DQ4 by setting $Q = 6$ for the same discrete grid as before, resulting in two additional subsets of light and heavy BBHs with 280 cases each. We have opted to present this comparison separately from the $Q \leq 4$ cases of the previous subsections since, as Ref. [77]

advises, we must exercise caution when using NRSUR7DQ4 in its extrapolation region, i.e., $4 < Q < 6, 0.8 < \chi_i < 1$. Nonetheless, comparisons with 100 sxs simulations with $Q = 6$ and $\chi_i \leq 0.8$ have yielded mismatches ranging from $\sim 10^{-4}$ to $\sim 10^{-2}$ with the median at $\sim 10^{-3}$. Overall, this is roughly an order of magnitude worse than NRSUR7DQ4's performance in its training region, $Q \leq 4, \chi_i \leq 0.8$, but good enough for our purposes here. A more direct approach would be to compare the four models with the aforementioned 100 sxs simulations, but we were unable to retrieve these from the sxs database. We present comparisons using different sets of sxs simulations in Sec. V.

Returning again to Fig. 5, where the $Q = 6$ $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ distributions are plotted as the rightmost half violins for each color, we see that the worsening trends observed for the ($Q = 4$)-case mismatches persist for the $Q = 6$ cases. As before, the mismatch distributions exhibit bi- or trimodalities with the larger-mismatch peaks of {SEOB, TEOB, XPHM}'s distributions coming from cases with either large planar spins, i.e., $\chi_p \gtrsim 0.7$ or with moderate χ_p values combined with negative χ_{eff} values. TPHM once again exhibits mismatches that are more symmetric with respect to χ_{eff} . Overall, the $Q = 6$ $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ values for the four models are slightly worse than the $Q = 4$ values. Interestingly, the $\chi_{\text{eff}} \lesssim -0.5, \chi_p \lesssim 0.2$ corner of the parameter space that previously yielded high mismatches for XPHM has shrunk to a single point that has the most negative χ_{eff} value of the entire $Q = 6$ set. This minor improvement in XPHM's performance is partly due to the fact that we no longer have the near-transitional precession cases encountered for $Q = 4$ where $|\mathbf{J}_{N,0}| \ll |\mathbf{L}_{N,0}|$, which were causing the MSA prescription to break down forcing the model to default to the NNLO prescription for precession dynamics.

Taking $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ as a gauge of inspiral-only model faithfulness, we find that {SEOB, TEOB, TPHM, XPHM} are reasonably faithful at this mass ratio with the percentage of cases of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}} > 0.035$ being {0%, 0%, 5%, 28%} for the light- and {0%, 14%, 2%, 11%} for the heavy-mass sets. As for the full mismatch $\bar{\mathcal{M}}_{\text{opt}}$, the percentage of cases greater than 0.035 is, respectively, {0%, 34%, 37%, 75%} for the light- and {16%, 41%, 46%, 77%} for the heavy-mass set.

B. Behavior of model unfaithfulness under changing inclination

Our aim in this section is to investigate how the inclusion of the precessing (2,1) multipole, i.e., h_{21} , affects the faithfulness. Equation (1) tells us that the contribution of h_{21} is maximal for $\vartheta_{\text{LN}} = \pi/2$ yielding $|_{-2}Y^{21}|/|_{-2}Y^{22}| = 2$. Though $|h_{21}^{\text{AS}}| < |h_{22}^{\text{AS}}|$ always holds true, for precessing multipoles, one can observe that $|h_{21}| \sim |h_{22}|$ (see, e.g., Fig. 1 of Ref. [77]). This is due to the power from the AS (2, ± 2) multipoles rotating into

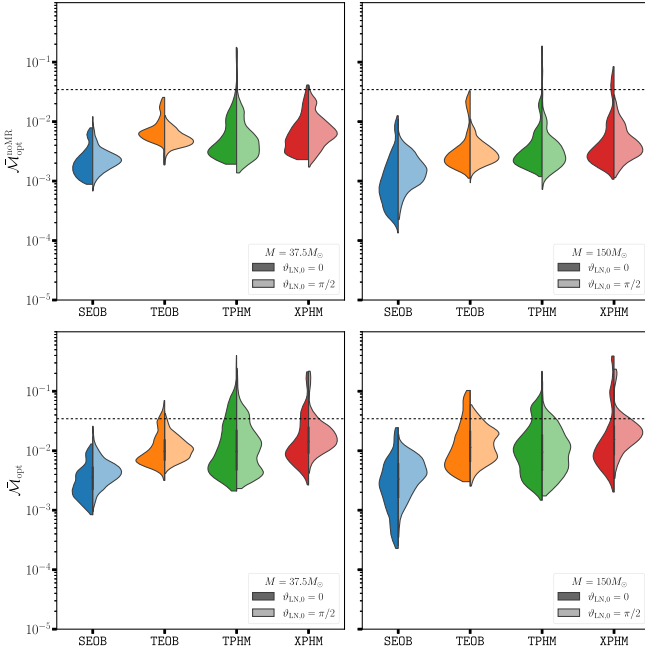


FIG. 8. Comparison of the mismatch distributions between face-on ($\vartheta_{\text{LN},0} = 0$) and edge-on ($\vartheta_{\text{LN},0} = \pi/2$) configurations. In the top row, we display violin plots of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ vs $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2)$ for the $M = 37.5M_{\odot}$ ($M = 150M_{\odot}$) sets in the left (right) panels. Likewise, the bottom row shows $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)$ vs $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2)$ for the same sets. Note that we exclude the $Q = 6$ subset here, but we include it in Fig. 9.

h_{21} which is time (frequency) dependent, as is ϑ_{LN} . Accordingly, the ${}_{-2}Y^{21}h_{21}$ term in the strain (1) can be dominant. Therefore, comparisons of strain mismatches for cases with large inclinations offer an indirect way to assess the faithfulness of “subdominant” multipoles such as h_{21} which, as we just explained, may not be subdominant at all for certain configurations. The analysis almost equally applies to $h_{2,-1}$ due to the precessing multipoles approximately inheriting the $m \leftrightarrow -m$ symmetry of the AS multipoles. For a more detailed investigation of the mismodeling of h_{21} and higher multipoles ($\ell > 2$), see Ref. [182].

This analysis also indirectly gauges the performance of the Euler angles used in the frame rotations. Since, as shown, the faithfulness of the coprecessing $(2, \pm 2)$ multipoles matters much more than that of the coprecessing $(2, \pm 1)$ multipoles, the cases for which the former are very faithful, but the resulting $\bar{\mathcal{M}}_{\text{opt}}$ values are high may tell us something about the mismodeling of the Euler angles.

We proceed by reemploying the discrete parameter set of Sec. IV A with $\vartheta_{\text{LN},0} = \pi/2$, and compute the mismatches $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$, $\bar{\mathcal{M}}_{\text{opt}}$ as before. Our results are presented in Fig. 8 where we show the distributions of the mismatches between the $\vartheta_{\text{LN},0} = 0$ and $\vartheta_{\text{LN},0} = \pi/2$ sets for the $Q \leq 4$ portion for both light and heavy binaries. Several features stand out in the figure which we highlight next.

First, as in the $\vartheta_{\text{LN},0} = 0$ case, the distributions of $\bar{\mathcal{M}}_{\text{opt}}$ are in general broader than $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ distributions and shifted to higher values. This shift is larger for the heavy masses. Second, the inspiral-only faithfulness of all approximants remains mostly unchanged between $\vartheta_{\text{LN},0} = 0$ and $\vartheta_{\text{LN},0} = \pi/2$, with TEOB even showing a slight improvement for $\vartheta_{\text{LN},0} = \pi/2$. There is also an overall trend of the high-mismatch tails being reduced.

In Fig. 9, we show a decomposition of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ and $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2)$ in terms of mass ratio, where the upper (lower) panels correspond to the $M = 37.5M_{\odot}$ ($M = 150M_{\odot}$) set. We note several observations. First, the $\vartheta_{\text{LN},0} = \pi/2$ distributions exhibit less or no modality compared with the $\vartheta_{\text{LN},0} = 0$ distributions. Second, $Q \geq 4$ $\vartheta_{\text{LN},0} = \pi/2$ mismatches have narrower distributions and mostly occupy lower values than their $\vartheta_{\text{LN},0} = 0$ counterparts. The converse holds true for the $Q \leq 2$ mismatches with the exception of TEOB. Third, the medians of the $\vartheta_{\text{LN},0} = \pi/2$ distributions shift to lower values than the medians of the $\vartheta_{\text{LN},0} = 0$ distributions as Q increases, especially for the EOB models. The especially remarkable feature is the improved inspiral-only faithfulness for all models for $Q = 4$ at $\vartheta_{\text{LN},0} = \pi/2$ as compared with $\vartheta_{\text{LN},0} = 0$.

For $Q = 6$, there is mostly a reduction of the bi/trimodalities which we consider to be an improvement. However, the situation is somewhat more complicated for $Q = 6$ as the values of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ of the leftmost peaks are mostly lower than the smallest values of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2)$ whereas the values of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ of the rightmost peaks are mostly higher than the largest values of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2)$. So, the question of whether or not the models show improved inspiral-only faithfulness for $Q = 6$ may have a subjective, rather than an objective answer. Therefore, let us focus on the $Q = 4$ cases next.

Given the above observations, an obvious question is, what is the cause of the improved faithfulness for $Q = 4$ in going from $\vartheta_{\text{LN},0} = 0$ to $\vartheta_{\text{LN},0} = \pi/2$? We hypothesize that this may be due to the $h_{2,\pm 1}$ multipoles improving the strain when they are more faithful than $h_{2,\pm 2}$. Given that at $\vartheta_{\text{LN},0} = \pi/2$, we have $|{}_{-2}Y^{2,\pm 1}|/|{}_{-2}Y^{2,\pm 2}| = 2$, this improvement more likely matters only for cases where $|h_{2,\pm 1}| \gtrsim |h_{2,\pm 2}|$. This may then result in $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2)$ being less than $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$. Henceforth, we will simply mention the $(2,1)$ and $(2,2)$ multipoles with the same results applying to the $(2,-1)$, $(2,-2)$ multipoles.

We test our hypothesis on the TEOB mismatches first since it shows the most $Q = 4$ improvement for $\vartheta_{\text{LN},0} = \pi/2$ as well as showing improvements for $Q = 1, \bar{1}, 2$. We begin by computing two new quantities: $\bar{\mathcal{M}}_{22}^{\text{noMR}}$, $\bar{\mathcal{M}}_{21}^{\text{noMR}}$, i.e., the inspiral-only $\{t_c, \varphi_c\}$ -maximized

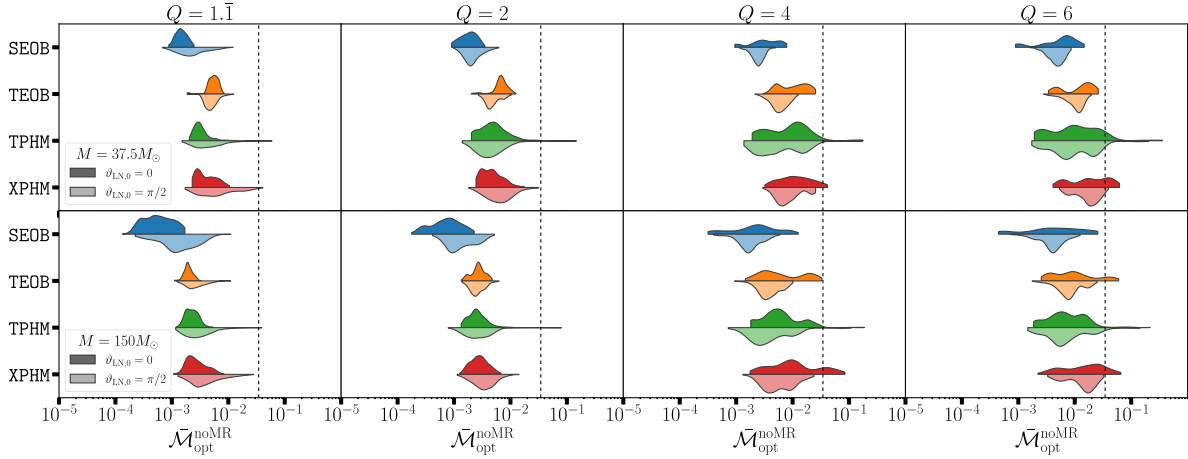


FIG. 9. Comparison of the mismatch distributions between “face-on” ($\vartheta_{\text{LN},0} = 0$) and “edge-on” ($\vartheta_{\text{LN},0} = \pi/2$) configurations separated by mass ratio. In particular, we present results here only for the MR-truncated, sky-angle optimized mismatch $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$. The top row displays results for the light ($M = 37.5M_{\odot}$) BBH set, and the bottom row the heavy ($M = 150M_{\odot}$) BBHs.

mismatches [see Eq. (13)] of the precessing (2,2) and (2,1) multipoles between NRSUR7DQ4 and TEOB. We also compute the ratio

$$R := \frac{2}{N} \sum_{i=0}^N \frac{|h_{21,i}|}{|h_{22,i}|}, \quad (24)$$

where we have $N + 1$ samples in the time domain labeled by i . R is the Riemann sum of the following integral

$$\int dt \frac{|_{-2}Y^{21}(\pi/2, 0)h_{21}(t)|}{|_{-2}Y^{22}(\pi/2, 0)h_{22}(t)|}. \quad (25)$$

$R > 1$ tells us that the $(2, \pm 1)$ terms have larger magnitudes than their $(2, \pm 2)$ counterparts in the waveform strain. Thus when $R > 1$ and $\bar{\mathcal{M}}_{21}^{\text{noMR}} < \bar{\mathcal{M}}_{22}^{\text{noMR}}$, we expect the more faithful h_{21} to result in improved faithfulness for $\vartheta_{\text{LN},0} = \pi/2$ as compared to $\vartheta_{\text{LN},0} = 0$, i.e., $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2) \leq \bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ in the respective distributions.

We proceed by counting the cases satisfying the condition $[R > 1] \wedge [\bar{\mathcal{M}}_{21}^{\text{noMR}} < \bar{\mathcal{M}}_{22}^{\text{noMR}}]$ and comparing the resulting distributions for $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2)$ with those of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$. The expectation is that we should see the distributions of the affirmative cases satisfying $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2) \leq \bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ and that there are enough such cases to affect the overall distributions as seen in Fig. 9. We find our hypothesis validated by the $M = 37.5M_{\odot}$, $Q = 4$ subsets of SEOB and TEOB as well as the $M = 150M_{\odot}$, $Q = 4$ subset of TPHM. We chose these subsets because it is evident from Fig. 9 that this improvement is most prominent for them. We also confirmed our hypothesis for the entire $M = 37.5M_{\odot}$, $Q \leq 4$ TEOB set. We did not use XPHM because it is nontrivial to obtain its

precessing multipoles in the time domain as it is a frequency domain approximant.

The MR-included version of Fig. 9 (which we do not include) shows less modality in the $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2)$ distributions. However, there are fewer Q subsets where $\vartheta_{\text{LN},0} = \pi/2$ mismatches are lower than the $\vartheta_{\text{LN},0} = 0$ mismatches. The exceptions to this relation are SEOB’s light and heavy $Q = 4, 6$ mismatches. For these, we verified that the above arguments held.

The fixed angle ϑ_{JN} is also at play here: Precession-induced modulations to the waveform decrease as $\vartheta_{\text{JN}} \rightarrow 0$ [37]. Using the LALSIMULATION function SIMINSPIRAL-TRANSFORMPRECESSINGWVF2PE, we computed ϑ_{JN} and observe that as the $Q = 4$ subset’s spin tilt angles go from pointing “northward” to southward, $\vartheta_{\text{JN}}|_{\vartheta_{\text{LN},0}=\pi/2}$ drops below $\vartheta_{\text{JN}}|_{\vartheta_{\text{LN},0}=0}$. In other words, the higher inclination cases yield smaller precession-induced modulations, which ties in with the discussion above, but does not explain why the precessing $(2, \pm 1)$ multipoles are sometimes more faithful than $(2, \pm 2)$ when both have the same coprecessing (AS) multipole content dominated by $h_{2,\pm 2}^{\text{coprec}}$ ($h_{2,\pm 2}^{\text{AS}}$).

To answer this question, one would have to dismantle the individual coprecessing (AS) multipole contribution to the twist expression (3) and compare the interplay between various multipoles. We do not pursue this matter any further here. Reference [182] presents additional analysis in this direction, where they compare waveform strain and individual precessing multipole mismatches for cases in which all (useful) AS multipole content has been included and for cases where some AS multipoles, e.g., $(2, \pm 1)$, have been excluded.

The final observation relevant to this section is that just as in the $\vartheta_{\text{LN},0} = 0$ case, the distributions for $\bar{\mathcal{M}}_{\text{opt}}(\pi/2)$ are shifted upward, i.e., to worse values, as compared to $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2)$ distributions, especially for $Q = 4, 6$.

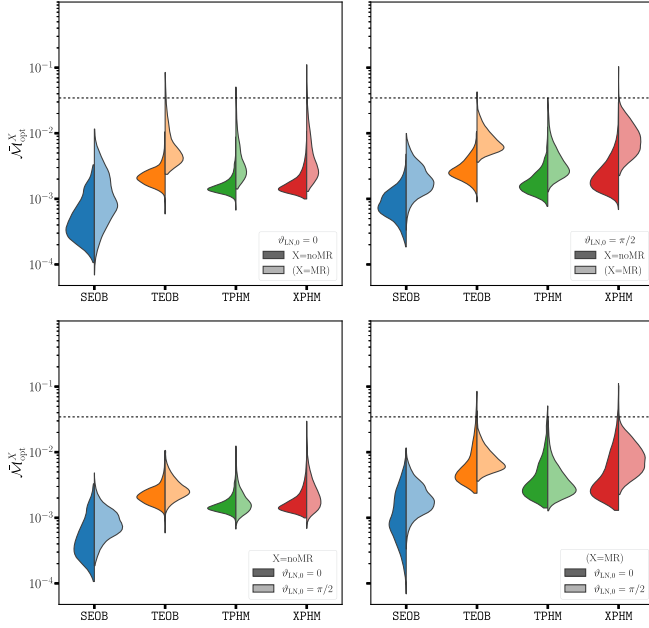


FIG. 10. Mismatches between NRSUR7DQ4 and SEOB, TEOB, TPHM, XPHM for the uniformly filled parameter space of 1000 cases. Upper panels compare $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ with $\bar{\mathcal{M}}_{\text{opt}}$ at inclinations of $\vartheta_{\text{LN},0} = 0$ (left) and $\pi/2$ (right) and the lower panels compare $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ with $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2)$ (left) and $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)$ with $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2)$ (right).

The fact that $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}} \leq \bar{\mathcal{M}}_{\text{opt}}$ is not surprising since the merger-ringdown part of the signal is more complex as we already mentioned. We can nonetheless ask whether or not there might be regions in the spin space where there is a larger gap between $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and $\bar{\mathcal{M}}_{\text{opt}}$ than other regions. To this end, we again looked at how the ratio $\mathcal{R} = \bar{\mathcal{M}}_{\text{opt}} / \bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ is distributed in the spin space. The approximants show a lot of variability with respect to each other and across Q values. However, one nearly common feature is this ratio reaching $\mathcal{O}(10)$ for $\theta_{1,2} = \pi/2$ with XPHM as the exception. Additionally, SEOB has the smallest maxima for the ratio while XPHM has the largest. XPHM has more occurrences for large ratios in the $\chi_{\text{eff}} < 0$ half whereas the other three models yield more large ratios in the $\chi_{\text{eff}} > 0$ half. XPHM also behaves differently for the $Q = 4$ subset with the ratio showing very little variation in the spin space and tending to the smallest values at $\theta_{1,2} = \pi/2$.

C. Comparisons over a uniformly filled parameter space

We now consider a uniformly populated parameter space consisting of 1000 BBHs with $M \in [35, 225]M_{\odot}$, $Q \in [1, 4]$, $\chi_i \in [0.1, 0.8]$. We choose to make the spin tilt angles uniform in their cosine, i.e., $\cos\theta_i \in [-1, 1]$ with $\phi_1 = 0$, $\phi_2 \in [0, 2\pi)$. To ensure once again that the resulting inspirals are not longer than 4300M, we set f_0 equal to a linearly decreasing function of M from 35 to 5.75 Hz as

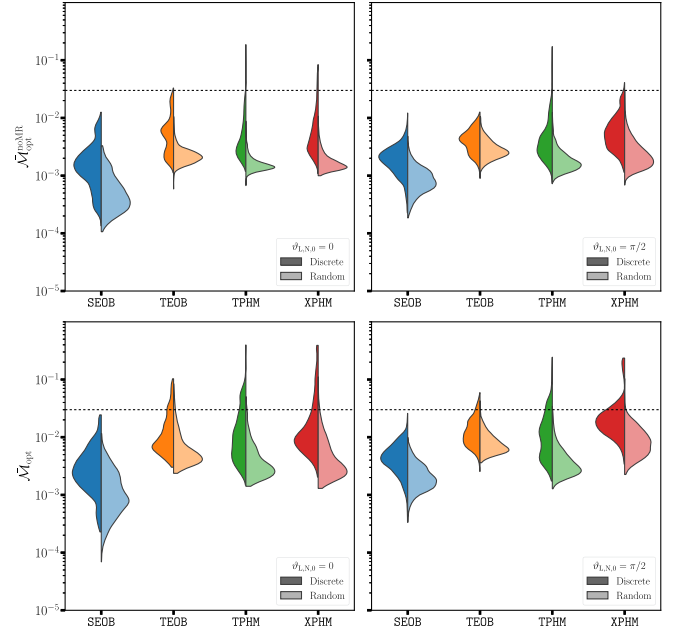


FIG. 11. Comparison of model mismatches to NRSUR7DQ4 between the discrete parameter set and the random-uniformly-filled one. The left half violins are the mismatch distributions from the entire $M = \{37.5M_{\odot}, 150M_{\odot}\}$, $Q = \{1, 2, 4\}$ discrete set while the right ones are the mismatch distributions from the 1000-case random-uniformly-filled set. In the top (bottom) row, we plot $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ ($\bar{\mathcal{M}}_{\text{opt}}$) at inclinations of 0 (left panels) and $\pi/2$ (right panels).

M increases from $35M_{\odot}$ to $225M_{\odot}$. We fix $f_i = f_0 + 3$ Hz (11 Hz) for $M \lesssim 210M_{\odot}$ ($\gtrsim 210M_{\odot}$) in the integral (14). In the following, we present the mismatches similarly to what we have done for the discrete set of Sec. IV A so that we can illustrate the differences in the approximant performance between a more traditional parameter space (uniformly filled) and one for which certain parameters are tailored to potentially better expose precession-related modeling systematics.

