Accuracy of binary black hole waveform models for aligned-spin binaries

Prayush Kumar,[1] Tony Chu,[2] Heather Fong,[1,3] Harald P. Pfeiffer,[1,4,5] Michael Boyle,[6]
Daniel A. Hemberger,[7] Lawrence E. Kidder,[6] Mark A. Scheel,[7] and Bela Szilagyi[7,8]

[1]*Canadian Institute for Theoretical Astrophysics, 60 St. George Street, University of Toronto, Toronto, Ontario M5S 3H8, Canada*
[2]*Department of Physics, Princeton University, Jadwin Hall, Princeton, New Jersey 08544, USA*
[3]*Department of Physics, University of Toronto, 60 St. George Street, Toronto, Ontario M5S 3H8, Canada*
[4]*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, 14476 Potsdam-Golm, Germany*
[5]*Canadian Institute for Advanced Research, 180 Dundas Street West, Toronto, Ontario M5G 1Z8, Canada*
[6]*Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, New York 14853, USA*
[7]*Theoretical Astrophysics 350-17, California Institute of Technology, Pasadena, California 91125, USA*
[8]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA*

(Received 28 January 2016; published 25 May 2016)

Coalescing binary black holes are among the primary science targets for second generation ground-based gravitational wave detectors. Reliable gravitational waveform models are central to detection of such systems and subsequent parameter estimation. This paper performs a comprehensive analysis of the accuracy of recent waveform models for binary black holes with aligned spins, utilizing a new set of 84 high-accuracy numerical relativity simulations. Our analysis covers comparable mass binaries (mass-ratio $1 \leq q \leq 3$), and samples independently both black hole spins up to a dimensionless spin magnitude of 0.9 for equal-mass binaries and 0.85 for unequal mass binaries. Furthermore, we focus on the high-mass regime (total mass $\gtrsim 50 M_\odot$). The two most recent waveform models considered (PhenomD and SEOBNRv2) both perform very well for signal *detection*, losing less than 0.5% of the recoverable signal-to-noise ratio ρ, except that SEOBNRv2's efficiency drops slightly for both black hole spins aligned at large magnitude. For *parameter estimation*, modeling inaccuracies of the SEOBNRv2 model are found to be smaller than systematic uncertainties for moderately strong GW events up to roughly $\rho \lesssim 15$. PhenomD's modeling errors are found to be smaller than SEOBNRv2's, and are generally irrelevant for $\rho \lesssim 20$. Both models' accuracy deteriorates with increased mass ratio, and when at least one black hole spin is large and aligned. The SEOBNRv2 model shows a pronounced disagreement with the numerical relativity simulation in the merger phase, for unequal masses and simultaneously both black hole spins very large and aligned. Two older waveform models (PhenomC and SEOBNRv1) are found to be distinctly less accurate than the more recent PhenomD and SEOBNRv2 models. Finally, we quantify the bias expected from all four waveform models during parameter estimation for several recovered binary parameters: chirp mass, mass ratio, and effective spin.

DOI: 10.1103/PhysRevD.93.104050

I. INTRODUCTION

Gravitational-wave (GW) astronomy is entering an exciting time with a concerted global effort to detect gravitational waves with ground-based facilities. In North America, the Advanced Laser Interferometer Gravitational-Wave Observatory (aLIGO) operates two 4-km scale GW detectors [1,2], located in Hanford, Washington, and Livingston, Louisiana. Both of these instruments began their first observation run "O1" in September 2015, which is scheduled to last for four months [3], operating at more than three times the strain sensitivity of the initial LIGO detectors [4]. In addition, the upgrades to the Virgo detector [5], construction of the KAGRA detector [6,7], and planning of the LIGO-India detector [8] are under way.

Binary black holes (BBHs) are among the most promising GW sources for detection with aLIGO. Compact binary merger rate estimates suggest a GW detection rate of approximately a few tens of binary black holes (BBH) every year [9]. The actual masses of astrophysical black holes are uncertain, but observations and population synthesis studies suggest that BHs formed from stellar core collapse can have masses up to and higher than $34 M_\odot$ [10,11]. Also, recent measurements using continuum fitting and x-ray reflection fitting suggest that black holes can have high spin, with the BH angular momentum in dimensionless units exceeding 0.8 [12–18]. Therefore the observations of GWs emitted by spinning BBHs will allow us to understand the spin-spin and spin-orbit dynamics of the two-body system, apart from allowing us to test strong-field dynamics of general relativity. Unlocking the full scientific potential of BBH GW observations, however, will require us to detect as many such GWs as possible, and to accurately characterize and classify the BBH systems that emitted them.

Optimal GW searches for stellar-mass BBH signals are based on matched-filtering the detector data with modeled waveforms. Past LIGO-Virgo searches for compact binaries used models of nonspinning BBH inspirals as filtering templates, e.g. [19–22] (with the exception of [23]). Recent progress has moved the collaboration towards using inspiral-merger-ringdown models of aligned-spin BBHs as filters. It has been shown that doing so will significantly increase search efficiency against generically oriented binaries [24]. Furthermore, it has been shown that complete inspiral-merger-rindown (IMR) waveforms are needed for the observation of BBHs with $M \gtrsim 12M_\odot$ [25]. It is therefore important for the aligned-spin candidate waveform models to be carefully examined for accuracy in capturing the entire coalescence process, including merger and ringdown. Early work on assessing the accuracy of different waveform models has focused on model-model comparisons [26–31]. In the absence of more accurate reference waveforms, such studies have been limited by the most accurate model they consider, and have used model-model agreement to make statements about model accuracy. More recently, there have been extensive studies of waveform models involving high-accuracy numerical relativity (NR) simulations [32–41]. However, most of these investigations have focused on binaries with zero spins or modest spin magnitudes. Furthermore, while recently developed waveform models [42,43] have used an unprecedented amount of information from NR to increase the accuracy of their merger description, their accuracy has not been investigated in a systematic manner over the BBH parameter space.

In this paper, we explore the accuracy of recent BBH waveform models using new high-accuracy NR simulations, from the perspective of their application to GW astronomy. The 84 numerical waveforms were computed with the Spectral Einstein Code (SpEC) [44] and are presented in detail in a companion paper [45]. This catalog covers nonprecessing configurations, i.e. BBHs with spin vectors parallel or antiparallel to the orbital angular momentum. More specifically, it spans mass ratios $q \equiv m_1/m_2 \in [1, 3]$, and spin projection $\chi_i \equiv \vec{\chi}_i \cdot \hat{L} \in [-0.9, +0.9]$, where $i = 1$, 2 labels the two black holes, with mass m_i and dimensionless angular momentum $\vec{\chi}_i \equiv \vec{S}_i/m_i^2$, and where $\hat{L}$ denotes the unit vector along the direction of the orbital angular momentum. The median length of these simulations is 24 orbits, allowing us to extend our comparisons down to binary masses as low as 40–70$M_\odot$ (depending on configuration, cf. Fig. 1) while still covering aLIGO's frequency band above 15 Hz. We restrict probed total masses below $150M_\odot$.

The waveform models we investigate include two recently calibrated Effective-One-Body (EOB) models (namely, SEOBNRv1 and SEOBNRv2) [42,46], and two recent phenomenological models (namely IMRPhenomC and IMRPhenomD) [43,47]. Both EOB and IMRPhenom models are constructed using (different) extensions of post-Newtonian (PN) dynamics of compact binaries, with free parameters that are calibrated to NR simulations. Figure 2 shows the parameters of the NR simulations used in the construction of each of these models. We probe their accuracy in different corners of the component spin space in this work. We model detector sensitivity using the zero-detuning high-power noise power spectral density for aLIGO [48], and use $f_{\mathrm{low}} = 15$ Hz as the lower frequency cutoff for filtering.