We summarize our main results in Fig. 10, where in the top row we compare $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ with $\bar{\mathcal{M}}_{\text{opt}}$ at inclinations of $\vartheta_{\text{LN},0} = 0$ ($\pi/2$) in the left (right) panels. As expected, the relation $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}} \leq \bar{\mathcal{M}}_{\text{opt}}$ holds likely due to the combination of MR modeling systematics and simpler signal morphology for the inspiral. The shifts between the peaks of the $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and $\bar{\mathcal{M}}_{\text{opt}}$ distributions are larger for $\vartheta_{\text{LN},0} = \pi/2$ than for $\vartheta_{\text{LN},0} = 0$ as can be discerned from the upper left and right panels of the figure. This was also the case for the discrete set though this is somewhat hard to extract from Fig. 8. This is a reaffirmation of the fact that the modeling of the plunge-merger-ringdown stages of the precessing ($2, \pm 1$) multipoles remains challenging.

In the bottom panels of Fig. 10, we show how the mismatches shift upward more for SEOB and XPHM than for TEOB and TPHM as the inclination goes from 0 to $\pi/2$, more consistent with the $M = 150M_{\odot}$ (right) panels of Fig. 8

than for the $M = 37.5M_\odot$ (left) ones. This somewhat makes sense as $150M_\odot$ roughly equals the median of $M \in [35, 225]M_\odot$, whereas $37.5M_\odot$ is at the lower end.

Figure 11 better highlights the differences in model faithfulness resulting from using a uniformly filled set vs our purpose-built discrete set of Secs. IV A and IV B. In this figure, the left half violins (darker shaded) are the mismatch distributions from the entire $M = \{37.5M_\odot, 150M_\odot\}$, $Q = \{1, \bar{1}, 2, 4\}$ discrete set, while the lighter-shaded right violins correspond to the mismatch distributions from the 1000-case random-uniformly-filled set. One immediately notices that the distributions for the latter (uniform) set have smaller minima, means, and maxima than the former (discrete) set. Though the shifts between the respective minima are not so large, the offset between the means and the maxima are quite prominent. The distributions for the uniform set are also smoother, narrower, and unimodal owing to the nature of the uniformly filled intrinsic parameter space.

The main conclusion of Fig. 11 is that the model faithfulness is better for the uniform set than for our purpose-built discrete set. This is, of course, not surprising since intrinsic parameters of the discrete set are such that it contains many more cases with strong precession than the uniform set, where $\chi_p(t_0) \geq 0.6$ for 60% of the cases for the former vs 13% for the latter. We already documented this contrasting coverage of the spin space in Fig. 2. Additionally, one third of the discrete set is made up of $Q = 4$ cases whereas the uniform set has $\sim 17\%$ of its cases with $3.5 \leq Q < 4$. Recall that these high- Q cases appear to be the most challenging for the models that we consider here, as can be gathered from Figs. 5 and 9.

Perhaps the difference between the two sets is best illustrated by considering the ‘‘extremal corner’’ of the five-dimensional $\{Q, \chi_1, \chi_2, \theta_1, \theta_2\}$ parameter space. Let us as, an example, assign to this region the following bounds: $Q \in [3.5, 4]$, $\chi_i \in [0.7, 0.8]$, and $\cos(\theta_i) \in [-0.1, 0.1]$. A straightforward computation shows that the uniform set needs to consist of 20379 points in order to have a 50% chance that one case will be inside this extremal region, whereas the discrete parameter set has eight out of 840 cases in this extremal corner (see Table I).

Let us elaborate why the differences shown in Fig. 11 matter. Our concern is for strongly precessing events like GW200129_065458 for which the PE routines should mostly draw samples from the large χ_p region of the spin space. This in itself is not a problem. However, if one employs a model whose faithfulness, to a target model such as NRSUR7DQ4, has been documented over a uniformly filled parameter space, then one may overestimate the faithfulness of that model. For example, a figure akin to our Fig. 10 may lead to the conclusion that model ‘‘X’’ is better than 0.035 faithful in the entire parameter space, but it may turn out that this model X actually has a large-mismatch tail extending to ~ 0.1 in the relevant corner of the parameter

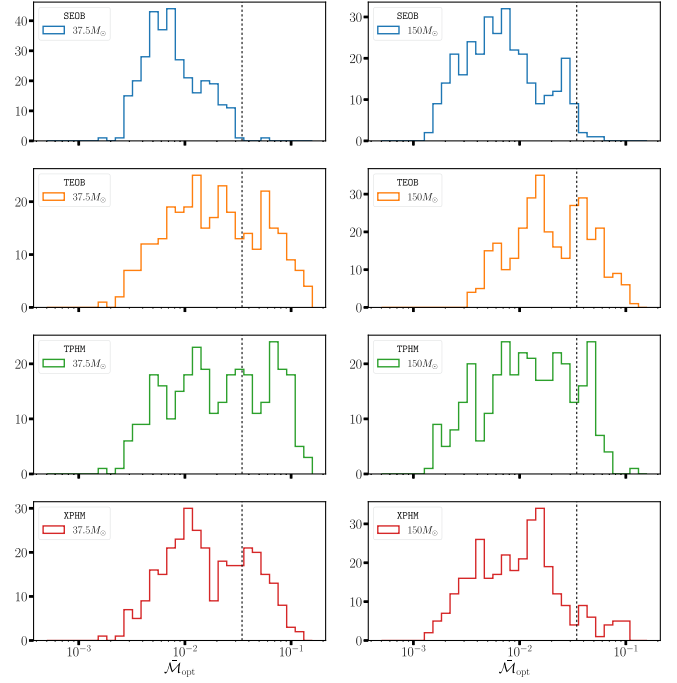


FIG. 12. Mismatches defined by Eq. (21) between the four approximants and the 317 short sxs waveforms shown at an inclination of $\vartheta_{\text{LN},0} = 0$ for a direct comparison with Fig. 3. The left and right columns of panels correspond to mismatches from the light ($M = 37.5M_\odot$) and heavy ($M = 150M_\odot$) binaries.

space as illustrated in our Fig. 11. This would be quite problematic for reliable parameter estimation from strongly precessing events. In short, uniformly sampled intrinsic parameter spaces may not contain the best set of parameters to properly ‘‘stress test’’ precessing approximants.

Let us conclude this section by providing additional metrics. As done in the previous section, we once again look at the ratio $\mathcal{R} = \bar{\mathcal{M}}_{\text{opt}} / \bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$. For $\vartheta_{\text{LN},0} = 0$, we observe that $\mathcal{R} \lesssim 5$ everywhere in the parameter space for all four models. For $\vartheta_{\text{LN},0} = \pi/2$, we record that $\mathcal{R} \lesssim 10$. For both inclinations, these upper bounds are roughly half of the corresponding bounds of \mathcal{R} for the discrete parameter set. As before XPHM has the largest values of \mathcal{R} and SEOB the smallest with TPHM yielding very similar magnitudes for \mathcal{R} in the parameter space. One major difference with respect to the discrete set is that as we uniformly fill the mass ratio space from $Q = 1$ to 4, the clear-cut regions of large \mathcal{R} observed for the discrete set at, e.g., $Q = 4$, are no longer there. Overall, the values of \mathcal{R} are more uniformly spread in the spin space and show less variation.

Finally, we count the percentage of cases with $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}} > 0.035$ and $\bar{\mathcal{M}}_{\text{opt}} > 0.035$ for both inclinations. As is clear from Fig. 10, this is 0% for all SEOB mismatches and 0% for all other $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ values regardless of inclination. As for $\bar{\mathcal{M}}_{\text{opt}}$ percentages, we record $\{2\%, 0.4\%, 0.8\%\}$ for $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0) > 0.035$, and $\{0.1\%, 0\%, 0.2\%\}$ for

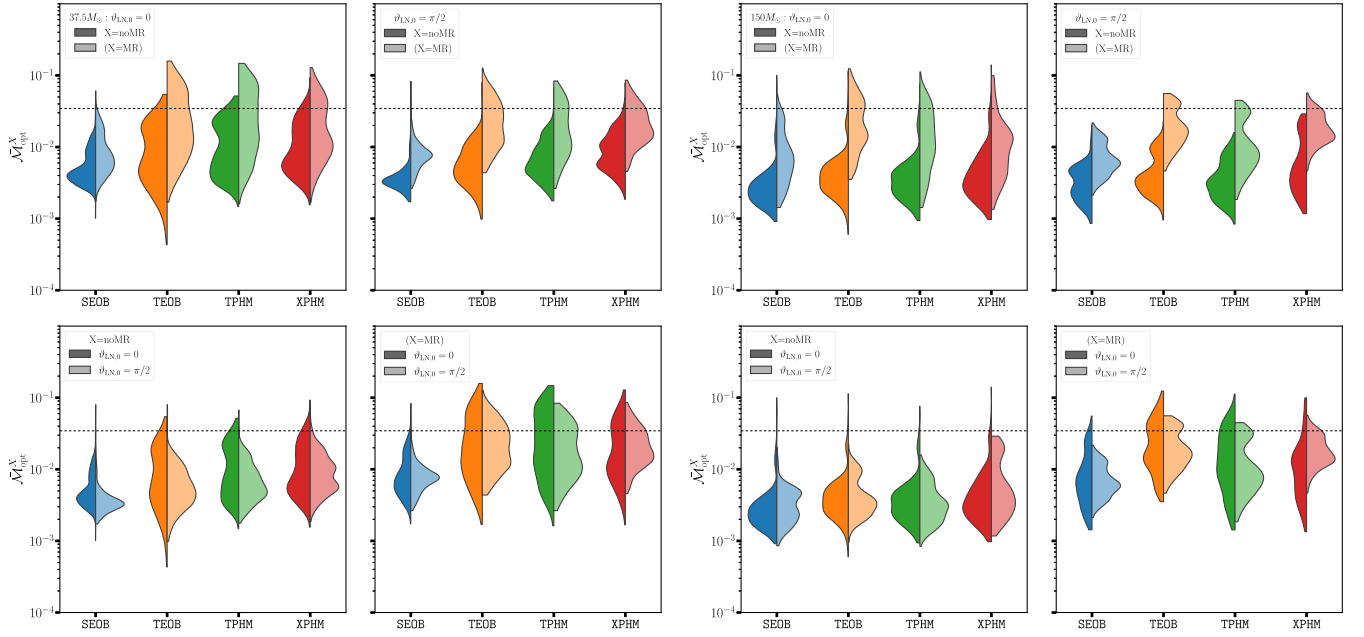


FIG. 13. Merger-ringdown truncated and full mismatches between the four approximants and the 317 short SXS waveforms shown for inclinations of 0 and $\pi/2$. The left (right) 2×2 blocks of figures correspond to light, $M = 37.5M_\odot$ (heavy, $M = 150M_\odot$) binaries. The top panels compare $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ with $\bar{\mathcal{M}}_{\text{opt}}^{\text{MR}}$ for $\vartheta_{\text{LN},0} = 0$ and $\pi/2$ and the bottom panels compare $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ with $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2)$ and $\bar{\mathcal{M}}_{\text{opt}}^{\text{MR}}(\vartheta_{\text{LN},0} = 0)$ with $\bar{\mathcal{M}}_{\text{opt}}^{\text{MR}}(\vartheta_{\text{LN},0} = \pi/2)$.

$\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2) > 0.035$, respectively for {TEOB, TPHM, XPHM}. These can be compared with the ones from the discrete set provided at the end of Sec. IV A 2.

V. FAITHFULNESS SURVEY II: COMPARISONS WITH NUMERICAL RELATIVITY WAVEFORMS

In the second part of our survey, we assess the faithfulness of {SEOB, TEOB, TPHM, XPHM} to numerical relativity waveforms from the SXS [202] catalog. Specifically, we compare against 317 short SXS waveforms and 23 long SXS waveforms of the LVCNR catalog [169], where short (long) means that the total number of GW cycles in the simulations is $\lesssim 70$ ($\gtrsim 125$). As such, the short SXS waveforms are more comparable in length to our discrete and uniform set NRSUR7DQ4 waveforms. For consistency with the previous sections, we set $M = 37.5M_\odot, 150M_\odot$ once again for the short waveform comparisons. For the long waveforms, we consider several other values for M as there are only 23 cases to compare.

A. Short sxs waveforms

We employ a specific dataset containing 317 short SXS waveforms, with $\chi_p \lesssim 0.89$, $\chi_{\text{eff}} \in [-0.6, 0.66]$, and $Q \lesssim 6$. These have between 13 and 70 GW cycles. Though there are roughly 1500 SXS BBH simulations from which we can extract waveforms, we purposefully formed a specific subset similar to that of the discrete set of Secs. IV A and IV B. Specifically, our short SXS set has

98 cases with $Q < 1.3$, 114 cases with $1.9 < Q < 2.1$, and 101 cases with $3.99 < Q < 4.01$, thus having nearly the same mass ratio split as our discrete set. As there are only three $Q \approx 5$ and one $Q \approx 6$ simulations in our set, we discuss them separately from the $313Q \lesssim 4$ cases. The relevant spin space parameters of these simulations are plotted in Fig. 2, where it can be seen that there are many cases with $\chi_p \approx 0.8$ and $|\chi_{\text{eff}}| \ll 1$. We obtained the parameters for each simulation using LALSimulation [203]⁴ which requires the recasting of SXS data into the LVCNR catalog format for which we employed the tools of the SXS package [204]. We provide the list of the 317 SXS simulations that we use in the linked GIT repository.

For the match computation, we retain the same values of $d_L = 500$ Mpc, $f_f = 1024$ Hz, and the same grid over the $\{\kappa, \phi_{\text{ref}}\}$ space. As SXS simulations start from differing frequencies, we do not employ a fixed value for f_0 , but use instead the F_LOWER_AT_1MSUN attribute [169] of each simulation's data to obtain the frequency in hertz for $1M_\odot$ then rescale for $M = 37.5M_\odot, 150M_\odot$. As before, we set $f_i = f_0 + 3$ Hz (11 Hz) in the match integral for light (heavy) systems.

The main results for zero inclination ($\vartheta_{\text{LN},0} = 0$) are shown in Fig. 12 where we plot the distributions of $\bar{\mathcal{M}}_{\text{opt}}$. What stands out the most is the superior performance of

⁴We employ the tools at <https://git.ligo.org/waveforms/lvcnrpy/-/tree/master> with public SXS data.

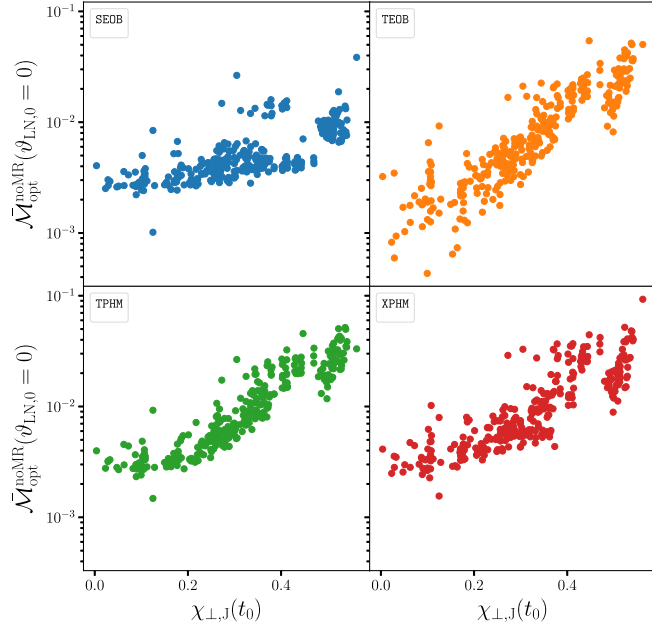


FIG. 14. Inspiral-only mismatches at $\vartheta_{\text{LN},0} = 0$, i.e., $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ between our entire $Q \lesssim 4$ short SXS set of waveforms and SEOB, TEOB, TPHM, XPHM plotted as a function of $\chi_{\perp,j}$ [Eq. (12)]. Since $\chi_{\perp,j}$ is maximized for $\theta_{1,2} \rightarrow \pi/2$, we observe increasing mismatches as the spins become more planar, i.e., for stronger precession.

SEOB as compared with the other three models. It yields only a handful of cases of $\bar{\mathcal{M}}_{\text{opt}} > 0.035$ whereas TEOB, TPHM, and XPHM have many mismatches above 0.035, especially for the light-mass BBHs, with a few cases of even $\bar{\mathcal{M}}_{\text{opt}} > 0.1$, which have $\{Q, \chi_p\} = \{1.1, 0.89\}$, $\{2, 0.85\}$, $\{4, 0.8\}$, $\{6, 0.77\}$. A quick count yields that $\{0.3, 30, 35, 27\}\%$ ($\{1, 31, 17, 9\}\%$) of the 317 cases have $\bar{\mathcal{M}}_{\text{opt}} > 0.035$ for $\{\text{SEOB, TEOB, TPHM, XPHM}\}$ for the light (heavy) binaries.

1. Effects of merger ringdown on faithfulness

As we did in Sec. IV A 1, we compare $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ with $\bar{\mathcal{M}}_{\text{opt}}$ for all four models in order to investigate how the model faithfulness is affected by the inclusion of the MR regime. When considering the inspiral-only mismatch $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$, the percentages above drop significantly, to $\{0.3, 4, 4, 5\}\%$ ($\{0.6, 1, 0.6, 1\}\%$). This drop is shown across the top row of Fig. 13 where we display the distributions for both $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and $\bar{\mathcal{M}}_{\text{opt}}$ at $\vartheta_{\text{LN},0} = 0$ and $\pi/2$ for $M = 37.5M_{\odot}, 150M_{\odot}$. As was the case with our discrete set, we observe an increased shift between the $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and $\bar{\mathcal{M}}_{\text{opt}}$ distributions as the binary mass increases from $37.5M_{\odot}$ to $150M_{\odot}$. The upward shift also increases more when the inclination goes from 0 to $\pi/2$ though less clearly so for the heavy binaries. We also observe that all four models behave extremely similarly for inspiral-only

heavy BBHs, for which we show the distributions of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ in the upper left panel of Fig. 14. This is not the case for the longer-inspiral light-mass binaries except for SEOB for which we observe very similar $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ distributions at both masses. We discuss our results for $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ further below.

Looking, as before, at the spread of the ratio $\mathcal{R} = \bar{\mathcal{M}}_{\text{opt}}/\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ over spin space, we see that $\mathcal{R} \lesssim 4$ for light SEOB, TPHM, and XPHM, with TEOB's range up to 7. For heavy BBHs, SEOB, XPHM yield $\mathcal{R} \lesssim 7$, with $\mathcal{R} \lesssim 14$ for TPHM and $\mathcal{R} \lesssim 22$ for TEOB. The near-maximal and maximal values cluster around the $\theta_{1,2} \approx \pi/2$ region of the parameter space. Moreover, the $\theta_1 > \pi/2$ half of the parameter space yields lower values for \mathcal{R} than the $\theta_1 < \pi/2$ half though with plenty of variability in the $\{\theta_1, \theta_2\}$ space. The range of \mathcal{R} shows a weaker dependence on the mass ratio here than was observed in Sec. IV A 1. This is due to the fact that we had previously used the same spin parameters at all three mass ratios, whereas in this section, the spin parameters and the mass ratios are dictated by the SXS simulations.

The mismatch distributions also exhibit some multimodality regardless of histogram binning, which is more prominent for TEOB, TPHM, and XPHM for $M = 37.5M_{\odot}$ as evident in Fig. 12. This was not the case for the $Q \leq 4$ discrete set as can be gathered from Fig. 3 though multimodalities emerged for its $Q = 4$ subset shown in Figs. 5 and 9. More multimodalities are seen in Fig. 13 in the various distributions such as TEOB and TPHM's $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2)$ for $M = 150M_{\odot}$ (lower right panel). Some of the modes seem to be due entirely to the MR part of the mismatch. For example, neither the $M = 37.5M_{\odot}$, nor the $M = 150M_{\odot}$ distributions for SEOB's $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ show any of the secondary modes present in the $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)$ distributions as seen in Fig. 12, which stand out more clearly using different histogram binning. We observe similar, MR-related modes appearing in all the distributions for $\bar{\mathcal{M}}_{\text{opt}}$ for all models. Therefore, we narrow our focus to $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ in the next few paragraphs.

2. Dependence of inspiral-only faithfulness on intrinsic parameters

As done in Sec. IV A, we investigate whether or not specific modes in the $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ distributions can be attributed specific regions in the $\{Q, \theta_1, \theta_2\}$ space. As briefly mentioned, we observe unimodal, positively skewed distributions for the heavy-mass $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ for all models as well as for the light-mass $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ SEOB distribution. However, the same distributions for light-mass $\{\text{TEOB, TPHM, XPHM}\}$ exhibit a clear secondary mode peaking between $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}} = 0.02$ and 0.03,

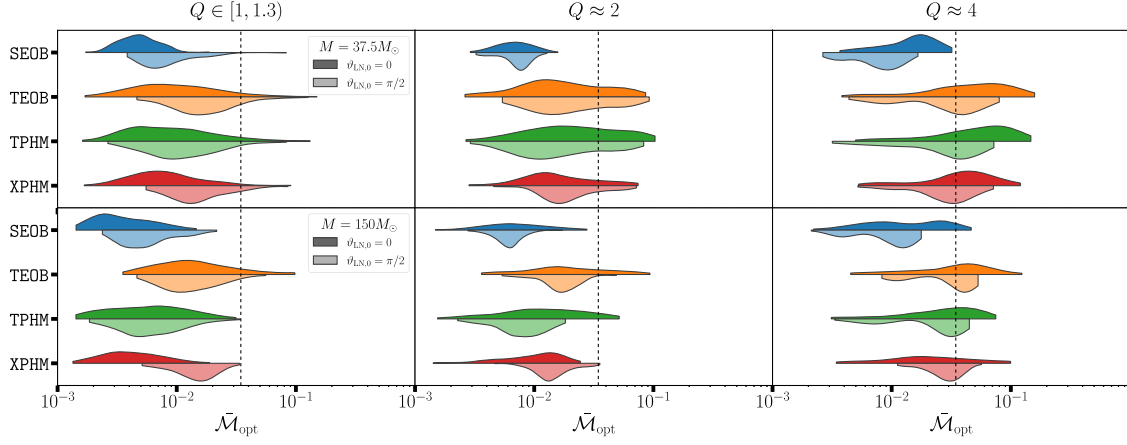


FIG. 15. Sky-optimized mismatches [Eq. (21)] between {SEOB, TEOB, TPHM, XPHM} and the 317 short SXS waveforms shown for inclinations of $\vartheta_{\text{LN},0} = 0$ and $\pi/2$, separated into three subsets in terms of the mass ratio Q . The top (bottom) panels show the mismatch distributions for binaries with total mass $M = 37.5 M_{\odot}$ ($150 M_{\odot}$).

which emerges at $Q \approx 2$, and becomes dominant at $Q \approx 4$. This behavior is consistent for {TEOB, TPHM, XPHM} indicating very similar inspiral performance for these models.

The tilt angles corresponding to the $Q \approx 2$ secondary mode all come from cases with $\theta_{1,2} \approx \pi/2, \chi_{1,2} \approx 0.8$ similar to our results in Sec. IV A 2. For $Q \approx 4$, it is mostly the cases with $\theta_1 \approx \pi/2$ and $\chi_1 \approx 0.8$ that yield the mismatches in the now-dominant mode of the $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ distributions for {TEOB, TPHM, XPHM}. There is no strong θ_2 dependence as $|\mathbf{S}_2| \ll |\mathbf{S}_1|$ for $Q \geq 4$. There is however a second smaller region of high mismatches clustered around $\theta_1 \approx 3\pi/4$. The common theme in the above analysis is that strongly precessing cases remain challenging for the models. This is even true for comparable mass ratios, where we found that the highest $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ values come from cases with the largest planar spin projections. We demonstrate this in Fig. 14, where we plot $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ vs $\chi_{\perp,J}$ for the entire $M = 37.5 M_{\odot}$, $Q \lesssim 4$ short SXS set. We opt for $\chi_{\perp,J}$ as χ_p yields many degenerate values and χ_p^{gen} correlates less clearly. Recall that $\chi_{\perp,J}$ becomes maximum for $\theta_{1,2} \rightarrow \pi/2$; thus the figure tells us that the highest mismatches come from cases with $\theta_{1,2} \approx \pi/2$.

3. Model faithfulness under increasing inclination

Returning to Fig. 13, we first focus on the bottom row where we compare $\vartheta_{\text{LN},0} = 0$ mismatches with the $\vartheta_{\text{LN},0} = \pi/2$ ones. For light BBHs, we observe once again that the distributions for $\vartheta_{\text{LN},0} = \pi/2$ are narrower, consistent with the left panels of Fig. 8. Overall the inspiral-only performance of the models is comparable modulo SEOB's superior performance for light BBHs. TEOB and TPHM yield very similar $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2)$ distributions. For light BBHs, XPHM's distribution is also very

similar to these two. For heavy BBHs, there is less similarity in the distributions. Nonetheless, the ranges of the mismatches for all models are very comparable modulo a secondary mode appearing for XPHM seen in the upper right panel of Fig. 8. A quick count yields {0.3, 0.6, 0.6, 1}% of the light mass, and {0, 0, 0, 0}% of the heavy-mass cases resulting in $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2) > 0.035$. These percentages are lower than their $\vartheta_{\text{LN},0} = 0$ counterparts. A similar count returns {0.3, 28, 23, 20}% ({0, 22, 4, 8}%) for the percentage of light (heavy) cases yielding $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2) > 0.035$, which are also lower than the corresponding $\vartheta_{\text{LN},0} = 0$ percentages.

The large offsets between the $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2)$ and $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2)$ distributions for both light and heavy masses shown in Fig. 13 are consistent with our previous results of Sec. IV B. As before, we also observe that changing inclination has a subdominant effect on model faithfulness compared to the MR portion of the waveforms. As for the ratio \mathcal{R} , we observe that $\mathcal{R} \lesssim \{3, 8, 5, 5\}$ for the light and $\lesssim \{5, 10, 9, 14\}$ for the heavy {SEOB, TEOB, TPHM, XPHM} mismatches. Comparing these with the $\vartheta_{\text{LN},0} = 0$ ratios of Sec. VA 1 tells us that \mathcal{R} 's range has barely changed for the light binaries. In the case of the heavy binaries, \mathcal{R} 's maximal values have decreased for SEOB, TEOB, and TPHM, while it has increased for XPHM. The large changes in the maximal values of \mathcal{R} are due to a few cases, each of which might be challenging to a particular model, while not to others, and vice versa.

The $\theta_{1,2} \approx \pi/2$ region still results in larger values of \mathcal{R} . And the $\theta_1 > \pi/2$ half of the parameter space still yields lower values for \mathcal{R} than the $\theta_1 < \pi/2$ half though this is less distinct than it was for $\vartheta_{\text{LN},0} = 0$, and with XPHM violating this trend with the region $\theta_{1,2} \gtrsim 2\pi/3$ also returning large values of \mathcal{R} . Consistent with our previous findings, the maximal values of \mathcal{R} do not vary much with changing mass ratio.

A striking feature in Fig. 13 is the prominent bimodalities in the $\vartheta_{\text{LN},0} = \pi/2$ mismatch distributions for $M = 150M_{\odot}$, which are not present for $M = 37.5M_{\odot}$. This bimodality is most prominent for the heavy-mass TEOB and TPHM mismatches, and is seen, albeit less strongly, for the heavy SEOB and the light TEOB, TPHM mismatches (referring specifically to $\bar{\mathcal{M}}_{\text{opt}}$). Decomposing these mismatches further into their mass ratio subsets as shown in Fig. 15 clarifies that the bimodalities are due to the mismatches clustering into two separate regions for $Q \lesssim 2$ vs $Q \approx 4$. We had already observed a similar behavior for the $\vartheta_{\text{LN},0} = 0$ mismatches as Q increased from 1 to 4, which we replot in Fig. 15, mirroring our $\vartheta_{\text{LN},0} = \pi/2$ results. The cases causing these large mismatches have $\theta_1 \in [1.45, 1.6]$ and $\chi_p \approx 0.8$, which, for $Q \approx 4$, result in values of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = \pi/2)$ much larger than the ones coming from the $Q \lesssim 2$ cases. These inspiral-only mismatches further degrade by up to an order of magnitude when we consider the full mismatches $\bar{\mathcal{M}}_{\text{opt}}$. Note that the values of θ_1, θ_2 do not alone explain the increased mismatches for $Q \approx 4$ as the $Q \approx 2$ subset has even more cases with $\theta_{1,2} \approx \pi/2$. The number of cycles do not provide a satisfactory explanation either as our chosen $Q \approx 4$ simulations are comparable in length to the $Q \approx 2$ subset.

Another interesting feature of Fig. 15 is the fact that for $Q \approx 4$ the $\vartheta_{\text{LN},0} = \pi/2$ mismatches are mostly lower than the $\vartheta_{\text{LN},0} = 0$ ones. Specifically, for all light BBHs and the heavy SEOB and TPHM, we clearly have $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2) \leq \bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)$. Focusing on the $Q \approx 4$ SEOB subsets alone, we confirmed that the improved faithfulness at $\vartheta_{\text{LN},0} = \pi/2$ is due to the arguments made in Sec. IV B; i.e., for cases with $|h_{21}| \gtrsim |h_{22}|$ and $\bar{\mathcal{M}}_{21} \lesssim \bar{\mathcal{M}}_{22}$, we see $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2) \leq \bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)$ for the distributions. In short, as we had observed in Sec. IV B, when the $h_{2,\pm 1}$ multipoles are more NR faithful than $h_{2,\pm 2}$ and their amplitudes are comparable to $|h_{2,\pm 2}|$, the resulting high-inclination strain matches are mostly better than zero-inclination matches.

4. Effects of coprocessing multipoles on faithfulness

As we had done in Sec. IV A 3, we investigate once again whether or not the unfaithfulness of the AS (coprocessing) multipoles have any effect on model performance. As before, we find a correlation between $\log(\bar{\mathcal{M}}_{22,\text{AS}})$ and $\log(\bar{\mathcal{M}}(\vartheta_{\text{LN},0} = 0))$, where $\bar{\mathcal{M}}_{22,\text{AS}}$ is the mismatch between the SXS coprocessing (2,2) multipole⁵ and each model's AS (2,2) multipole. The correlation for our short SXS set is stronger than the discrete set of Sec. IV A and persists whether or not we consider $\log \bar{\mathcal{M}}$ or $\log \bar{\mathcal{M}}_{\text{opt}}$ or

⁵We employ the SCRIP package [205–208] to extract the SXS coprocessing multipoles.

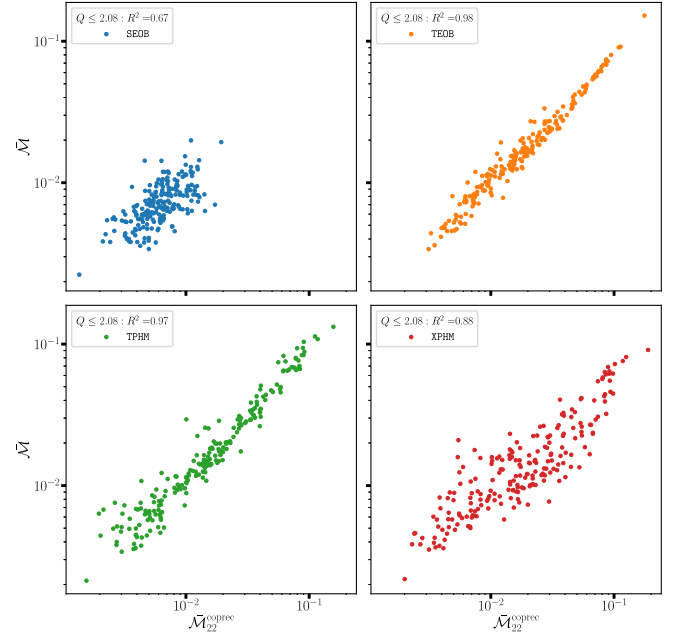


FIG. 16. The correlation between $\bar{\mathcal{M}}_{22,\text{AS}}$ and $\bar{\mathcal{M}}(\vartheta_{\text{LN},0} = 0)$ for the $M = 37.5M_{\odot}$, $Q \lesssim 2$ cases of our short SXS set. $\bar{\mathcal{M}}_{22,\text{AS}}$ is the mismatch [Eq. (13)] between the SXS coprocessing (2,2) multipole and each AS (2,2) multipole of the four models. Each legend also gives the R^2 value for the log-log datasets. This figure is akin to the top panel of Fig. 7.

inspiral-only vs full mismatches. It is also seen for the $M = 150M_{\odot}$ waveforms, which for the SXS set are in general longer than the discrete set, which were limited by NRSUR7DQ4's length restrictions. As before, the correlation is strongest for the $M = 37.5M_{\odot}$ TEOB and TPHM mismatches, but also persists for SEOB and XPHM as can be seen in Fig. 16, where we plot only the data from the $Q \lesssim 2$ subset. Thus, this figure can be directly compared with the top row of Fig. 7. In Fig. 16, we also provide the R^2 values for the log-log scatter plots, which for TEOB and TPHM are near unity, and equals almost 0.9 for XPHM.

Following Sec. IV A 3, we additionally compute the slopes of the linear fits to $\bar{\mathcal{M}}(\vartheta_{\text{LN},0} = 0)$ as a function of $\bar{\mathcal{M}}_{22,\text{AS}}$ for TEOB, TPHM, and XPHM. We find that the entire $Q \lesssim 4$ set of $\bar{\mathcal{M}}(\vartheta_{\text{LN},0} = 0)$ (313 out of 317 cases) for TEOB and TPHM can be fit by lines with slopes of approximately 0.9, and 0.5 for XPHM. This is consistent with our findings in Sec. IV A 3 that once the AS (2, ± 2) multipole unfaithfulness exceeds a certain value, it becomes the dominant source of unfaithfulness for the strain mismatch.

Since the (2, ± 2) AS/coprocessing multipoles also make up the dominant contribution to the precessing (2, ± 1) multipoles, we expect the correlation to persist for $\bar{\mathcal{M}}(\vartheta_{\text{LN},0} = \pi/2)$, albeit less strongly. Indeed, the R^2 values provided in Fig. 16 drop to $\{0.2, 0.77, 0.85, 0.6\}$, respectively for four models. And the slopes of the linear

fits are comparable. Finally, as we have already seen in Sec. IV A 3 the AS ($2, \pm 1$) multipoles do not yield a similar correlation: Their magnitudes are too small. As done in Sec. IV A 3, we also checked how much these modes matter for strain faithfulness. Specifically, we reconstructed the SXS strains without the $h_{2,\pm 1}^{\text{coprec}}$ content and computed the mismatch to the unaltered SXS strains. As before, we found that the mismatch is at most 0.015 (0.07) at $\vartheta_{\text{LN},0} = 0$ ($\vartheta_{\text{LN},0} = \pi/2$). At both inclinations, it is only few cases that yield such mismatches with the majority of the values below 0.001 (0.01).

5. $Q \geq 5$ comparisons

As already noted, we only have four cases with $Q \geq 5$ in our short SXS set. Three of these, namely SXS:0057, 0058, 0062, have $Q \approx 5$, $-0.18 \lesssim \chi_{\text{eff}} \lesssim 0$ and $0.45 \lesssim \chi_p \lesssim 0.5$, and one, SXS:0165, has $Q \approx 6$, $\chi_{\text{eff}} \approx -0.45$, $\chi_p \approx 0.77$. We should caution that we detected “V-shaped” kinks in the phases of both the precessing and coprecessing ($2,2$), ($2,1$) multipoles. The first kinks appear at $\approx 100M$ before the peak of the multipole amplitude. Additionally, SXS:0057 is very short; i.e., its inspiral lasts $\approx 800M$. For these reasons, we present only the full wavelength mismatch, $\bar{\mathcal{M}}_{\text{opt}}$ for these cases.

We list the values of $\bar{\mathcal{M}}_{\text{opt}}$ for $M = 37.5M_{\odot}, 150M_{\odot}$, and $\vartheta_{\text{LN},0} = 0, \pi/2$ for each model in Table II. One interesting finding is that $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2) < \bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)$ for all models except for the $M = 150M_{\odot}$ mismatches for SXS:0058. Most of the improved faithfulness at $\vartheta_{\text{LN},0} = \pi/2$ is due to a more faithful h_{21} multipole satisfying the relation $|h_{21}| \gtrsim |h_{22}|$ as we have previously explored in detail. As we have only four cases with $Q \gtrsim 5$, we move on to comparisons of longer waveforms.

B. Long SXS waveforms

For this assessment, we employ 23 long SXS waveforms, namely simulations 1389–1411 [141]. The total number of GW cycles in these simulations varies approximately between 128 and 305, and the relevant parameters have the following ranges: $Q \leq 4$, $\chi_p \lesssim 0.49$, $\chi_{\text{eff}} \in [-0.2, 0.3]$ though only 1409–1411 have $Q = 4$ with the rest having $Q < 2$. With only 23 cases as compared with $\mathcal{O}(10^3)$ in Sec. IV and $\mathcal{O}(10^2)$ in Sec. VA, we could afford the computational time to expand our total mass sample to the following values: $M = \{37.5, 75, 112.5, 150, 187.5, 225\}M_{\odot}$, with $f_i = f_0 + 3$ Hz for all match integrals. Note that because the amplitude spectral density drops steeply from $\mathcal{O}(10)$ to 20 Hz (see Fig. 1), only one to two inspiral cycles contribute to the match integrals for the $M = 225M_{\odot}$ cases even though $f_i \in [4, 5]$ Hz for these. Similarly, we record 2.5 to 4 inspiral cycles for $M = 187.5M_{\odot}$ and 4.5 to 7 cycles for $M = 150M_{\odot}$.

TABLE II. The sky-optimized mismatches between the four $Q > 4$ SXS waveforms and the four models that we consider in this work.

SXS ID	Model	$\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)$		$\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2)$	
		$37.5M_{\odot}$	$150M_{\odot}$	$37.5M_{\odot}$	$150M_{\odot}$
0057	SEOB	0.027	0.032	0.011	0.0076
	TEOB	0.024	0.038	0.015	0.010
	TPHM	0.029	0.032	0.012	0.0081
	XPHM	0.036	0.038	0.015	0.0098
0058	SEOB	0.0069	0.0041	0.0055	0.0050
	TEOB	0.022	0.014	0.018	0.019
	TPHM	0.026	0.0087	0.013	0.0086
	XPHM	0.012	0.010	0.013	0.018
0062	SEOB	0.014	0.019	0.0039	0.0026
	TEOB	0.020	0.038	0.0079	0.0076
	TPHM	0.016	0.023	0.0051	0.0045
	XPHM	0.025	0.047	0.0062	0.0077
0165	SEOB	0.061	0.055	0.017	0.016
	TEOB	0.088	0.076	0.028	0.023
	TPHM	0.063	0.11	0.028	0.027
	XPHM	0.13	0.093	0.080	0.057

Accordingly, we observe that (MR-only SNR) \gtrsim (inspiral-only SNR) for $M \gtrsim 175M_{\odot}$.

We present the resulting mismatches as trace plots in Fig. 17. In the top two rows of the figure, we display $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ at $\vartheta_{\text{LN},0} = 0$ and $\pi/2$ then repeat this for $\bar{\mathcal{M}}_{\text{opt}}$ in the bottom two rows with the four columns corresponding to the four approximants in the usual order. First, we note that with the exception of one TPHM $\vartheta_{\text{LN},0} = \pi/2$ case (corresponding to SXS:1395), all mismatches are less than 0.035. Moreover, we also observe that $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0) < 0.01$ holds nearly universally and that this inequality still holds at $\vartheta_{\text{LN},0} = \pi/2$ for the EOB models. Our finding of $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}} \leq \bar{\mathcal{M}}_{\text{opt}}$ from the previous sections persists here as well, which can be gathered by comparing row one with row three, and row two with row four. The dependence of the mismatches on inclination also seems to be consistent with our previous results as the relation $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0) \leq \bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2)$ is retained as seen by comparing row one with row two, and row three with row four of the same figure.

As for the dependence of the mismatches on the total mass, we observe mostly a flattening of each mismatch curve as M exceeds a certain threshold, more strongly so for the $\vartheta_{\text{LN},0} = 0$ cases. This flattening is consistent with the results of, e.g., Refs. [70,98,110,131]. Interestingly we observe it for both $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and $\bar{\mathcal{M}}_{\text{opt}}$. For the flattening of the former, our reasoning is as follows: As the binary gets heavier, the signal shifts toward lower frequencies, e.g., $f_{\text{peak}} \lesssim 50$ Hz for $M = 150M_{\odot}$ and $f_{\text{peak}} \lesssim 30$ Hz for $M = 225M_{\odot}$. In this regime, the detectors are less sensitive

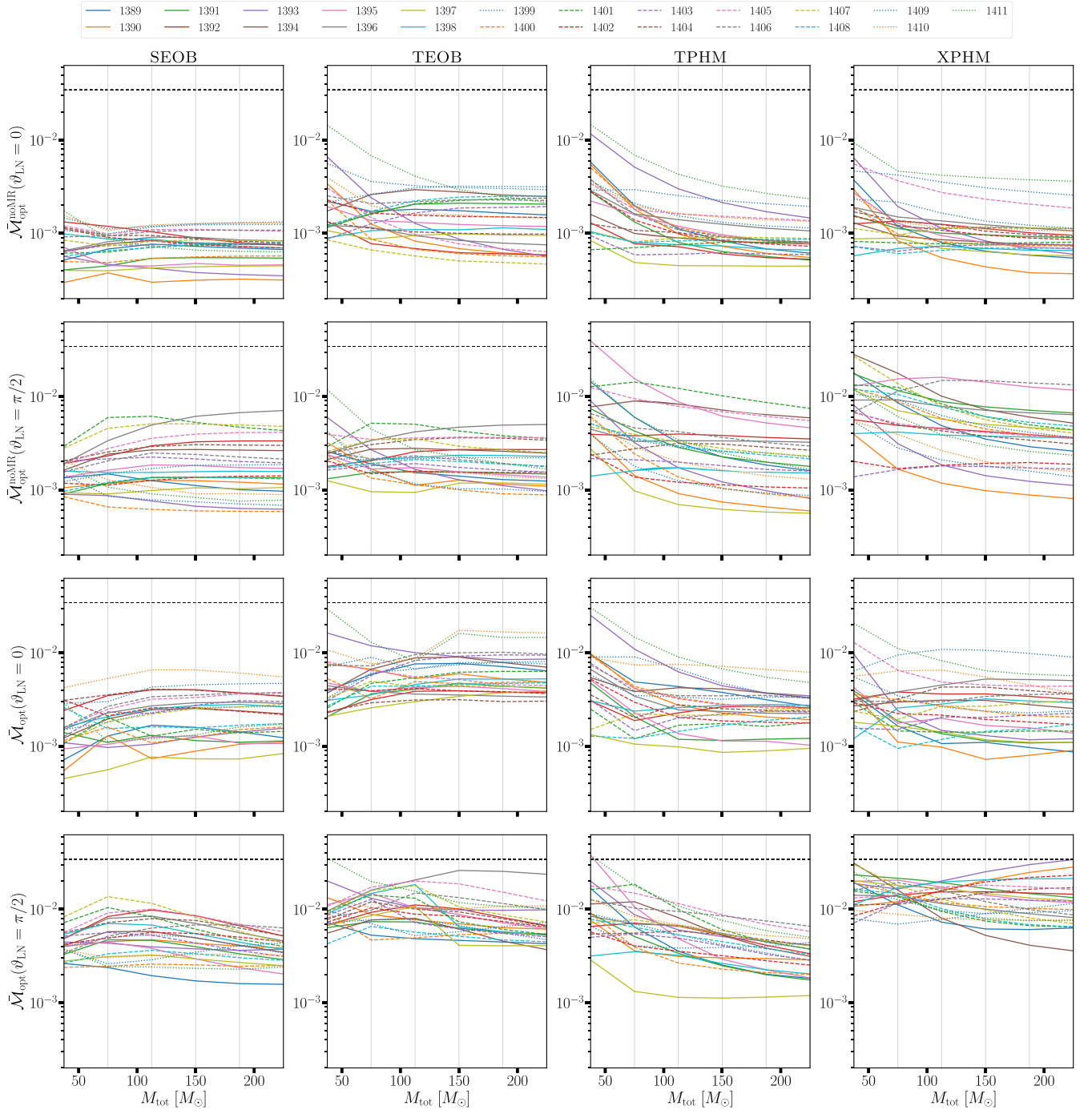


FIG. 17. Trace plots of the mismatches between {SEOB, TEOB, TPHM, XPHM} and numerical relativity waveforms from 23 long SXS simulations as a function of the total binary mass. We computed the mismatches at the values of $M = \{37.5, 75, 112.5, 150, 187.5, 225\}M_{\odot}$ (marked by the faint vertical lines) and linearly connected the data points. The top (bottom) two rows show $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ ($\bar{\mathcal{M}}_{\text{opt}}$) with odd (even) rows corresponding to $\vartheta_{\text{LN},0} = 0$ ($\vartheta_{\text{LN},0} = \pi/2$) results. The four columns represent, from left to right, the results for SEOB, TEOB, TPHM, XPHM.

(see Fig. 1); therefore the mismatch expression is also less sensitive to the differences between waveforms. So the dependence on the total mass becomes irrelevant for $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ beyond a certain threshold.

The flattening of the $\bar{\mathcal{M}}_{\text{opt}}$ curves are partly due to this reason combined with the fact that for the cases with only a few inspiral cycles, i.e., $M \gtrsim 150M_{\odot}$, the MR part of the signal becomes the dominant contribution to the mismatch.

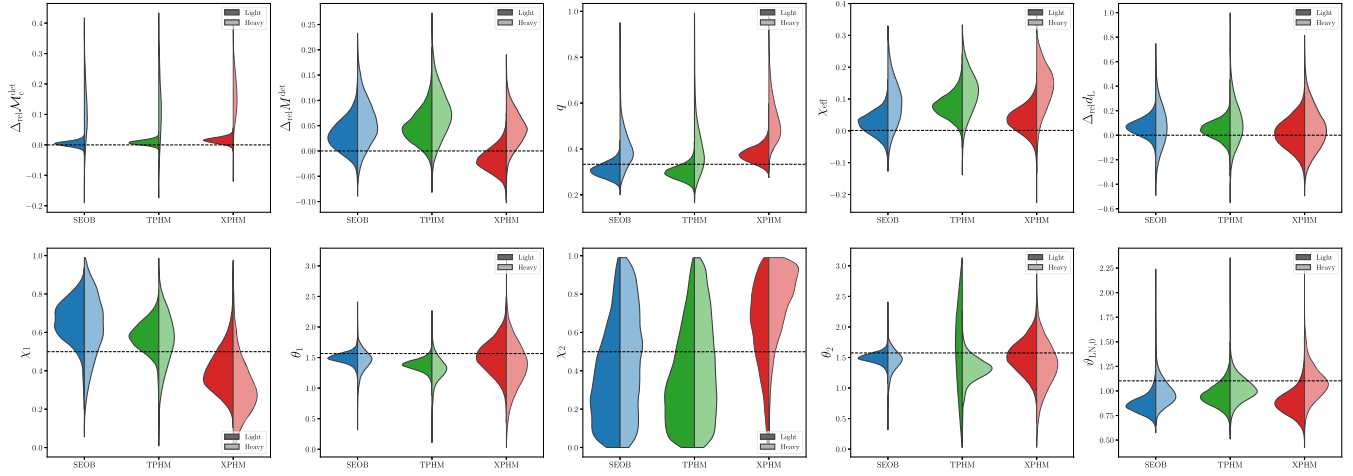


FIG. 18. Recovery performances of SEOB, TPHM, and XPHM for a synthetic signal generated by a “zero-noise” injection of the sxs simulation 0050 into the O4a LIGO Hanford-Livingston network. In each panel, we present the results from both the light- and the heavy-mass runs as the darker and lighter halves of each violin, respectively. Shown are the recovered posteriors for $\{\mathcal{M}_c^{\text{det}}, M^{\text{det}}, q, \chi_{\text{eff}}, d_L, \chi_1, \theta_1, \chi_2, \theta_2, \vartheta_{\text{LN},0}\}$ with the dashed horizontal lines marking the injected values. For $\{\mathcal{M}_c^{\text{det}}, M^{\text{det}}, d_L\}$, we present posteriors for the relative difference between the injected and the recovered values, i.e., $\Delta_{\text{rel}}X := (X_r - X_{\text{inj}})/X_r$. The first panel, i.e., $\Delta_{\text{rel}}\mathcal{M}_c^{\text{det}}$, is somewhat hard to read because the models produced very narrow posteriors for the light-mass run, but much wider ones for the heavy one.

In other words, the contribution to the mismatch from $\lesssim 5$ inspiral cycles does not change the overall $\bar{\mathcal{M}}_{\text{opt}}$ significantly enough. However, the MR part of some of the heavier cases falls in the $\mathcal{O}(100)$ Hz region, where the detectors are more sensitive, as is, accordingly, our full-band mismatch $\bar{\mathcal{M}}_{\text{opt}}$. This is why we observe flatter curves for $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}(\vartheta_{\text{LN},0} = 0)$ than for $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)$ in Fig. 17. Also note that some curves do not flatten out at all. We are unable to provide a single universal, “one size fits all” explanation for these as the shape and mass dependence of each curve varies from one model to another.

One final prominent feature of the figure is the change in XPHM’s faithfulness, $\bar{\mathcal{M}}_{\text{opt}}$, in going from $\vartheta_{\text{LN},0} = 0$ to $\pi/2$, for which we see that the values of the mismatches shift from mostly being below 0.01 to mostly being above it, while still remaining below 0.035. This behavior might be indicative of more severe MR-related issues for XPHM’s precessing $(2, \pm 1)$ multipoles than for other models. We recomputed $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2)$ using the updated SPINTAYLOR version of XPHM [136] which showed improvement: The range of $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2)$ dropped from $[0.004, 0.03]$ to $[0.003, 0.02]$. More noticeably, the $M \gtrsim 100M_\odot$ values of $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2)$ obtained with the SPINTAYLOR version are all $\lesssim 0.02$ in contrast to many cases of $0.02 < \bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = \pi/2) \lesssim 0.035$ seen in the lower right panel of Fig. 17.

VI. INJECTION/RECOVERY STUDY

As the final part of our survey, we investigate the performance of the precessing approximants in an

injection-recovery PE study. Specifically, we inject two different sxs ($\ell = 2$)-only strains into LIGO noise and recover the parameters of the injected waveforms with SEOB, TPHM, and XPHM. We are grateful to Charlie Hoy and Lorenzo Pompili for assisting us with the SEOB runs. Our omission of the TEOB results is due to (i) the precessing model can only be used for PE with the BAJES library [209] which is external to the computational infrastructure that we employed for the PE runs with $\{\text{SEOB, TPHM, XPHM}\}$; (ii) BAJES runs crashed with waveform errors when the upper bounds of the $\chi_{1,2}$ priors were above 0.8 despite TEOB generating reasonable waveforms in that regime. This error first arose when attempting to analyze GW200129_065458 with TEOB and persisted when injection/recovery runs were carried out. We hope to collaborate with BAJES developers in the future so that these issues can be resolved.

Returning to our study here. We performed zero-noise injections of the NR waveforms sxs:0050 (henceforth 0050) and sxs:0628 (henceforth 0628) while using an estimated O4a PSD [210] for our likelihood computations. These simulations have key parameters $\{Q, \chi_{\text{eff}}, \max(\chi_p)\} \approx \{3, 0.001, 0.5\}, \{2, -0.174, 0.84\}$ respectively. The former has a higher component mass asymmetry and moderate spins, only in the orbital plane, whereas the latter has a moderate mass ratio, but significant precession. We perform each injection twice: for a low mass and a high mass BBH given in the detector frame, specifically, $M^{\text{det}} = 56.5M_\odot, 150M_\odot$ for 0050 and $M^{\text{det}} = 67M_\odot, 150M_\odot$ for 0628. The values for the lighter masses are chosen so that the sxs reference frequency corresponds to $f_0^{\text{det}} \approx 20$ Hz for both simulations. The values for the

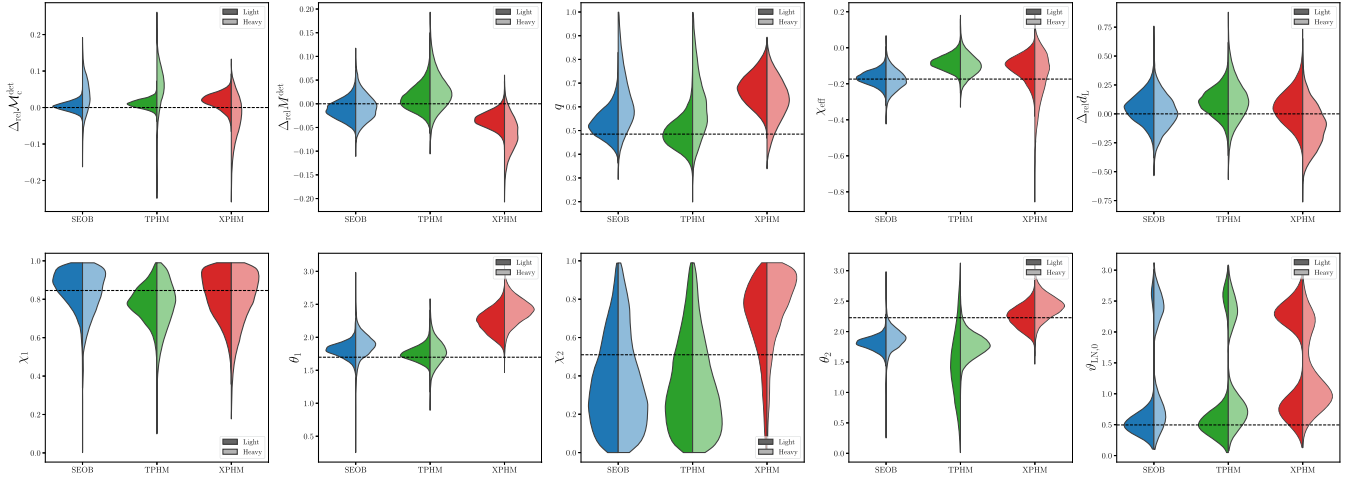


FIG. 19. Same as Fig. 18, but for the sxs simulation 0628.

heavier masses are chosen to put more emphasis on the plunge-merger-ringdown portions of the waveforms, consistent with the rest of this article. The source-frame masses are smaller by a factor of $(1+z)$ where z is the redshift of the sources at the injected luminosity distance which we set to $\{\{308, 649\}, \{555, 1020\}\}$ Mpc, respectively for light/heavy 0050 and light/heavy 0628.

The reason why the value of d_L for the $150M_\odot$ 0050 run differs from that of 0628 is due to our choice of extrinsic parameters. Specifically, we fine-tuned $\{\vartheta_{\text{LN},0}, \theta_s, \phi_s, \psi_s, d_L\}$ such that the total network SNR, ρ_{tot} , accumulated by the H1-L1 network is 40 for each case resulting in a precession SNR of $\rho_p = \{\{11, 6.4\}, \{15, 10\}\}$, respectively for light/heavy 0050 and light/heavy 0628. This quantity was first defined in Ref. [211] and represents the contribution to the SNR coming from the second most significant precessing harmonic. It is proportional to the total SNR, thus becomes larger for louder events, and has to be greater than 2.1 for the effects of precession on the signal to not be attributed to noise alone [211]. Let us add that we opted here for $\rho_{\text{tot}} = 40$ as our trial runs with $\rho_{\text{tot}} = 20$ resulted in the posteriors of precession-related quantities, such as χ_p , being prior dominated, hence uninformative. The specific values for all our injected parameters can be found in the config files uploaded to our GIT repository.

We employ the BILBY [212,213] and BILBY_PIPE [214] modules with the DYNESTY sampler [215–218]. We set the number of live points, NLIVE, to 1000, and NACCEPT to 60. Other DYNESTY attributes are set to their default values. When employing the IMRPHENOM models, we run two parallel chains and combine the results at the end. For SEOB, we run only one chain per injection.

The priors for the parameters are as follows. For the lighter mass runs, we use the BILBY function UNIFORMINCOMPONENTSCHIRPMASS with the range $\mathcal{M}_c \in [10, 35]M_\odot$, the UNIFORMINCOMPONENTSMASSRATIO function with

range $q \in [0.083, 1]$, and constraints of $m_i \in [1, 1000]M_\odot$. The luminosity distance prior range is given by $d_L \in [100, 5000]$ Mpc via the function UNIFORMSOURCEFRAME using Λ CDM cosmology parametrized by Planck 2015 data [219]. The spin magnitudes are uniform in the range $\chi_i \in [0, 0.99]$ while the spin tilt angles are uniform in their sines, $\sin \theta_i \in [0, 1]$, and the **J** frame inclination angle is uniform in cosine, $\cos \vartheta_{\text{JN},0} \in [-1, 1]$, where $\vartheta_{\text{JN},0}$ is the angle between **J** and $\hat{\mathbf{N}}$ at $f = f_0$. Other angles are all uniform in their respective ranges. For the heavier $M^{\text{det}} = 150M_\odot$ runs, we set the chirp mass prior range to $\mathcal{M}_c \in [45, 85]M_\odot$ ($[40, 90]M_\odot$) for 0050 (0628) with the other priors kept the same. We find that such an injection/recovery PE “job” consisting of two chains of IMRPHENOMTPHM and two chains of IMRPHENOMXPHM takes roughly one day using 100 processors. With the same resources, one SEOBNRV5PHM chain can be obtained in ~ 1.5 days.

A. Main results

Overall, we have 12 sets of posterior parameters to present resulting from $\{\text{SEOB, TPHM, XPHM}\} \times \{0050, 0628\} \times \{\text{light, heavy}\}$ runs. For the sake of brevity, we show only a subset of posteriors for important intrinsic and extrinsic parameters, namely the set $\{\mathcal{M}_c^{\text{det}}, M^{\text{det}}, q, \chi_{\text{eff}}, d_L, \chi_1, \theta_1, \chi_2, \theta_2, \vartheta_{\text{LN},0}\}$ in Figs. 18 and 19, where, for $\{\mathcal{M}_c^{\text{det}}, M^{\text{det}}, d_L\}$ we present the relative differences between the injected and the recovered values. When relevant, we also discuss the posteriors for other parameters besides these. Table IV presents the 5th, 50th, and 95th percentiles of the recovered posteriors for the relevant parameters. We discuss the recovery of precession-related quantities further below in Sec. VI B. For reference, we display in Table III, the mismatches between the injected waveforms and those generated by SEOB, TPHM, and XPHM using the injected parameters.

We start with Fig. 18, where we present both the light- and heavy-mass posteriors from the 0050 recovery runs as

TABLE III. Mismatches calculated between the injected SXS waveforms and the recovery models evaluated at the same parameters as the injected waveforms. The values presented here are calculated using Eq. (13).

Model	0050 light	0050 heavy	0628 light	0628 heavy
SEOB	1.28×10^{-2}	1.39×10^{-2}	1.05×10^{-2}	7.94×10^{-3}
TPHM	2.13×10^{-2}	1.00×10^{-2}	1.75×10^{-2}	9.81×10^{-3}
XPHM	2.43×10^{-2}	1.73×10^{-2}	1.67×10^{-2}	1.63×10^{-2}

violin plots. A common feature seems to be that SEOB and TPHM posteriors are more similar to each other than XPHM. In general, one model recovers a particular set of parameters better than the other two. For example, for the light-mass case, SEOB recovers $\{\mathcal{M}_c^{\text{det}}, q, \chi_{\text{eff}}\}$ better, whereas XPHM recovers $\{\chi_1, \theta_1, \theta_2, d_L\}$ better. TPHM recovers $\vartheta_{\text{LN},0}$ better than the other two models, and is sometimes better than SEOB or XPHM for other parameters, but not both. χ_2 is recovered at roughly the same confidence by the models though the XPHM posterior is skewed oppositely. Looking at the median values and the 90% confidence intervals (CIs) for m_1, m_2 , i.e., the source-frame component masses, in Table IV, we see SEOB/TPHM slightly overestimating m_1 while underestimating m_2 with XPHM behaving in the opposite manner. Additional corner plots of the posteriors can be found in our GIT repository.

SEOB's and TPHM's performances are even more similar for the heavy-mass case. SEOB/TPHM recover $\{\mathcal{M}_c^{\text{det}}, q, \chi_1\}$ better, whereas XPHM does so for $\{\theta_1, \theta_2, d_L, \vartheta_{\text{LN},0}\}$. χ_2 posteriors for SEOB/TPHM look nearly flat indicating the dominance of the uniform prior in the posterior, whereas XPHM's posterior rails against the $\chi = 1$ Kerr bound. $\{M^{\text{det}}, \chi_{\text{eff}}\}$ are recovered equally marginally by SEOB and XPHM. The recovery posteriors for $\mathcal{M}_c^{\text{det}}$ are rather wide with the 90% CIs ranging roughly from $54M_\odot$ to $68M_\odot$ for SEOB, $54M_\odot$ to $70M_\odot$ for TPHM, and from $58M_\odot$ to $70M_\odot$ for XPHM, given an injected value of $55M_\odot$.

The recovery of 0628's parameters reveals yet more differences in model performance, but with the similarity in SEOB and TPHM's posteriors persisting. Let us recall that 0628 has stronger precession than 0050, but nonzero parallel spin projections. For the light-mass run, SEOB recovers $\{\mathcal{M}_c^{\text{det}}, \chi_{\text{eff}}\}$ better while TPHM does so for $\{M^{\text{det}}, q, \theta_1\}$ with XPHM recovering θ_2 better. SEOB, TPHM recover $\vartheta_{\text{LN},0}$ very similarly, and SEOB, XPHM do so for d_L . SEOB/XPHM both rail against the $\chi_1 = 1$ bound, but not TPHM. On the other hand, SEOB/TPHM rail against the $\chi_2 = 0$ bound, with XPHM somewhat railing against $\chi_2 = 1$. Though XPHM recovers $\mathcal{M}_c^{\text{det}}$ in the 90% CI, its recovery of the other mass parameters is rather poor. For example, its posteriors for both the detector and the source-frame (total, primary, secondary) masses do not contain the injected parameters within their respective 90% CIs, whereas SEOB's and TPHM's posteriors do, with the

injected values recovered near the peaks of the distributions. This can also be partly gathered by looking at the recovered $\{m_1, m_2\}$ posteriors in Table IV under the 0628 Light rows. The fact XPHM under(over)estimates $m_1(m_2)$ is why it recovers \mathcal{M}_c within its 90% CI.

As for the heavy-mass 0628, SEOB/TPHM recover both the detector and source-frame masses robustly. XPHM underestimates the injected values for m_1, m_1^{det} by 3 standard deviations or more. Despite this, its q posteriors are comparable to that of SEOB's. Overall, SEOB recovers $\{M^{\text{det}}, \chi_{\text{eff}}, d_L\}$ better, while TPHM does so for $\{q, \chi_1, \theta_1\}$, and with XPHM doing so for $\{\mathcal{M}_c^{\text{det}}, \theta_2\}$, but cannot recover θ_1 with any confidence. The various railings of the $\chi_{1,2}$ posteriors against the $\chi_{1,2} = 0, 1$ bounds, seen for the light-mass injection, are exacerbated. As was the case with 0050, the recovered posteriors for $\mathcal{M}_c^{\text{det}}$ are wider for the heavy-mass run with the 90% CIs ranging roughly from $58M_\odot$ to $67M_\odot$ for SEOB, $59M_\odot$ to $69M_\odot$ for TPHM, and from $53M_\odot$ to $64M_\odot$ for XPHM for an injected value of $60.5M_\odot$. Another prominent feature to note is the bimodality in the $\vartheta_{\text{LN},0}$ posteriors which we did not observe for 0050. Given that we only inject and recover the $\ell = 2$ strain, some bimodality in the inclination is expected due to the distance-inclination degeneracy. However, when there is precession, the $(2, \pm 1)$ multipoles may contain enough power to break this, and more so for larger Q . Indeed, we find that the ratio $|-2Y_{21}h_{21}|/|-2Y_{22}h_{22}|$ is larger for 0050 than for 0628, thus explaining why we observe no bimodality in the inclination posteriors for 0050.

An interesting result that warrants further discussion is the consistent overestimation of χ_{eff} for the 0050 run where the injection has $\chi_{\text{eff}} = 0.001$. This can be seen in the upper right panel of Fig. 18 (also upper left panels of Figs. 20, 21) and in Table IV. This is consistent with the broad findings of Ref. [220]. Given that the total SNRs are fixed to 40 with precession SNRs all exceeding 6, this is somewhat surprising. Perhaps, this may be due to the fact that the injected value is small. Interestingly, Appendix D of Ref. [220] contradicts us, mostly finding that the posteriors slightly underestimate the injected values of $\chi_{\text{eff}} = 0$. We believe the disagreement is mostly due to the facts that (i) they set $\theta_{1,2} = 10^\circ, 30^\circ$ thus have to fix $\chi_{1,2} = 0$ to obtain $\chi_{\text{eff}} = 0$, which is different from our situation which has $\theta_{1,2} \simeq 90^\circ$; (ii) they employ the IMRPHENOMPV2 model which is no longer considered the state of the art. It might be case that the recovery of small injected values of χ_{eff} with large χ_p is in general more challenging. Otherwise, we can expect χ_{eff} to be well recovered by the current models as seen in, e.g., Table I of Ref. [131] for TPHM and Table II of Ref. [128] for XPHM. A systematic injection/recovery campaign would certainly find a trend (if there is any). Let us reiterate that χ_{eff} is an effective measure of the impact of spin in the inspiral, which may not necessarily hold for high-mass systems, i.e., few inspiral cycles. Indeed, we observe that the

TABLE IV. The results of our injection/recovery parameter estimation runs. For a selected subset of parameters given in the first row, we present the medians and {5th, 95th} percentile error bars for the posteriors recovered by SEOB, TPHM, and XPHM for four injections of sxs numerical relativity waveforms into LIGO Hanford-Livingston O4a sensitivity. The labels 0050, 0628 in the first column are the sxs simulation numbers. Light and Heavy refer to the total mass of the binary black hole system. The row label “inj.” denotes the injected value. The {5th, 95th} error bars are simply the following percentile differences: {5th-50th, 95th-50th}.

	$m_1(M_\odot)$	$m_2(M_\odot)$	χ_1	χ_2	χ_{eff}	χ_p	θ_1	θ_2	d_L (Mpc)	$\vartheta_{\text{LN},0}$
0050 Light										
inj.	39.74	13.25	0.50	0.50	0.001	0.50	1.57	1.57	308	1.05
SEOB	41.56 ^{+3.79} _{-3.21}	12.75 ^{+1.08} _{-0.93}	0.66 ^{+0.16} _{-0.18}	0.32 ^{+0.48} _{-0.29}	0.023 ^{+0.056} _{-0.059}	0.66 ^{+0.16} _{-0.18}	1.49 ^{+0.14} _{-0.15}	1.78 ^{+0.92} _{-1.13}	328 ⁺³⁵ ₋₃₆	0.86 ^{+0.17} _{-0.12}
TPHM	42.57 ^{+4.08} _{-3.33}	12.63 ^{+1.06} _{-0.97}	0.58 ^{+0.12} _{-0.12}	0.31 ^{+0.48} _{-0.28}	0.077 ^{+0.068} _{-0.057}	0.57 ^{+0.12} _{-0.13}	1.38 ^{+0.16} _{-0.16}	1.67 ^{+1.01} _{-1.15}	322 ⁺³⁶ ₋₃₉	0.95 ^{+0.20} _{-0.16}
XPHM	37.84 ^{+3.00} _{-2.85}	14.17 ^{+1.09} _{-1.00}	0.39 ^{+0.21} _{-0.15}	0.64 ^{+0.29} _{-0.45}	0.034 ^{+0.065} _{-0.067}	0.38 ^{+0.22} _{-0.15}	1.51 ^{+0.45} _{-0.44}	1.44 ^{+0.89} _{-0.88}	311 ⁺⁶¹ ₋₆₂	0.88 ^{+0.23} _{-0.19}
0050 Heavy										
inj.	99.24	33.08	0.50	0.50	0.001	0.50	1.57	1.57	649	1.05
SEOB	98.85 ^{+10.75} _{-9.01}	39.01 ^{+11.26} _{-8.04}	0.64 ^{+0.24} _{-0.29}	0.49 ^{+0.43} _{-0.44}	0.085 ^{+0.121} _{-0.111}	0.62 ^{+0.24} _{-0.29}	1.46 ^{+0.25} _{-0.36}	1.36 ^{+1.12} _{-0.94}	694 ⁺¹⁸⁶ ₋₁₇₄	0.97 ^{+0.25} _{-0.17}
TPHM	100.63 ^{+15.24} _{-12.94}	38.68 ^{+15.13} _{-8.97}	0.57 ^{+0.21} _{-0.25}	0.43 ^{+0.48} _{-0.38}	0.122 ^{+0.105} _{-0.110}	0.54 ^{+0.20} _{-0.24}	1.32 ^{+0.31} _{-0.38}	1.43 ^{+1.12} _{-1.01}	705 ⁺¹⁹⁸ ₋₁₆₄	1.01 ^{+0.23} _{-0.19}
XPHM	90.36 ^{+7.01} _{-8.13}	46.31 ^{+13.09} _{-7.91}	0.30 ^{+0.29} _{-0.20}	0.80 ^{+0.17} _{-0.39}	0.139 ^{+0.106} _{-0.123}	0.36 ^{+0.22} _{-0.14}	1.38 ^{+0.62} _{-0.67}	1.16 ^{+0.70} _{-0.61}	679 ⁺²⁰² ₋₁₆₈	1.08 ^{+0.31} _{-0.24}
0628 Light										
inj.	40.44	19.61	0.85	0.51	-0.174	0.84	1.70	2.23	555	0.52
SEOB	38.65 ^{+3.54} _{-3.57}	20.36 ^{+2.25} _{-1.88}	0.87 ^{+0.10} _{-0.18}	0.33 ^{+0.44} _{-0.29}	-0.167 ^{+0.065} _{-0.066}	0.84 ^{+0.10} _{-0.17}	1.81 ^{+0.17} _{-0.19}	1.89 ^{+0.82} _{-1.12}	584 ⁺¹¹⁴ ₋₁₁₄	0.52 ^{+2.05} _{-0.22}
TPHM	40.44 ^{+4.66} _{-4.06}	19.62 ^{+2.27} _{-1.96}	0.78 ^{+0.14} _{-0.15}	0.33 ^{+0.48} _{-0.28}	-0.083 ^{+0.071} _{-0.065}	0.77 ^{+0.13} _{-0.15}	1.75 ^{+0.20} _{-0.16}	1.46 ^{+1.00} _{-0.87}	612 ⁺¹⁰⁸ ₋₁₁₀	0.55 ^{+2.11} _{-0.28}
XPHM	34.46 ^{+2.85} _{-2.57}	23.26 ^{+2.22} _{-2.51}	0.83 ^{+0.14} _{-0.23}	0.72 ^{+0.22} _{-0.50}	-0.102 ^{+0.091} _{-0.131}	0.63 ^{+0.21} _{-0.20}	2.24 ^{+0.29} _{-0.30}	0.79 ^{+0.56} _{-0.47}	585 ⁺¹⁴⁵ ₋₁₃₄	2.01 ^{+0.57} _{-1.53}
0628 Heavy										
inj.	84.07	40.77	0.85	0.51	-0.174	0.84	1.70	2.23	1020	0.52
SEOB	76.22 ^{+8.92} _{-8.92}	46.75 ^{+12.28} _{-8.17}	0.84 ^{+0.13} _{-0.24}	0.32 ^{+0.52} _{-0.28}	-0.179 ^{+0.094} _{-0.093}	0.79 ^{+0.15} _{-0.24}	1.89 ^{+0.27} _{-0.30}	1.73 ^{+0.91} _{-1.13}	1029 ⁺²⁵⁹ ₋₂₅₅	0.95 ^{+1.74} _{-0.56}
TPHM	78.37 ^{+13.73} _{-10.79}	46.87 ^{+11.36} _{-8.94}	0.78 ^{+0.17} _{-0.25}	0.29 ^{+0.50} _{-0.26}	-0.100 ^{+0.102} _{-0.096}	0.75 ^{+0.19} _{-0.26}	1.78 ^{+0.33} _{-0.31}	1.62 ^{+0.98} _{-1.03}	1130 ⁺³²³ ₋₂₈₅	0.85 ^{+1.77} _{-0.44}
XPHM	73.16 ^{+6.51} _{-5.59}	45.84 ^{+8.28} _{-9.45}	0.82 ^{+0.15} _{-0.30}	0.81 ^{+0.16} _{-0.46}	-0.139 ^{+0.163} _{-0.269}	0.53 ^{+0.20} _{-0.17}	2.42 ^{+0.31} _{-0.33}	0.78 ^{+0.71} _{-0.50}	909 ⁺³¹⁵ ₋₂₉₈	1.13 ^{+1.27} _{-0.53}

overestimation is more severe for our heavy-mass injection. For the 0628 run, SEOB recovers χ_{eff} very well, but TPHM and XPHM overestimate it once again.

Overall, the least informative recovery is for χ_2 as can be gathered from Figs. 18 and 19 and Table IV. Though the models capture the injected values within their 90% CIs, these are 2 to 3 times wider than the ones corresponding to χ_1 , thus hardly informative. The only clear result is the opposite behavior between the SEOB/TPHM and XPHM posteriors with the former favoring low values and the latter high values. We should also remark that the heavy 0050 SEOB/TPHM posteriors are mostly prior dominated. To better quantify this, we compute the Jensen-Shannon (JS) divergences, D_{JS} , between the posteriors and the priors. Given two distributions $p_1(x)$, $p_2(x)$, the JS divergence is a measure of how similar they are, given by [221]

$$D_{\text{JS}} = \frac{1}{2} \left[\sum_x p_1(x) \ln \left(\frac{p_1(x)}{\bar{p}(x)} \right) + p_2(x) \ln \left(\frac{p_2(x)}{\bar{p}(x)} \right) \right], \quad (26)$$

where $\bar{p}(x) = (p_1(x) + p_2(x))/2$. $D_{\text{JS}} = 0$ means identical distributions. We find that for the heavy-mass 0050 run, both SEOB and TPHM χ_2 posteriors yield $D_{\text{JS}} < 0.01$

indicating high resemblance to the flat χ_2 priors as can be seen from Fig. 18.

Similarly, for θ_2 , the SEOB/TPHM posteriors are mostly prior dominated. For three TPHM runs, the JS divergence between the θ_2 posteriors and the prior are all $\lesssim 0.01$, indicative of very similar distributions. Even for the light-mass 0628 run, TPHM yields $D_{\text{JS}} \approx 0.019$. SEOB posteriors give $D_{\text{JS}} \approx 0.03$ for the same run, and $D_{\text{JS}} < 0.015$ for the other three runs. On the other hand, the JS divergences between XPHM’s posteriors and the priors are all greater than 0.1 except for the light 0050 run which still yields $D_{\text{JS}} > 0.02$, which is usually taken to be the threshold for similarity between two distributions.

B. Inferring precession from the injections

In the previous section, we deliberately omitted the discussion on the recovery of χ_p as we want to present it along with the other precession scalars, namely χ_p^{Gen} and $\chi_{\perp, \text{J}}$. Recall that our injected signals have total SNRs of 40 each, part of which is due to the precession SNR, specifically, $\rho_p = \{\{11, 6.4\}, \{15, 10\}\}$, respectively for light/heavy 0050 and light/heavy 0628.

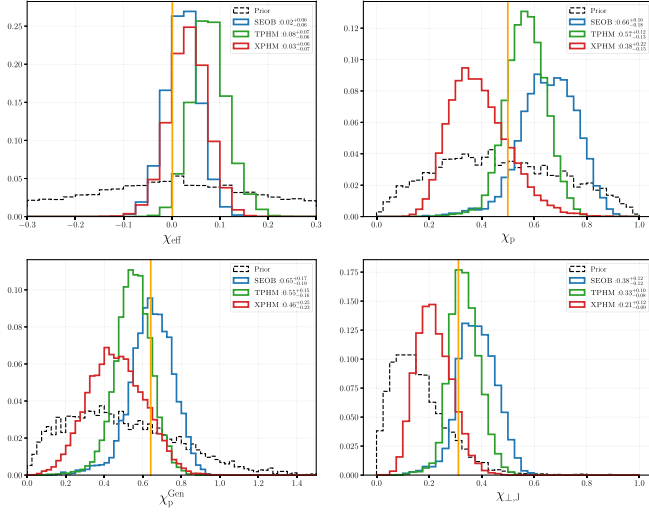


FIG. 20. Posterior distributions recovered by TPHM and XPHM for $\{\chi_{\text{eff}}, \chi_p, \chi_p^{\text{Gen}}, \chi_{\perp,J}\}$ for the injection of the SXS simulation 0050 with $M^{\text{det}} = 56.5M_{\odot}$. The black dashed distribution is the prior for each quantity. The orange line marks the injected value. The color coding of each model is consistent with the entire article.

We present the recovered posteriors for $\{\chi_p, \chi_p^{\text{Gen}}, \chi_{\perp,J}\}$ (along with χ_{eff}) from the light/heavy 0050 PE runs in Figs. 20 and 21. Similar results for the light/heavy 0628 recovery are shown in Figs. 22 and 23. In all the figures, we observe posteriors not dominated by the priors (dashed histograms) consistent with our having injected $\rho_p \geq 6$ into all runs. We also see the tallest and narrowest posteriors for the $\rho_p = 15$ run, i.e., the light 0628. As was the case in the previous section, the posteriors coming from the heavy-mass recovery runs are in general wider than their light-mass counterparts. They are additionally not as centered on the injected values and less symmetric.

Starting with the χ_p posteriors, we note that TPHM consistently recovers the injected value very close to its median with SEOB performing similarly with some disagreement between the two models' posteriors for the light-mass injections. XPHM consistently underestimates χ_p , but still recovers it within its 90% CI except for heavy 0628 (see Table IV). The underestimation by XPHM bias was not observed in Ref. [128], where the recovery of a $\{Q = 6, \chi_{\text{eff}} < 0, \chi_p \gtrsim 0.75\}$ injection (SXS:0165) was performed, albeit at an SNR of 26 with no information as to what ρ_p is. This might explain why a bias was not observed there though we should be cautious about drawing conclusions regarding systematic biases from a few PE runs [131].

Turning our attention to the χ_p^{Gen} posteriors, we find that all models recover the injected values within their 90% CIs as can be discerned from Figs. 20–23. The consistent underestimation of χ_p by XPHM is not exhibited here for χ_p^{Gen} . SEOB/TPHM recover the injected values closer to

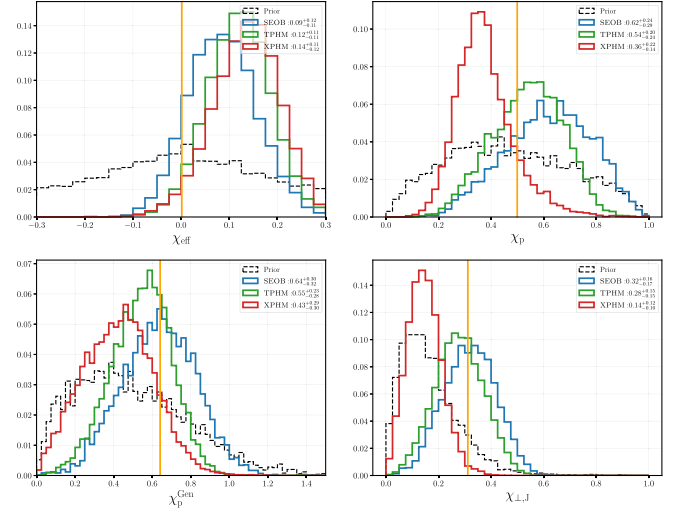


FIG. 21. Same as Fig. 20, but for the heavy 0050 PE run with an injected value of $M^{\text{det}} = 150M_{\odot}$. Note that the vertical scale is different from Fig. 20, which is why the histograms for the priors may give the illusion of looking different, when they are in fact identical.

the medians than XPHM except for the heavy 0628 run. In any case, we do not see the bias observed in the lower panels of Fig. 7 of Ref. [191], where injections of NRSUR7DQ4 were carried out. The reason for this most likely is that their injected SNRs were greater than 75.

As for the $\chi_{\perp,J}$ recovery, we find that both SEOB and TPHM also recover this quantity rather well, whereby the injected values almost line up with the peaks of nearly symmetric posterior distributions. TPHM's recovery is especially outstanding for the light-mass runs. XPHM underestimates the injected value every time and either barely recovers within its 90% CI or just outside it. As far as we are aware, ours are the first PE results for this quantity so we can not compare our findings with the literature.

Overall, both SEOB and TPHM seem to recover any one of the three χ_{\perp} 's well at these relatively high SNRs. The models especially recover $\chi_p^{\text{Gen}}, \chi_{\perp,J}$ posteriors well though this is likely just a coincidence, and nonetheless warrants a more systematic investigation. XPHM seems to best recover χ_p^{Gen} of the three χ_{\perp} 's. Focusing only on the χ_p recovery, we see that SEOB, TPHM perform reliably, with perhaps the latter being most reliable in terms of consistent recovery closest to its median.

Looking at the $\{t_c, \varphi_c\}$ -maximized mismatches of Table III, we can see some support for why we expect SEOB, TPHM to perform better: For the heavy injections, SEOB/TPHM waveforms with injected parameters yield mismatches that are nearly half of the corresponding XPHM waveforms. As for the light cases, SEOB mismatches are lower, while the TPHM mismatch is lower for 0050, and roughly the same for 0628. We can compare these mismatches with what the indistinguishability criterion of

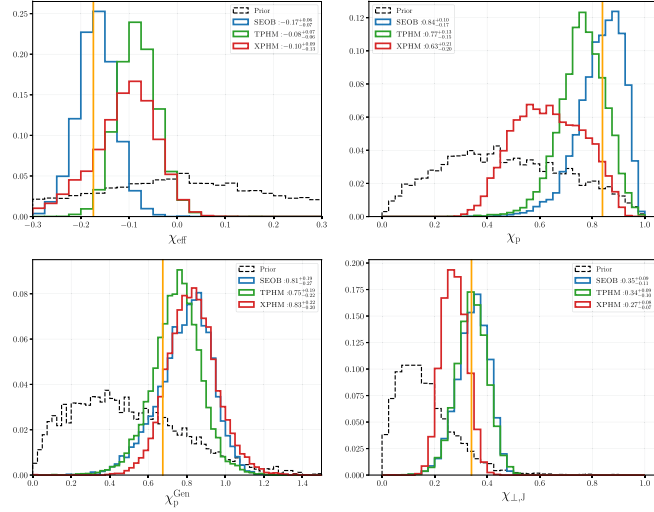


FIG. 22. Same as Fig. 20, but for the light 0628 run with an injected value of $M^{\text{det}} = 67M_{\odot}$.

Ref. [51] gives us. Namely, for k degrees of freedom and an SNR of ρ , the required mismatch below which two waveforms will be indistinguishable at $100p\%$ confidence is given by [64]

$$\bar{\mathcal{M}} \leq \frac{\chi_k^2(1-p)}{2\rho^2}. \quad (27)$$

A precessing BBH system has 8 degrees of freedom. This reduces to 4 if we minimize the mismatches over the planar spin components ϕ_1, ϕ_2 . To quantify waveform indistinguishability at 90% confidence, we use $\chi_8^2(0.1) \simeq 13.362$ and $\chi_4^2(0.1) \simeq 7.79$. Accordingly, for an SNR of 40, two waveforms will be indistinguishable if their mismatch is less than 4.2×10^{-3} assuming $k = 8$ and $\bar{\mathcal{M}} \leq 2.4 \times 10^{-3}$ assuming $k = 4$. The mismatches we have in Table III are larger than these bounds, but SEOB's are only twice to thrice the $k = 8$ bound, with TPHM's slightly larger. Thus, we expect the results of our PE recovery to be somewhat biased, but less so for SEOB and TPHM. Perhaps this is manifest mostly in the precession-related posteriors. We can also solve Eq. (27) for an “indistinguishability SNR” given a mismatch. For the fiducial value of $\bar{\mathcal{M}} = 0.01$, we obtain $\rho \approx 26$. Thus, any injection with less than this SNR would not pick up modeling-related systematics in PE.

VII. SUMMARY

We conduct a long survey of four precessing waveform models: SEOBNRV5PHM, TEOBRESUMS, IMRPHENOMTPHM, and IMRPHENOMXPHM. Our survey has a large part pertaining to assessing model faithfulness to the surrogate model NRSUR7DQ4 and to numerical relativity waveforms from the SXS simulation catalog. We quantify model faithfulness

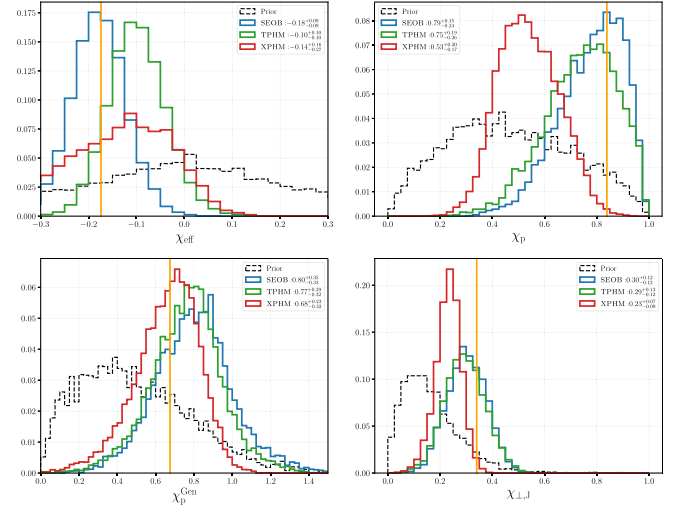


FIG. 23. Same as Fig. 20, but for the heavy 0628 run with an injected value of $M^{\text{det}} = 150M_{\odot}$.

via the optimized mismatch, $\bar{\mathcal{M}}_{\text{opt}}$ [Eqs. (20) and (21)], and the merger-ringdown truncated (inspiral-only) mismatch, $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ [Eq. (23)]. The second shorter part of our survey involves checking model performance in several injection/recovery parameter estimation runs for which we cannot employ TEOBRESUMS.

The assessment of model faithfulness to NRSUR7DQ4 has two separate parts distinguished by the chosen set of intrinsic parameters. In Secs. IV A and IV B, we employ a discrete grid with emphasis on having large planar components for the spin vectors $\mathbf{S}_1, \mathbf{S}_2$ to “stress test” the precessing models. In Sec. IV C, we use a random-uniformly filled intrinsic parameter space. The latter is commonly encountered in reviews of waveform models as well as in parameter estimation runs, where the parameter space is randomly sampled. As can be seen from Fig. 2, the two parameter sets cover different regions. For the discrete grid, we consider both a light- and a heavy-mass binary with total source-frame masses of $37.5M_{\odot}$ and $150M_{\odot}$. The cases with the latter mass contain fewer inspiral cycles; thus the plunge-merger-ringdown portions contribute more to the signal than for the light-mass cases. We further decompose each mass set into four subsets of mass ratio, $Q = \{1, 2, 4, 6\}$, the last of which is in the so-called extrapolation region of NRSUR7DQ4.

In Sec. V, we extend the above survey to comparisons with numerical relativity waveforms from the SXS catalogs. We separate our investigation into a part involving “short” waveforms, i.e., fewer than 70 GW cycles (Sec. V A) and into a part with longer waveforms ($\gtrsim 125$ GW cycles) in Sec. V B. For the former, we once again set the binary mass to $37.5M_{\odot}$ and $150M_{\odot}$. For the latter, we consider a total of six values for the total mass ranging from $37.5M_{\odot}$ to $225M_{\odot}$ in steps of $37.5M_{\odot}$. For the short SXS set, we pick

317 simulations out of the larger ensemble of approximately 1600 in order to have roughly 100 precessing waveforms per mass ratio subsets: $\{Q \lesssim 1.3, Q \approx 2, Q \approx 4\}$. This way, we are better able to draw parallels with the assessment conducted using NRSUR7DQ4 in Secs. IV A and IV B.

We compute all mismatches at inclinations of $\vartheta_{\text{LN},0} = 0$ and $\vartheta_{\text{LN},0} = \pi/2$. For each case, we also compute the merger-ringdown truncated, i.e., inspiral-only, mismatch denoted by the noMR superscript. Our findings are as follows:

- (i) For each case in our parameter sets, $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}} < \bar{\mathcal{M}}_{\text{opt}}$. The corresponding distributions can be seen in Figs. 4, 10, 13, and 17. The ratio $\mathcal{R} = \bar{\mathcal{M}}_{\text{opt}} / \bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ can be $\mathcal{O}(10)$ or larger, but remains $\lesssim 7$ for SEOB.
- (ii) The offset between the $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ distributions and the $\bar{\mathcal{M}}_{\text{opt}}$ distributions is larger for the heavier-mass binaries as can be seen by comparing the left and right panels of Figs. 4 and 13.
- (iii) Model faithfulness deteriorates with increasing mass asymmetry as shown in Figs. 5, 9, and 15.
- (iv) For short waveforms, model faithfulness, specifically $\bar{\mathcal{M}}_{\text{opt}}$, does not significantly degrade at higher inclinations as exhibited in Figs. 8, 10, 13, and 15. However, for long waveforms, i.e., $\gtrsim 125$ cycles, faithfulness decreases as inclination goes from 0 to $\pi/2$ as shown in Fig. 17. The models' faithfulness under changing inclination also varies when considering the inspiral only ($\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$) and different mass ratios.
- (v) Models exhibit higher unfaithfulness to NRSUR7DQ4/SXS in regions with strong precession.
- (vi) When large enough, the unfaithfulness of the coprocessing (aligned-spin) $(2, \pm 2)$ multipoles to the corresponding NRSUR7DQ4/SXS multipoles can become the dominant systematic in the precessing waveform unfaithfulness. This is especially the case for TEOBRESUMS and IMRPHENOMTPHM as detailed in Secs. IV A 3 and VA 4, and is manifested as a strong correlation shown in Figs. 7 and 16.
- (vii) Sufficiently faithful precessing $(2, \pm 1)$ multipoles can make high-inclination waveforms more faithful than their low-inclination counterparts as explained in Secs. IV B and VA 3.

Finally, in Sec. VI, we conduct a parameter estimation study by performing zero-noise injections of two precessing SXS simulations (0050 and 0628) and recovering the injected parameters with the models SEOBNRV5PHM, IMRPHENOMTPHM, and IMRPHENOMXPHM. Our work in this regard complements those of Refs. [98,128,131], where different SXS simulations were injected. We perform each injection twice, at a low and high total binary mass. We set the extrinsic parameters so that in each case the total network SNR equals 40 with 15% to 40% coming from

precession depending on the case. The recovered posterior distributions for a subset of key parameters are shown in Figs. 18 and 19 for SXS:0050 and SXS:0628, respectively. These are supplemented by Table IV where we present the median values as well as the $\{5, 95\}$ percentile error bars of the posteriors for a complementary subset of parameters with some overlap.

Overall, we observe mixed model performance where one model may recover certain parameters better than the other. For example, SEOBNRV5PHM consistently recovers χ_{eff} better while IMRPHENOMTPHM could be argued to better recover χ_1 (no railing), whereas IMRPHENOMXPHM better recovers θ_2 . However, when it comes to the recovery of the effective precession scalars, $\chi_p, \chi_p^{\text{Gen}}$ and the perpendicular spin component $\chi_{\perp,J}$, SEOBNRV5PHM and IMRPHENOMTPHM outperform IMRPHENOMXPHM, which we present in Figs. 20 and 21 for SXS:0050, and in Figs. 22 and 23 for SXS:0628. χ_p^{Gen} and $\chi_{\perp,J}$ are especially rather well recovered by SEOBNRV5PHM and IMRPHENOMTPHM as can be garnered from the bottom panels of these figures. We must nonetheless repeat our cautionary remark from Sec. VI that results from a few parameter estimation runs that are insufficient for drawing general, reliable conclusions regarding model performance. Therefore, many more such studies are needed with special attention paid to the precession SNR.

VIII. DISCUSSION

Our findings indicate that at present, the MR portion of the waveforms along with the precessing $(2, \pm 1)$ multipoles need to be further improved to attain an overall waveform faithfulness comparable to that of the inspiral-only $(2, \pm 2)$ multipoles for all inclinations. As already mentioned, part of the increased unfaithfulness when considering the full waveform is simply due to changing signal morphology of the MR phase. One can, in principle, determine how much of the increase is due to the actual mismodeling of the MR regime by smoothly stitching the inspiral part of, e.g., an SEOBNRV5PHM waveform with the MR part of an NRSUR7DQ4/SXS waveform and computing the mismatch $\bar{\mathcal{M}}_{\text{opt}}$ between this hybrid waveform and the full SEOBNRV5PHM waveform. A comparison of this mismatch with the already computed $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ values with respect to NRSUR7DQ4/SXS then roughly tells us how much of the mismatch comes from waveform systematics in the MR regime. This comparison might provide an estimation for an achievable waveform model unfaithfulness that can be targeted.

As for the modeling of the precessing $(2, \pm 1)$ multipoles, we have seen already that the main contribution to these comes from the AS $(2, \pm 2)$ multipoles which differ most from NRSUR7DQ4/SXS's coprocessing $(2, \pm 2)$ multipoles for cases with $60^\circ \leq \theta_{1,2} \leq 120^\circ$ and with increasing mass asymmetry. This region in $\theta_{1,2}$ space is also where we

believe the assumed $m \leftrightarrow -m$ symmetry of the AS multipoles is most unfaithful. The obvious remedy is the abandonment of this symmetry, at least for the AS $(2, \pm 2)$ multipoles which has been done for the new model IMRPHENOMXO4A [135].

Though the AS $(2, \pm 1)$ multipoles do not matter nearly as much as their $(2, \pm 2)$ counterparts, we have seen that their omission can cause mismatches as large as 0.01 for $\vartheta_{\text{LN},0} = 0$ and 0.03 for $\vartheta_{\text{LN},0} = \pi/2$. As detector sensitivity steadily improves such margins will become crucial to overcome. More work needs to be undertaken to fully determine how much the unfaithfulness of the AS $(2, \pm 1)$ multipoles matters, while it is clear that this is subdominant to the unfaithfulness of the AS $(2, \pm 2)$ multipoles. However, the ratio $|h_{2,\pm 1}^{\text{AS}}|/|h_{2,\pm 2}^{\text{AS}}|$ increases with increasing mass asymmetry implying that unfaithfulness of the AS $(2, \pm 1)$ multipoles becomes non-negligible beyond a certain mass ratio.

The repeatedly observed fact that faithfulness, in general, deteriorates with increasing mass asymmetry ($Q \gtrsim 4$) should be addressed more systematically. Though it is the case that these waveforms have more GW cycles than the ones with $Q \leq 2$, we had argued that waveform length alone does not explain the increased unfaithfulness. Part of the degradation is due to the AS multipoles as we saw in Sec. IVA 3: They become less faithful to NRSUR7DQ4's coprocessing multipoles as Q increases. On the other hand, some deterioration could be due to the mismodeling of the AS $(2, \pm 1)$ multipoles which have increasing contributions as discussed above. Another part could be due to the mismodeling of the Euler angles which brings us to the second part of this discussion.

Our survey is by no means a complete one, so let us expand on possible ways to extend it here. The twist prescription has two essential ingredients: the AS multipoles and the Euler angles used to rotate the multipoles from the $\mathbf{L}_N(t)$ frame (coprocessing) to an inertial (e.g., $\mathbf{L}_{N,0}$) frame. Though we investigated how the unfaithfulness of the AS multipoles might affect the strain faithfulness, we did not conduct a study of how the systematics in the Euler angles affect the faithfulness. We propose the following hybridized approach for such a study: Euler rotation of NRSUR7DQ4's coprocessing multipoles for which the angles are to be input from various models under investigation. Such a study was performed in Ref. [132] (Fig. 18) for assessing the MSA angles and the angles from their new prescription.

In our case, we could, for example, use TEOBRESUMS Euler angles to twist the coprocessing NRSUR7DQ4 multipoles then compute the mismatches between the resulting hybrid waveforms and NRSUR7DQ4 waveforms. This investigation can also be extended to the choice of coprocessing frame; i.e., since $\mathbf{L} \neq \mathbf{L}_N$, the Euler angles for the $\mathbf{L}(t) \rightarrow \mathbf{L}_0$ rotation will differ from those of the $\mathbf{L}_N(t) \rightarrow \mathbf{L}_{N,0}$ rotation, albeit slightly. Moreover, neither

direction coincides with the direction of maximum GW emission which, in the time domain, differs depending on whether one uses the GW strain, the Bondi news function, or the Weyl scalar ψ_4 [49,132,206,222,223].

It would be useful to extend our total mass coverage in Secs. IV and VA to more values as we did for our assessment using the long sxs waveforms. A range of $M \approx 10M_\odot$ to $\approx 250M_\odot$ in steps of $10M_\odot$ is common. Additionally, for random-uniformly-filled parameter spaces, 5000 to $\mathcal{O}(10^4)$ cases are customary. However, given that we have to compute $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and $\bar{\mathcal{M}}_{\text{opt}}$ separately at two different inclinations for each single point in the BBH parameter space, we have 4 times the computational burden. This is a considerable increase as each optimized mismatch requires up to $\mathcal{O}(10^5)$ waveform generations which we must then repeat for 42 points in our $\{\varphi_{\text{ref}}, \kappa\}$ grid [Eq. (20)].

Another obvious way to extend our survey is by including higher multipoles in the strain. Reference [182] already conducted such a study with the multipoles $(3, \pm 3)$, (3 ± 2) , $(4, \pm 4)$, $(4, \pm 3)$ included. We can extend our AS (2 ± 2) , (2 ± 1) multipole assessment of Sec. IVA 3 to these multipoles. We can also reapply all the analyses that we performed on the $\ell = 2$ strain to the $\ell > 2$ strain.

Yet another extension is to employ as many precessing numerical relativity waveforms as feasible as was done in the LVK study using $\gtrsim 1500$ NR waveforms that we mentioned in Sec. I. A shorter version of this at a single inclination of $\pi/3$ was presented in Ref. [98]. We can supplement these results with our $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and $\bar{\mathcal{M}}_{\text{opt}}$ computed at inclinations of 0 and $\pi/2$ and possibly more. All of these data could then be presented in terms of mass ratio as in our Figs. 5, 9, and 15. We should also add that the latest trend in the literature is to present model unfaithfulness in terms of an SNR-weighted match [95,132]. This quantity is different from the sky-optimized match \mathcal{M}_{opt} of Eq. (19). Therefore, a comparison of these two faithfulness measures may also be informative.

A large-scale injection/recovery parameter estimation campaign would make certain vague trends either more concrete or obsolete. However, such studies are computationally expensive which is why most model reviews suffice with one or two of these. Nevertheless, we have found that using the PARALLEL_BILBY library, we can complete two chains of IMRPHENOMTPHM and two chains of IMRPHENOMXPHM injection/recovery runs per day using $\mathcal{O}(100)$ processors. With the same resources, we can alternatively obtain one SEOBNRV5PHM chain per ~ 1.5 days. So, $\gtrsim 30$ injection/recovery PE runs can be completed in approximately one month with the IMRPHENOM models.

The conclusions of injection/recovery studies are dependent on the choice of injected SNRs. In our case, we opted for a total SNR of 40 which may seem unrealistic given that

95% of the O1, O2, O3 events had $\text{SNR} \lesssim 23$.⁶ However, there will be a few events with large enough SNRs for which accurate modeling of precession will be crucial. GW200129_065458, at an SNR of 26, was such an event where a reanalysis of the data with the state-of-the-art precessing model NRSUR7DQ4 discovered strong precession [64], consistent with an earlier analysis from an IMRPHENOMXPHM run, but not with the one from SEOBNRV4PHM [22]. The precession SNR ρ_p should also be chosen carefully in such studies. Even if an SNR = 40 system is very strongly precessing, if the extrinsic parameters conspire to minimize ρ_p , the imprint of precession on the received GWs will be weak.

Finally, let us add that a survey such as ours risks becoming outdated by the time it may be complete. Indeed, during the writing of this article, the model SEOBNR got upgraded from v4 to v5. Additionally, the MSA prescription of BBH precession dynamics in IMRPHENOMXPHM got replaced by one based on SPINTAYLORT4 [136]. Finally, an upgraded IMRPHENOMX (O4A) model was released [135]. The inclusion of any new models into a similar survey inevitably delays the completion of the survey itself. However, once a new model is incorporated into LALSIMULATION it should, in principle, be straightforward to assess its faithfulness within the framework that we have built.

The supporting data for this paper are openly available on GitHub [143]. The following software have been used for this work. The NumPy [225], Matplotlib [226], PANDAS [227,228], and SEABORN [201] libraries of PYTHON. LALSuite (7.10) [203], LALInference [229], PyCBC (2.0.6) [230] BILBY (2.1.2) [212,213], the SXS package (v2022.5.2) [204], the SCRI package (2022.8.8) [208], and TEOBResumSv4.1.5-GIOTTO 84b8f10 (July 6, 2023). These results were found to be in good agreement with the more recent TEOBRESUMS bd3452e (September 7, 2023).

ACKNOWLEDGMENTS

We thank Rossella Gamba for sharing codes and help with TEOBRESUMS, Matteo Breschi for help with BAJES, and Marta Colleoni for answering questions regarding IMRPHENOMXP. We are grateful to Charlie Hoy for helping us with the troubleshooting of the SEOBNRV5PHM PE runs and sharing the results of his validation runs using the Sciamia High Performance Compute cluster which is supported by the ICG, SEPNet, and the University of Portsmouth. We are also grateful to Lorenzo Pompili for pointing out the final fix to our SEOBNRV5PHM PE runs and sharing, in particular, the results of the heavy-mass PE injection/recovery with SEOBNRV5PHM. This work makes

⁶We extracted this value from the list of events provided in Ref. [224].

use of the Black Hole Perturbation Toolkit [231] and the “pesummary” package [232]. S. A. and J. M. U. acknowledge support from the University College Dublin Ad Astra Fellowship. J. T. acknowledges support from the NASA LISA Preparatory Science Grant No. 20-LPS20-0005. We thank Niels Warburton for providing access to the “chirp2” computer, which is funded by Royal Society—Science Foundation Ireland University Research Fellowship Grant No. RGF/R1\180022. This research has made use of data, software, and/or web tools obtained from the Gravitational Wave Open Science Center [233,234], a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. We have additionally employed the computational resources of LIGO Laboratory (CIT cluster) supported by the U.S. National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459. LIGO Laboratory and Advanced LIGO are funded by the United States NSF as well as the Science and Technology Facilities Council of the United Kingdom, the Max-Planck-Society, 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, by the French Centre National de Recherche Scientifique, the Italian Istituto Nazionale di Fisica Nucleare, and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, and Spain. K. A. G. R. A. is supported by Ministry of Education, Culture, Sports, Science and Technology, Japan Society for the Promotion of Science in Japan; National Research Foundation and Ministry of Science and ICT in Korea; Academia Sinica and National Science and Technology Council in Taiwan.

APPENDIX A: MULTIDIMENSIONAL FITS OF MISMATCHES FOR BAYESIAN MODEL SELECTION

As briefly mentioned in Sec. I, we can use the results of our faithfulness survey to inform Bayesian model selection. For example, one can employ the waveform mismatch, $\bar{\mathcal{M}}_{\text{opt}}$ in our case, as a type of weight to generate a weighted categorical prior. This then can act as a custom prior in a joint Bayesian analysis [142]. Therefore, by constructing fits to $\bar{\mathcal{M}}_{\text{opt}}$ over the intrinsic parameter space, we can effectively provide a parameter-space-dependent prior for Bayesian model selection. We take a first step toward building this categorical prior here by constructing multi-dimensional fits to $\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)$.

Though we have at our disposal several sets of intrinsic parameters, namely the discrete grid of Sec. IV A, the uniformly filled space of Sec. IV C, and the SXS sets of

Sec. V, we choose the discrete grid, as this set seems to cover the high-spin regions of the parameter space better (see Fig. 2). We should keep in mind that we fixed $\chi_1 = \chi_2 = 0.8$ for this set.

In more detail, we construct multidimensional polynomial fits to $\log_{10}[\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)]$ as functions of η, χ_{eff} and either $\chi_{\text{p}}^{\text{Gen}}$ or $\chi_{\perp,\text{J}}$. We choose these perpendicular projections as opposed to χ_{p} since these have better coverage of the parameter space as shown in Fig. 2. Using the symmetric mass ratio η allows us to work with a compact fit domain since $0 < \eta \leq 0.25$. We do not include the $Q = 6(\eta = 6/49)$ subset for our fits here though this can be included straightforwardly. For our specific fits, we use the values for $\{\chi_{\text{eff}}, \chi_{\text{p}}^{\text{Gen}}, \chi_{\perp,\text{J}}\}$ at $f = f_0$, but have dropped the subscript 0 to reduce clutter. The total mass, M , is the fourth parameter in our fits, but since we have only two points in mass space, i.e., $M = 37.5M_{\odot}, 150M_{\odot}$, we construct separate 3D fits for each mass. Future work in this direction will have additional data coming from other values of the total mass, thus necessitating 4D fits.

For fitting, we employ simple polynomials of the form

$$\mathcal{P}_{\mathcal{M}}(x, y, z) = \sum_{j=0}^{n_y} \sum_{k=0}^{n_z} \sum_{i=0}^{n_{\text{max}}^{kj}} a_{ijk} x^i y^j z^k, \quad (\text{A1})$$

where a_{ijk} are the fitting coefficients, $\{y, z\} = \{\chi_{\text{eff}}, \eta\}$, $x = \chi_{\text{p}}^{\text{Gen}}$ or $\chi_{\perp,\text{J}}$, and $n_{\text{max}}^{kj} = \max\{n_x, n_y, n_z\} - k - j$. We

expect the highest variation in mismatches to be along the planar spin direction and the least variation in the mass ratio space. Therefore, we impose $n_x > n_y > n_z$ with $n_x \leq 7$, $n_y \leq 5$, $n_z \leq 4$. For our fits, we use a routine that picks the best values for n_x, n_y, n_z that minimize (i) $1 - \bar{R}^2$, where \bar{R}^2 is the adjusted R-squared value, (ii) the Bayesian information criterion (BIC), (iii) the Akaike information criterion with the finite sample correction (AICc), and (iv) the average relative difference between the fit $\mathcal{P}_{\mathcal{M}}$ and the validation data $d_i = (x_i, y_i, z_i, w_i)$ which we define as

$$\Delta_{\text{rel}}^{\text{av}} := \frac{1}{N} \left(\sum_{i=1}^N \left[1 - \frac{w_i}{\mathcal{P}_{\mathcal{M}}(x_i, y_i, z_i)} \right]^2 \right)^{1/2}, \quad (\text{A2})$$

where w_i represent $\log_{10}[\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)]$. We employ two thirds of the dataset for fitting and the remaining one third for validation, where the assignment fit/validation is made randomly. There are likely more informative ways of choosing the fit/validation subsets, but we leave this for future work. There is also not necessarily one set of unique values of $\{n_x, n_y, n_z\}$ that satisfy all four criteria listed above. If more than one set of values satisfy the criteria, we pick the one with smaller values of $\{n_x, n_y, n_z\}$. On the other hand, if the routine cannot meet all four criteria, it looks for $\{n_x, n_y, n_z\}$ that minimize three out of the four.

Using this routine and the fitting function (A1), we construct 3D fits of $\log_{10}[\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)]$ for all four approximants for $M = 37.5M_{\odot}, 150M_{\odot}$. We find that in

TABLE V. List of the three-dimensional fits to the mismatches between a given waveform model and NRSUR7DQ4. Specifically, we have fitted to $\log_{10}[\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)]$ values resulting from the discrete parameter set of Sec. IV A. From left to right, the columns are the model name, the total mass, the x variable used in the fits [see Eq. (A1)], the three-dimensional polynomial order of the fits (*ibid.*), the total number of parameters used, the adjusted R-squared value, the average relative difference between the fit and the validation data [see Eq. (A2)], the Bayesian information criterion value, and the value of the finite-sample size corrected Akaike information criterion.

Model	$M(M_{\odot})$	x	(n_x, n_y, n_z)	No. params.	\bar{R}^2	$\Delta_{\text{rel}}^{\text{av}}$	BIC	AICc
SEOB	37.5	$\chi_{\text{p}}^{\text{Gen}}$	(4,2,1)	21	0.777	3.33×10^{-3}	-696	-793
TEOB	37.5	$\chi_{\text{p}}^{\text{Gen}}$	(7,5,2)	81	0.959	2.49×10^{-3}	-1512	-1858
TPHM	37.5	$\chi_{\text{p}}^{\text{Gen}}$	(6,4,2)	60	0.874	1.54×10^{-2}	-304	-565
XPHM	150	$\chi_{\text{p}}^{\text{Gen}}$	(6,5,2)	63	0.895	4.97×10^{-3}	-481	-754
SEOB	150	$\chi_{\text{p}}^{\text{Gen}}$	(6,4,2)	60	0.813	5.10×10^{-3}	+76	-182
TEOB	150	$\chi_{\text{p}}^{\text{Gen}}$	(6,5,2)	63	0.893	5.96×10^{-3}	-479	-751
TPHM	150	$\chi_{\text{p}}^{\text{Gen}}$	(6,4,2)	60	0.869	6.89×10^{-3}	-319	-579
XPHM	150	$\chi_{\text{p}}^{\text{Gen}}$	(7,4,2)	75	0.914	6.79×10^{-3}	-248	-570
SEOB	37.5	$\chi_{\perp,\text{J}}$	(7,5,2)	81	0.884	3.17×10^{-3}	-823	-1171
TEOB	37.5	$\chi_{\perp,\text{J}}$	(7,3,2)	66	0.965	1.72×10^{-3}	-1646	-1929
TPHM	37.5	$\chi_{\perp,\text{J}}$	(6,5,3)	73	0.898	5.63×10^{-3}	-353	-666
XPHM	37.5	$\chi_{\perp,\text{J}}$	(7,4,1)	55	0.904	3.96×10^{-3}	-606	-846
SEOB	150	$\chi_{\perp,\text{J}}$	(6,5,2)	63	0.884	5.10×10^{-3}	-203	-475
TEOB	150	$\chi_{\perp,\text{J}}$	(6,5,2)	63	0.918	4.85×10^{-3}	-680	-954
TPHM	150	$\chi_{\perp,\text{J}}$	(6,4,2)	60	0.908	5.79×10^{-3}	-572	-834
XPHM	150	$\chi_{\perp,\text{J}}$	(7,3,2)	66	0.933	5.35×10^{-3}	-457	-742

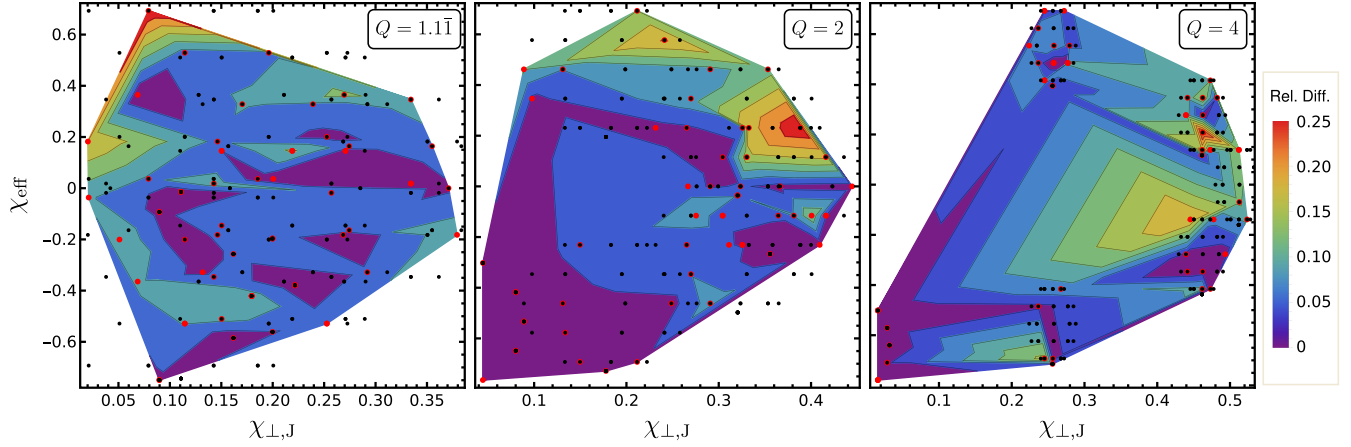


FIG. 24. Contour plots of the relative difference between the validation data (red points) and the third TEOB fit of Table V, which is obtained from Eq. (A1) with $\{n_x, n_y, n_z\} = \{7, 3, 2\}$. This particular fit is a function of $\chi_{\perp,J}, \chi_{\text{eff}}$ and η and represents the $M = 37.5M_{\odot}$, $\log_{10}[\bar{\mathcal{M}}_{\text{opt}}(\vartheta_{\text{LN},0} = 0)]$ data, where $\bar{\mathcal{M}}_{\text{opt}}$ is the sky-optimized mismatch between NRSUR7DQ4 and TEOB. The three panels from left to right show contour plots of the relative differences separated by mass ratios of $Q = 1.1, 2, 4$. The black dots are the data points used to construct the fit and are not covered by the contour plot shown here which only applies to the red dots.

most cases, the best fit functions have $n_x = 6$ or 7 , $n_y = 4$ or 5 and, not surprisingly, $n_z \leq 2$ (since there are only three η values), thus resulting in roughly $\mathcal{O}(60)$ parameters as shown in Table V where we also report various goodness-of-fit metrics. As can be seen from the table, the mismatches for TEOB and XPHM can be fit rather well for both light and heavy masses with \bar{R}^2 values as high as 0.965 and an average relative disagreement as low as 1.7×10^{-3} . SEOB mismatches can still be fit, but not as well, with \bar{R}^2 below 0.9. TPHM mismatches prove to be hardest to fit with the corresponding $\Delta_{\text{rel}}^{\text{av}}$ values being some of the highest.

Table V also hints that fits constructed using $\chi_{\perp,J}$ yield slightly higher (lower) values for $\bar{R}^2(\Delta_{\text{rel}}^{\text{av}})$ than those using χ_p^{Gen} though the difference is rather marginal and requires a more thorough investigation to be conclusive. We show one of the fits of Table V in Fig. 24 in terms of 2D contour plots of the relative difference between the fit and the validation data for the $M = 37.5M_{\odot}$ TEOB mismatches. This particular fit uses $\chi_{\perp,J}$ so corresponds to the third TEOB row of Table V with $\bar{R}^2 = 0.965$. As the figure shows, for most of the plot regions, the relative difference between the validation data points (red) and the fit is less than 0.05. In fact, according to Table V, the average relative difference is $\approx 1.7 \times 10^{-3}$. Fits of this level of faithfulness may prove quite useful for Bayesian model selection.

We additionally briefly explored whether or not a more finely sampled set of parameters improves the goodness of the fits. In particular, we recomputed the $M = 150M_{\odot}$ NRSUR7DQ4-XPHM mismatches for $\mathcal{O}(3000)$ BBHs, i.e., nearly 3 times as many data points in the spin space. For both the χ_p^{Gen} and $\chi_{\perp,J}$ fits, we observed slight improvement with \bar{R}^2 almost equaling 0.95 and $\Delta_{\text{rel}}^{\text{av}}$ dropping to 2×10^{-3} which can be compared with the second and fourth XPHM

rows of Table V. We should add that this improvement comes at the expense of increasing the computational burden by a factor of 3.

In summary, we have illustrated that, using certain projections of the spins, it is possible to construct reasonably good three-dimensional fits of the mismatches, effectively capturing most of the behavior of model faithfulness over the seven-dimensional intrinsic parameter space. The fits can be improved by including a denser coverage of the parameter space, but also need to be extended along the fourth dimension (total mass) to be of actual use in Bayesian model selection. There are further additional improvements that can be made, but we leave it all to future work. Our goal here was to show that faithful fits can be constructed in principle by using either χ_p^{Gen} or $\chi_{\perp,J}$ and the goodness-of-fit metrics in Table V seem to affirm this.

APPENDIX B: A BRIEF COMPARISON OF SEOBNRV5PHM WITH SEOBNRV4PHM

As already mentioned, while in the middle of this project, the upgraded model SEOBNRV5PHM [98] was released along with a simple-to-use PYTHON package PYSEONBR [101] for it. As the newer model is considerably faster than its v4 predecessor, we were easily compelled to use it. We had however already generated a considerable dataset of SEOBNRV4PHM mismatches by then. Therefore, we show a brief comparison between the newest (v5) and the previous versions in Fig. 25 where we plot $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and $\bar{\mathcal{M}}_{\text{opt}}$ once again at inclinations of $\vartheta_{\text{LN},0} = 0$ and $\pi/2$ for the randomly uniformly filled 1000-case parameter set. We see that the upgraded model is more faithful to NRSUR7DQ4 than its predecessor.

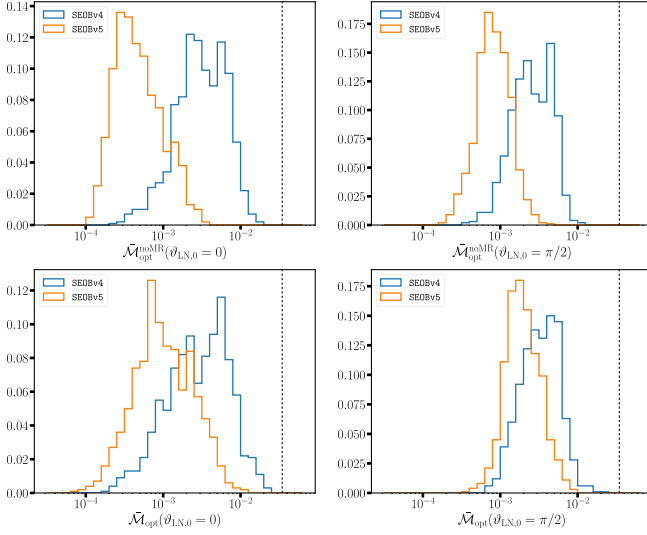


FIG. 25. Comparison of the faithfulness of SEOBv5PHM with that of SEOBv4PHM for the 1000-case random-uniformly-filled parameter space described in Sec. IV C. The baseline model used to measure faithfulness is NRSUR7DQ4.

APPENDIX C: ANOTHER COMPARISON: IMRPHEMOMXO4A VS MSA AND SPINTAYLOR IMPLEMENTATIONS OF IMRPHEMOMXPHM

Here, we provide a brief comparison between the original IMRPHEMOMXPHM (using MSA angles) [128] with the upgraded version that uses SPINTAYLORT4 dynamics to obtain the Euler angles [136]. We also include the new model IMRPHEMOMXO4A [135] in the comparison. We had briefly discussed the unfaithfulness performances of the MSA and the SPINTAYLORT4 versions at the end of Sec. V B. We supplement it here with Fig. 26 where we plot $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$ and $\bar{\mathcal{M}}_{\text{opt}}$ yet again at inclinations of $\vartheta_{\text{LN},0} = 0$ and $\pi/2$ for the random-uniform parameter set. We observe very similar performances between the MSA and SPINTAYLOR implementations, and XO4A. The performance of the MSA version is only slightly inferior at higher inclinations with the newer IMRPHEMOMX models slightly outperforming their predecessor. All three models yield very similar distributions for the inspiral-only mismatches $\bar{\mathcal{M}}_{\text{opt}}^{\text{noMR}}$, indicating that the differences exhibited may be more related to merger-ringdown modeling.

A more striking difference in the faithfulness of the MSA version vs the other two is exposed when considering the discrete parameter set of Secs. IV A and IV B where we had a small number of cases with parameters chosen such that $|\mathbf{J}_{\text{N},0}| \approx 0.025M^2$ which caused the MSA prescription to break down. This was exhibited as small secondary modes located at $\bar{\mathcal{M}}_{\text{opt}} \approx 0.1(0.3)$ for $M = 37.5M_{\odot}(150M_{\odot})$ in the mismatch distributions plotted in Figs. 3, 8, and replotted in Fig. 27, where we plot the histograms for $\bar{\mathcal{M}}_{\text{opt}}$ at $\vartheta_{\text{LN},0} = 0, \pi/2$ for $M = 37.5M_{\odot}, 150M_{\odot}$. We can

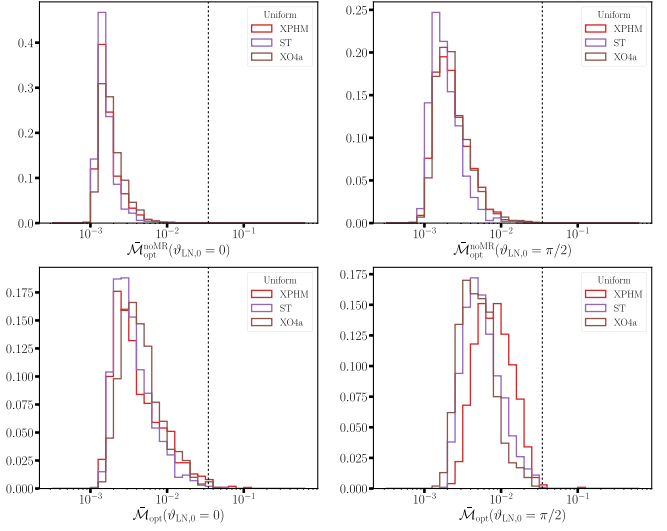


FIG. 26. Comparison of the faithfulness of the MSA implementation of IMRPHEMOMXPHM with that of its more recent SPINTAYLOR implementation (ST in the figure) and the new model IMRPHEMOMXO4A. The parameters used to generate the waveforms are those of the random-uniformly-filled set of Sec. IV C. The baseline model used to measure faithfulness is NRSUR7DQ4.

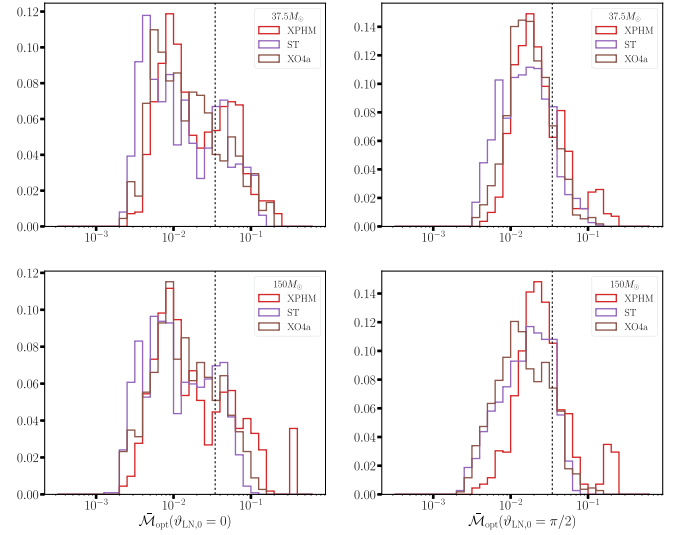


FIG. 27. Same comparison as Fig. 26, but for the discrete parameter set of Sec. IV A. Note the disappearance of a secondary mode containing high-mismatch values for the SPINTAYLOR version of XPHM and IMRPHEMOMXO4A.

see from the figure that the newer SPINTAYLOR version of XPHM and XO4A do not run into this issue.

APPENDIX D: TIMING BENCHMARKS

Though detailed timing benchmarks can be found in, e.g., Refs. [98,128,131], we nonetheless present our own results here. We choose three different $Q = 2$ configurations for our timing: (i) $\theta_{1,2} = 30^\circ$, (ii) $\theta_1 = 90^\circ, \theta_2 \approx 115^\circ$,

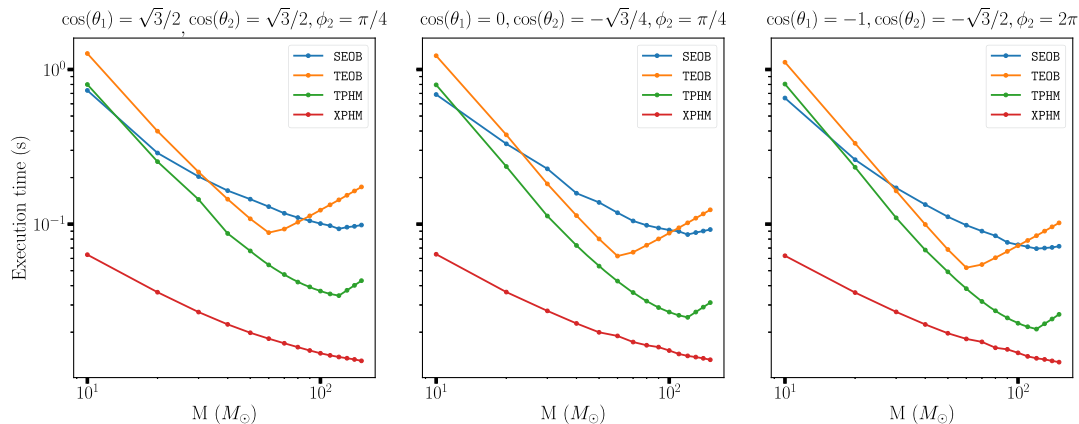


FIG. 28. Timing benchmarks for SEOB, TEOB, TPHM, and XPHM. Each panel gives results from runs with key parameters displayed above with $Q = 2$, $\vartheta_{\text{LN},0} = 0$, $f_0 = 20$ Hz, $d_L = 500$ Mpc. Each individual case was run 1000 times per model in a virtual environment on an AMD EPYC 7453 server.

and (iii) $\theta_1 = 180^\circ$, $\theta_2 = 150^\circ$. We fix $f_0 = 20$ Hz and vary the total mass from $10M_\odot$ to $150M_\odot$ in steps of $10M_\odot$. We further set $\vartheta_{\text{LN},0} = 0$ and $d_L = 500$ Mpc. We run each case 1000 times per model and quote the median values of the results. For the time-domain models, we use a sample

rate of 16384 Hz and for the frequency domain XPHM we use $df = 1/32$ Hz. All runs were performed on an AMD EPYC 7453 server, running under virtualization with `OMP_NUM_THREADS` and `NUMBA_NUM_THREADS` both set to 1. We present the timing results in Fig. 28.

-
- [1] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), *Phys. Rev. Lett.* **116**, 241102 (2016).
 - [2] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **93**, 122003 (2016).
 - [3] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **93**, 122004 (2016); **94**, 069903(E) (2016).
 - [4] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), *Phys. Rev. X* **6**, 041015 (2016).
 - [5] B. P. Abbott *et al.* (LIGO Scientific, Virgo, ASKAP, BOOTES, DES, Fermi-GBM, Fermi-LAT, GRAWITA, INTEGRAL, iPTF, InterPlanetary Network, J-GEM, La Silla-QUEST Survey, Liverpool Telescope, LOFAR, MASTER, MAXI, MWA, Pan-STARRS, PESSTO, Pi of the Sky, SkyMapper, Swift, TAROT, Zadko, Algerian National Observatory, C2PU, TOROS, and VISTA Collaborations), *Astrophys. J. Suppl. Ser.* **225**, 8 (2016).
 - [6] T. D. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), *Phys. Rev. X* **6**, 041014 (2016).
 - [7] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), *Phys. Rev. D* **94**, 064035 (2016).
 - [8] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), *Phys. Rev. Lett.* **116**, 241103 (2016).
 - [9] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), *Classical Quantum Gravity* **34**, 104002 (2017).
 - [10] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Ann. Phys. (Berlin)* **529**, 1600209 (2017).
 - [11] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), *Astrophys. J.* **851**, L35 (2017).
 - [12] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), *Phys. Rev. Lett.* **119**, 141101 (2017).
 - [13] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), *Phys. Rev. Lett.* **118**, 221101 (2017).
 - [14] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. X* **9**, 031040 (2019).
 - [15] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, *Phys. Rev. D* **100**, 023011 (2019).
 - [16] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, *Phys. Rev. D* **101**, 083030 (2020).
 - [17] A. H. Nitz, C. Capano, A. B. Nielsen, S. Reyes, R. White, D. A. Brown, and B. Krishnan, *Astrophys. J.* **872**, 195 (2019).
 - [18] A. H. Nitz, T. Dent, G. S. Davies, S. Kumar, C. D. Capano, I. Harry, S. Mozzon, L. Nuttall, A. Lundgren, and M. Tápai, *Astrophys. J.* **891**, 123 (2020).
 - [19] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **896**, L44 (2020).
 - [20] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **102**, 043015 (2020).
 - [21] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **125**, 101102 (2020).
 - [22] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Phys. Rev. X* **13**, 041039 (2023).

- [23] A. H. Nitz, S. Kumar, Y.-F. Wang, S. Kastha, S. Wu, M. Schäfer, R. Dhurkunde, and C. D. Capano, *Astrophys. J.* **946**, 59 (2023).
- [24] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), *Phys. Rev. Lett.* **119**, 161101 (2017).
- [25] B. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **892**, L3 (2020).
- [26] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **900**, L13 (2020).
- [27] R. Abbott *et al.* (LIGO Scientific, KAGRA, and Virgo Collaborations), *Astrophys. J. Lett.* **915**, L5 (2021).
- [28] S. Olsen, T. Venumadhav, J. Mushkin, J. Roulet, B. Zackay, and M. Zaldarriaga, *Phys. Rev. D* **106**, 043009 (2022).
- [29] D. Williams, [arXiv:2401.08709](https://arxiv.org/abs/2401.08709).
- [30] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. X* **11**, 021053 (2021).
- [31] I. Mandel and A. Farmer, *Phys. Rep.* **955**, 1 (2022).
- [32] M. Mapelli, Formation channels of single and binary stellar-mass black holes, in *Handbook of Gravitational Wave Astronomy*, edited by C. Bambi, S. Katsanevas, and K. D. Kokkotas (Springer, Singapore, 2021).
- [33] D. Gerosa, M. Kesden, E. Berti, R. O’Shaughnessy, and U. Sperhake, *Phys. Rev. D* **87**, 104028 (2013).
- [34] S. Vitale, R. Lynch, R. Sturani, and P. Graff, *Classical Quantum Gravity* **34**, 03LT01 (2017).
- [35] C. L. Rodriguez, M. Zevin, C. Pankow, V. Kalogera, and F. A. Rasio, *Astrophys. J. Lett.* **832**, L2 (2016).
- [36] S. Stevenson, C. P. L. Berry, and I. Mandel, *Mon. Not. R. Astron. Soc.* **471**, 2801 (2017).
- [37] T. A. Apostolatos, C. Cutler, G. J. Sussman, and K. S. Thorne, *Phys. Rev. D* **49**, 6274 (1994).
- [38] L. E. Kidder, *Phys. Rev. D* **52**, 821 (1995).
- [39] C. Cutler and E. E. Flanagan, *Phys. Rev. D* **49**, 2658 (1994).
- [40] E. Poisson and C. M. Will, *Phys. Rev. D* **52**, 848 (1995).
- [41] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), *Astrophys. J.* **818**, L22 (2016).
- [42] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **116**, 241103 (2016).
- [43] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **851**, L35 (2017).
- [44] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Phys. Rev. X* **13**, 011048 (2023).
- [45] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, *Phys. Rev. Lett.* **113**, 151101 (2014).
- [46] T. Damour, *Phys. Rev. D* **64**, 124013 (2001).
- [47] E. Racine, *Phys. Rev. D* **78**, 044021 (2008).
- [48] P. Ajith, *Phys. Rev. D* **84**, 084037 (2011).
- [49] P. Schmidt, M. Hannam, S. Husa, and P. Ajith, *Phys. Rev. D* **84**, 024046 (2011).
- [50] M. Pürrer, M. Hannam, P. Ajith, and S. Husa, *Phys. Rev. D* **88**, 064007 (2013).
- [51] E. Baird, S. Fairhurst, M. Hannam, and P. Murphy, *Phys. Rev. D* **87**, 024035 (2013).
- [52] M. Hannam, D. A. Brown, S. Fairhurst, C. L. Fryer, and I. W. Harry, *Astrophys. J.* **766**, L14 (2013).
- [53] S. Galaudage, C. Talbot, T. Nagar, D. Jain, E. Thrane, and I. Mandel, *Astrophys. J. Lett.* **921**, L15 (2021); **936**, L18 (E) (2022).
- [54] J. Golomb and C. Talbot, *Phys. Rev. D* **108**, 103009 (2023).
- [55] C. Adamcewicz and E. Thrane, *Mon. Not. R. Astron. Soc.* **517**, 3928 (2022).
- [56] H. Tong, S. Galaudage, and E. Thrane, *Phys. Rev. D* **106**, 103019 (2022).
- [57] S. Biscoveanu, T. A. Callister, C.-J. Haster, K. K. Y. Ng, S. Vitale, and W. M. Farr, *Astrophys. J. Lett.* **932**, L19 (2022).
- [58] T. A. Callister, S. J. Miller, K. Chatziioannou, and W. M. Farr, *Astrophys. J. Lett.* **937**, L13 (2022).
- [59] V. Baibhav, Z. Doctor, and V. Kalogera, *Astrophys. J.* **946**, 50 (2023).
- [60] P. Schmidt, F. Ohme, and M. Hannam, *Phys. Rev. D* **91**, 024043 (2015).
- [61] K. G. Arun, A. Buonanno, G. Faye, and E. Ochsner, *Phys. Rev. D* **79**, 104023 (2009); **84**, 049901(E) (2011).
- [62] P. Schmidt, M. Hannam, and S. Husa, *Phys. Rev. D* **86**, 104063 (2012).
- [63] S. J. Miller, M. Isi, K. Chatziioannou, V. Varma, and I. Mandel, *Phys. Rev. D* **109**, 024024 (2024).
- [64] M. Hannam, C. Hoy, J. E. Thompson, S. Fairhurst, V. Raymond *et al.*, *Nature (London)* **610**, 652 (2022).
- [65] E. Payne, S. Hourihane, J. Golomb, R. Udall, R. Udall, D. Davis, and K. Chatziioannou, *Phys. Rev. D* **106**, 104017 (2022).
- [66] R. Macas, A. Lundgren, and G. Ashton, *Phys. Rev. D* **109**, 062006 (2024).
- [67] R. Cotesta, A. Buonanno, A. Bohé, A. Taracchini, I. Hinder, and S. Ossokine, *Phys. Rev. D* **98**, 084028 (2018).
- [68] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, L. E. Kidder, and H. P. Pfeiffer, *Phys. Rev. D* **99**, 064045 (2019).
- [69] H. Estellés, S. Husa, M. Colleoni, D. Keitel, M. Mateu-Lucena, C. García-Quirós, A. Ramos-Buades, and A. Borchers, *Phys. Rev. D* **105**, 084039 (2022).
- [70] G. Pratten, S. Husa, C. Garcia-Quiros, M. Colleoni, A. Ramos-Buades, H. Estelles, and R. Jaume, *Phys. Rev. D* **102**, 064001 (2020).
- [71] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, *Phys. Rev. D* **102**, 064002 (2020).
- [72] G. Riemenschneider, P. Rettegno, M. Breschi, A. Albertini, R. Gamba, S. Bernuzzi, and A. Nagar, *Phys. Rev. D* **104**, 104045 (2021).
- [73] L. Pompili *et al.*, *Phys. Rev. D* **108**, 124035 (2023).
- [74] A. Nagar, P. Rettegno, R. Gamba, A. Albertini, and S. Bernuzzi, *Phys. Rev. D* **108**, 124018 (2023).
- [75] E. E. Flanagan and S. A. Hughes, *Phys. Rev. D* **57**, 4566 (1998).
- [76] M. Pürrer and C.-J. Haster, *Phys. Rev. Res.* **2**, 023151 (2020).
- [77] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, *Phys. Rev. Res.* **1**, 033015 (2019).
- [78] J. Blackman, B. Szilagyi, C. R. Galley, and M. Tiglio, *Phys. Rev. Lett.* **113**, 021101 (2014).

- [79] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, *Phys. Rev. D* **95**, 104023 (2017).
- [80] D. Williams, I. S. Heng, J. Gair, J. A. Clark, and B. Khamesra, *Phys. Rev. D* **101**, 063011 (2020).
- [81] M. Walker, V. Varma, G. Lovelace, and M. A. Scheel, *Classical Quantum Gravity* **40**, 055003 (2023).
- [82] J. Yoo *et al.*, *Phys. Rev. D* **108**, 064027 (2023).
- [83] T. Islam, A. Vajpeyi, F. H. Shaik, C.-J. Haster, V. Varma, S. E. Field, J. Lange, R. O’Shaughnessy, and R. Smith, [arXiv:2309.14473](https://arxiv.org/abs/2309.14473).
- [84] A. Buonanno and T. Damour, *Phys. Rev. D* **59**, 084006 (1999).
- [85] A. Buonanno and T. Damour, *Phys. Rev. D* **62**, 064015 (2000).
- [86] T. Damour, P. Jaranowski, and G. Schaefel, *Phys. Rev. D* **62**, 084011 (2000).
- [87] A. Buonanno, Y. Chen, and T. Damour, *Phys. Rev. D* **74**, 104005 (2006).
- [88] T. Damour, P. Jaranowski, and G. Schäfer, *Phys. Rev. D* **91**, 084024 (2015).
- [89] A. Bohé *et al.*, *Phys. Rev. D* **95**, 044028 (2017).
- [90] S. Babak, A. Taracchini, and A. Buonanno, *Phys. Rev. D* **95**, 024010 (2017).
- [91] T. Hinderer *et al.*, *Phys. Rev. Lett.* **116**, 181101 (2016).
- [92] J. Steinhoff, T. Hinderer, A. Buonanno, and A. Taracchini, *Phys. Rev. D* **94**, 104028 (2016).
- [93] B. D. Lackey, M. Pürrer, A. Taracchini, and S. Marsat, *Phys. Rev. D* **100**, 024002 (2019).
- [94] A. Matas *et al.*, *Phys. Rev. D* **102**, 043023 (2020).
- [95] S. Ossokine *et al.*, *Phys. Rev. D* **102**, 044055 (2020).
- [96] B. Gadre, M. Pürrer, S. E. Field, S. Ossokine, and V. Varma, [arXiv:2203.00381](https://arxiv.org/abs/2203.00381).
- [97] L. M. Thomas, G. Pratten, and P. Schmidt, *Phys. Rev. D* **106**, 104029 (2022).
- [98] A. Ramos-Buades, A. Buonanno, H. Estellés, M. Khalil, D. P. Mihaylov, S. Ossokine, L. Pompili, and M. Shiferaw, *Phys. Rev. D* **108**, 124037 (2023).
- [99] M. van de Meent, A. Buonanno, D. P. Mihaylov, S. Ossokine, L. Pompili, N. Warburton, A. Pound, B. Wardell, L. Durkan, and J. Miller, *Phys. Rev. D* **108**, 124038 (2023).
- [100] M. Khalil, A. Buonanno, H. Estelles, D. P. Mihaylov, S. Ossokine, L. Pompili, and A. Ramos-Buades, *Phys. Rev. D* **108**, 124036 (2023).
- [101] D. P. Mihaylov, S. Ossokine, A. Buonanno, H. Estelles, L. Pompili, M. Pürrer, and A. Ramos-Buades, [arXiv:2303.18203](https://arxiv.org/abs/2303.18203).
- [102] T. Damour and A. Nagar, *Phys. Rev. D* **90**, 024054 (2014).
- [103] S. Bernuzzi, A. Nagar, T. Dietrich, and T. Damour, *Phys. Rev. Lett.* **114**, 161103 (2015).
- [104] A. Nagar, G. Riemenschneider, and G. Pratten, *Phys. Rev. D* **96**, 084045 (2017).
- [105] A. Nagar *et al.*, *Phys. Rev. D* **98**, 104052 (2018).
- [106] S. Akçay, S. Bernuzzi, F. Messina, A. Nagar, N. Ortiz, and P. Rettegno, *Phys. Rev. D* **99**, 044051 (2019).
- [107] S. Akçay, R. Gamba, and S. Bernuzzi, *Phys. Rev. D* **103**, 024014 (2021).
- [108] A. Nagar, G. Riemenschneider, G. Pratten, P. Rettegno, and F. Messina, *Phys. Rev. D* **102**, 024077 (2020).
- [109] D. Chiaromello and A. Nagar, *Phys. Rev. D* **101**, 101501 (2020).
- [110] R. Gamba, S. Akçay, S. Bernuzzi, and J. Williams, *Phys. Rev. D* **106**, 024020 (2022).
- [111] A. Nagar and P. Rettegno, *Phys. Rev. D* **104**, 104004 (2021).
- [112] A. Albertini, A. Nagar, P. Rettegno, S. Albanesi, and R. Gamba, *Phys. Rev. D* **105**, 084025 (2022).
- [113] A. Nagar and S. Albanesi, *Phys. Rev. D* **106**, 064049 (2022).
- [114] A. Albertini, A. Nagar, A. Pound, N. Warburton, B. Wardell, L. Durkan, and J. Miller, *Phys. Rev. D* **106**, 084061 (2022).
- [115] A. Gonzalez, R. Gamba, M. Breschi, F. Zappa, G. Carullo, S. Bernuzzi, and A. Nagar, *Phys. Rev. D* **107**, 084026 (2023).
- [116] T. Andrade *et al.*, [arXiv:2307.08697](https://arxiv.org/abs/2307.08697).
- [117] P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan *et al.*, *Classical Quantum Gravity* **24**, S689 (2007).
- [118] P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan *et al.*, *Phys. Rev. D* **77**, 104017 (2008).
- [119] P. Ajith, M. Hannam, S. Husa, Y. Chen, B. Brügmann *et al.*, *Phys. Rev. Lett.* **106**, 241101 (2011).
- [120] L. Santamaria, F. Ohme, P. Ajith, B. Brügmann, N. Dorband *et al.*, *Phys. Rev. D* **82**, 064016 (2010).
- [121] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. Jiménez Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044006 (2016).
- [122] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Jiménez Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044007 (2016).
- [123] T. Dietrich, S. Bernuzzi, and W. Tichy, *Phys. Rev. D* **96**, 121501 (2017).
- [124] L. London, S. Khan, E. Fauchon-Jones, X. J. Forteza, M. Hannam, S. Husa, C. Kalaghatgi, F. Ohme, and F. Pannarale, *Phys. Rev. Lett.* **120**, 161102 (2018).
- [125] S. Khan, K. Chatziioannou, M. Hannam, and F. Ohme, *Phys. Rev. D* **100**, 024059 (2019).
- [126] T. Dietrich, A. Samajdar, S. Khan, N. K. Johnson-McDaniel, R. Dudi, and W. Tichy, *Phys. Rev. D* **100**, 044003 (2019).
- [127] S. Khan, F. Ohme, K. Chatziioannou, and M. Hannam, *Phys. Rev. D* **101**, 024056 (2020).
- [128] G. Pratten *et al.*, *Phys. Rev. D* **103**, 104056 (2021).
- [129] J. E. Thompson, E. Fauchon-Jones, S. Khan, E. Nitoglia, F. Pannarale, T. Dietrich, and M. Hannam, *Phys. Rev. D* **101**, 124059 (2020).
- [130] H. Estellés, A. Ramos-Buades, S. Husa, and C. García-Quirós, M. Colleoni, L. Haegel, and R. Jaume, *Phys. Rev. D* **103**, 124060 (2021).
- [131] H. Estellés, M. Colleoni, C. García-Quirós, S. Husa, D. Keitel, M. Mateu-Lucena, M. d. L. Planas, and A. Ramos-Buades, *Phys. Rev. D* **105**, 084040 (2022).
- [132] E. Hamilton, L. London, J. E. Thompson, E. Fauchon-Jones, M. Hannam, C. Kalaghatgi, S. Khan, F. Pannarale, and A. Vano-Vinuales, *Phys. Rev. D* **104**, 124027 (2021).
- [133] H. Yu, J. Roulet, T. Venumadhav, B. Zackay, and M. Zaldarriaga, *Phys. Rev. D* **108**, 064059 (2023).
- [134] seobnr5 code repository, <https://git.ligo.org/waveforms/software/pyseobnr>.

- [135] J.E. Thompson, E. Hamilton, L. London, S. Ghosh, P. Kolitsidou, C. Hoy, and M. Hannam, *Phys. Rev. D* **109**, 063012 (2024).
- [136] M. Colleoni *et al.* (to be published).
- [137] E. Hamilton *et al.*, *Phys. Rev. D* **109**, 044032 (2024).
- [138] S. Husa, J. A. González, M. Hannam, B. Brügmann, and U. Sperhake, *Classical Quantum Gravity* **25**, 105006 (2008).
- [139] B. Brügmann, J. A. Gonzalez, M. Hannam, S. Husa, U. Sperhake, and W. Tichy, *Phys. Rev. D* **77**, 024027 (2008).
- [140] A. Puecher, A. Samajdar, G. Ashton, C. Van Den Broeck, and T. Dietrich, *Phys. Rev. D* **109**, 023019 (2024).
- [141] M. Boyle *et al.*, *Classical Quantum Gravity* **36**, 195006 (2019).
- [142] C. Hoy, *Phys. Rev. D* **106**, 083003 (2022).
- [143] J. Mac Uilliam, S. Akcay, and J.E. Thompson, [10.5281/zenodo.10971581](https://arxiv.org/abs/10.5281/zenodo.10971581).
- [144] P.C. Peters and J. Mathews, *Phys. Rev.* **131**, 435 (1963).
- [145] I. M. Romero-Shaw, P. D. Lasky, and E. Thrane, *Mon. Not. R. Astron. Soc.* **490**, 5210 (2019).
- [146] A. H. Nitz, A. Lenon, and D. A. Brown, *Astrophys. J.* **890**, 1 (2019).
- [147] Q.-Y. Yun, W.-B. Han, G. Wang, and S.-C. Yang, [arXiv:2002.08682](https://arxiv.org/abs/2002.08682).
- [148] A. Ramos-Buades, S. Tiwari, M. Haney, and S. Husa, *Phys. Rev. D* **102**, 043005 (2020).
- [149] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **913**, L7 (2021).
- [150] A. K. Lenon, A. H. Nitz, and D. A. Brown, *Mon. Not. R. Astron. Soc.* **497**, 1966 (2020).
- [151] S. Wu, Z. Cao, and Z.-H. Zhu, *Mon. Not. R. Astron. Soc.* **495**, 466 (2020).
- [152] H. L. Iglesias *et al.*, [arXiv:2208.01766](https://arxiv.org/abs/2208.01766).
- [153] I. M. Romero-Shaw, P. D. Lasky, E. Thrane, and J. C. Bustillo, *Astrophys. J. Lett.* **903**, L5 (2020).
- [154] I. M. Romero-Shaw, P. D. Lasky, and E. Thrane, *Astrophys. J.* **940**, 171 (2022).
- [155] K. Breivik, C. L. Rodriguez, S. L. Larson, V. Kalogera, and F. A. Rasio, *Astrophys. J. Lett.* **830**, L18 (2016).
- [156] J. Samsing and E. Ramirez-Ruiz, *Astrophys. J. Lett.* **840**, L14 (2017).
- [157] J. Samsing, A. Askar, and M. Giersz, *Astrophys. J.* **855**, 124 (2018).
- [158] C. L. Rodriguez, P. Amaro-Seoane, S. Chatterjee, and F. A. Rasio, *Phys. Rev. Lett.* **120**, 151101 (2018).
- [159] L. Gondán and B. Kocsis, *Astrophys. J.* **871**, 178 (2019).
- [160] S. Banerjee, *Mon. Not. R. Astron. Soc.* **481**, 5123 (2018).
- [161] B. Liu, D. Lai, and Y.-H. Wang, *Astrophys. J.* **881**, 41 (2019).
- [162] M. Zevin, S. S. Bavera, C. P. L. Berry, V. Kalogera, T. Fragos, P. Marchant, C. L. Rodriguez, F. Antonini, D. E. Holz, and C. Pankow, *Astrophys. J.* **910**, 152 (2021).
- [163] E. Michaely and H. G. Perets, *Mon. Not. R. Astron. Soc.* **498**, 4924 (2020).
- [164] A. S. Hamers, G. Fragione, P. Neunteufel, and B. Kocsis, *Mon. Not. R. Astron. Soc.* **506**, 5345 (2021).
- [165] M. Spera, A. A. Trani, and M. Mencagli, *Galaxies* **10**, 76 (2022).
- [166] J. Samsing, I. Bartos, D. J. D’Orazio, Z. Haiman, B. Kocsis, N. W. C. Leigh, B. Liu, M. E. Pessah, and H. Tagawa, *Nature (London)* **603**, 237 (2022).
- [167] L. Hellström, A. Askar, A. A. Trani, M. Giersz, R. P. Church, and J. Samsing, *Mon. Not. R. Astron. Soc.* **517**, 1695 (2022).
- [168] M. Dall’Amico, M. Mapelli, S. Torniamenti, and M. A. Sedda, *Astron. Astrophys.* **683**, A186 (2023).
- [169] P. Schmidt, I. W. Harry, and H. P. Pfeiffer, [arXiv:1703.01076](https://arxiv.org/abs/1703.01076).
- [170] M. Maggiore, *Gravitational Waves: Theory and Experiments*, Oxford Master Series in Physics Vol. 1 (Oxford University Press, New York, 2007).
- [171] M. Boyle, R. Owen, and H. P. Pfeiffer, *Phys. Rev. D* **84**, 124011 (2011).
- [172] J. J. Sakurai, *Modern Quantum Mechanics*, revised ed. (Addison-Wesley, Reading, MA, 1994).
- [173] D. Brown, S. Fairhurst, B. Krishnan, R. A. Mercer, R. K. Kopparapu, L. Santamaria, and J. T. Whelan, [arXiv:0709.0093](https://arxiv.org/abs/0709.0093).
- [174] D. Gerosa and M. Kesden, *Phys. Rev. D* **93**, 124066 (2016).
- [175] D. Gerosa, G. Fumagalli, M. Mould, G. Cavallotto, D. P. Monroy, D. Gangardt, and V. De Renzi, *Phys. Rev. D* **108**, 024042 (2023).
- [176] M. Kesden, D. Gerosa, R. O’Shaughnessy, E. Berti, and U. Sperhake, *Phys. Rev. Lett.* **114**, 081103 (2015).
- [177] D. Gerosa, M. Kesden, U. Sperhake, E. Berti, and R. O’Shaughnessy, *Phys. Rev. D* **92**, 064016 (2015).
- [178] K. Chatziioannou, A. Klein, N. Yunes, and N. Cornish, *Phys. Rev. D* **95**, 104004 (2017).
- [179] A. Buonanno, Y.-b. Chen, and M. Vallisneri, *Phys. Rev. D* **67**, 104025 (2003); **74**, 029904(E) (2006).
- [180] A. Buonanno, B. Iyer, E. Ochsner, Y. Pan, and B. Sathyaprakash, *Phys. Rev. D* **80**, 084043 (2009).
- [181] K. Chatziioannou, A. Klein, N. Yunes, and N. Cornish, *Phys. Rev. D* **88**, 063011 (2013).
- [182] A. Ramos-Buades, P. Schmidt, G. Pratten, and S. Husa, *Phys. Rev. D* **101**, 103014 (2020).
- [183] D. Keppel, D. A. Nichols, Y. Chen, and K. S. Thorne, *Phys. Rev. D* **80**, 124015 (2009).
- [184] B. Bruegmann, J. A. Gonzalez, M. Hannam, S. Husa, and U. Sperhake, *Phys. Rev. D* **77**, 124047 (2008).
- [185] C. Kalaghatgi and M. Hannam, *Phys. Rev. D* **103**, 024024 (2021).
- [186] P. Kolitsidou, J. E. Thompson, and M. Hannam, [arXiv:2402.00813](https://arxiv.org/abs/2402.00813).
- [187] S. Ghosh, P. Kolitsidou, and M. Hannam, *Phys. Rev. D* **109**, 024061 (2024).
- [188] D. Gerosa, A. Lima, E. Berti, U. Sperhake, M. Kesden, and R. O’Shaughnessy, *Classical Quantum Gravity* **36**, 105003 (2019).
- [189] D. Gangardt, D. Gerosa, M. Kesden, V. De Renzi, and N. Steinle, *Phys. Rev. D* **106**, 024019 (2022).
- [190] D. Gerosa, M. Mould, D. Gangardt, P. Schmidt, G. Pratten, and L. M. Thomas, *Phys. Rev. D* **103**, 064067 (2021).
- [191] V. De Renzi, D. Gerosa, G. Pratten, P. Schmidt, and M. Mould, *Phys. Rev. D* **106**, 084040 (2022).
- [192] L. M. Thomas, P. Schmidt, and G. Pratten, *Phys. Rev. D* **103**, 083022 (2021).
- [193] N. K. Johnson-McDaniel, S. Kulkarni, and A. Gupta, *Phys. Rev. D* **106**, 023001 (2022).

- [194] M. Mould and D. Gerosa, *Phys. Rev. D* **105**, 024076 (2022).
- [195] J. Aasi *et al.* (LIGO Scientific Collaboration), *Classical Quantum Gravity* **32**, 074001 (2015).
- [196] Updated Advanced LIGO sensitivity design curve, <https://dcc.ligo.org/LIGO-T1800044/public>.
- [197] L1 calibrated sensitivity spectra, January 4, 2020, <https://dcc.ligo.org/LIGO-G2100675>.
- [198] I. Harry, S. Privitera, A. Bohé, and A. Buonanno, *Phys. Rev. D* **94**, 024012 (2016).
- [199] B. J. Owen, *Phys. Rev. D* **53**, 6749 (1996).
- [200] E. E. Flanagan and S. A. Hughes, *Phys. Rev. D* **57**, 4535 (1998).
- [201] M. L. Waskom, *J. Open Source Software* **6**, 3021 (2021).
- [202] SXS Collaboration catalog tools, <https://github.com/sxs-collaboration/catalog-tools>.
- [203] LIGO Scientific, Virgo, and KAGRA Collaborations, LVK Algorithm Library—LALSuite, Free software (GPL) (2018), 10.7935/GT1W-FZ16.
- [204] M. Boyle and M. Scheel, The sxs package, 2023, 10.5281/zenodo.4034006.
- [205] M. Boyle, *Phys. Rev. D* **87**, 104006 (2013).
- [206] M. Boyle, L. E. Kidder, S. Ossokine, and H. P. Pfeiffer, [arXiv:1409.4431](https://arxiv.org/abs/1409.4431).
- [207] M. Boyle, *Phys. Rev. D* **93**, 084031 (2016).
- [208] M. Boyle, D. Iozzo, and L. C. Stein, MOBLE/SCRI: v1.2, 2020.
- [209] M. Breschi, R. Gamba, and S. Bernuzzi, *Phys. Rev. D* **104**, 042001 (2021).
- [210] Noise curves used for simulations in the update of the observing scenarios paper, 2022, <https://dcc.ligo.org/LIGO-T2000012/public>.
- [211] S. Fairhurst, R. Green, C. Hoy, M. Hannam, and A. Muir, *Phys. Rev. D* **102**, 024055 (2020).
- [212] G. Ashton *et al.*, *Astrophys. J. Suppl. Ser.* **241**, 27 (2019).
- [213] I. M. Romero-Shaw *et al.*, *Mon. Not. R. Astron. Soc.* **499**, 3295 (2020).
- [214] BILBY_PIPE, 2020, <https://lscsoft.docs.ligo.org/bilby-pipe/master/index.html>.
- [215] J. Skilling, *AIP Conf. Proc.* **735**, 395 (2004).
- [216] J. Skilling, *Bayesian Anal.* **1**, 833 (2006).
- [217] J. S. Speagle, *Mon. Not. R. Astron. Soc.* **493**, 3132 (2020).
- [218] S. Koposov, J. Speagle, K. Barbary, G. Ashton, E. Bennett, J. Buchner, C. Scheffler, B. Cook, C. Talbot, J. Guillochon, P. Cubillos, A. A. Ramos, B. Johnson, D. Lang, Ilya, M. Dartiailh, A. Nitz, A. McCluskey, and A. Archibald, JOSHSPEAGLE/DYNESTY v2.1.3, 2023.
- [219] P. A. R. Ade *et al.* (Planck Collaboration), *Astron. Astrophys.* **594**, A13 (2016).
- [220] K. K. Y. Ng, S. Vitale, A. Zimmerman, K. Chatziioannou, D. Gerosa, and C.-J. Haster, *Phys. Rev. D* **98**, 083007 (2018).
- [221] J. Lin, *IEEE Trans. Inf. Theory* **37**, 145 (1991).
- [222] E. Ochsner and R. O’Shaughnessy, *Phys. Rev. D* **86**, 104037 (2012).
- [223] E. Hamilton and M. Hannam, *Phys. Rev. D* **98**, 084018 (2018).
- [224] Wikipedia, List of gravitational wave observations, <https://en.wikipedia.org/w/index.php?title=List-of-gravitational-wave-observations&oldid=1188064092> (December 5, 2023).
- [225] C. R. Harris *et al.*, *Nature (London)* **585**, 357 (2020).
- [226] J. D. Hunter, *Comput. Sci. Eng.* **9**, 90 (2007).
- [227] Pandas Development Team, pandas-dev/pandas: Pandas, 2020.
- [228] Wes McKinney, in *Proceedings of the 9th Python in Science Conference*, edited by Stéfan van der Walt and Jarrod Millman, pp. 56–61, 10.25080/Majora-92bf1922-00a.
- [229] J. Veitch *et al.*, *Phys. Rev. D* **91**, 042003 (2015).
- [230] C. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, *Publ. Astron. Soc. Pac.* **131**, 024503 (2019).
- [231] Black Hole Perturbation Toolkit, <https://bhptoolkit.org/>.
- [232] C. Hoy and V. Raymond, *SoftwareX* **15**, 100765 (2021).
- [233] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *SoftwareX* **13**, 100658 (2021).
- [234] <https://www.gw-openscience.org>.