We perform the following studies. First, we measure the faithfulness of different waveform models by calculating their noise-weighted overlaps against the new NR waveforms. We find that (i) both SEOBNRv2 and IMRPhenomD are faithful to our NR simulations over most of the spin and mass-ratio parameter space (overlaps $>99\%$), with overlaps falling to 97–98% when component spins are antiparallel to each other. However, when both BHs have large positive-aligned spins, IMRPhenomD fares significantly better, while the overlaps between SEOBNRv2 and NR fall to 80%; (ii) both SEOBNRv1 and IMRPhenomC show larger disagreement with NR, and we clearly show that they have been superseded by their more recent versions in accuracy. Specifically, we find that SEOBNRv1 deteriorates when the spin on the *larger* BH is $\gtrsim +0.5$ (with overlaps falling to 80%), while IMRPhenomC performs poorly when the spin magnitude on the *smaller* BH exceeds ≈0.5, with overlaps falling below 80%. While we do not find a strong correlation

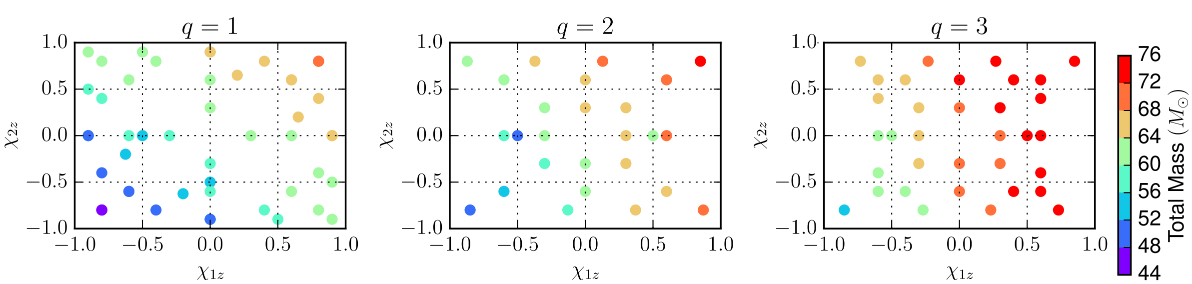

FIG. 1. Parameter space coverage of the simulations considered here. For mass ratio $q = \{1, 2, 3\}$ we indicate the spin components (χ_1, χ_2) projected onto the orbital angular momentum. Each point is color coded by the lowest total mass to which the waveform can se scaled, such that the initial GW frequency remains $\gtrsim 15$ Hz. In the $q = 1$ panel, each simulation is plotted twice at $(\chi_1, \chi_2) \to (\chi_2, \chi_1)$ to represent the symmetry under exchange of the two objects.

between model accuracy and mass ratio for the more recent SEOBNRv2 and IMRPhenomD models, we do find that both SEOBNRv1 and IMRPhenomC deteriorate in accuracy with increasing binary mass ratio. This accuracy enhancement in both model families comes in part from improvements in parametrization, and in part from extensive calibration to NR simulations. While SEOBNRv1 was calibrated to only two equal-mass spinning NR simulations, SEOBNRv2 absorbs information from as many as 30. On the other hand, IMRPhenomD improves over IMRPhenomC by adopting a double-spin parametrization, and using information from NR simulations with mass ratios up to $q = 18$. In addition, we also show that the IMRPhenomD and SEOBNRv2 models are indistinguishable from NR simulations in large regions of the considered parameter space up to effective signal-to-noise ratios (SNR) of 20 and 15, respectively, albeit with significant dependence on the mass ratio and spins.

In our second study, we assess the viability of waveform models for aLIGO *detection* searches for high-mass BBHs. We compute the overlaps between each rescaled NR waveform and a large set of model waveforms that sample the binary mass and spin parameter space densely. From this, we recover the maximum fraction of the optimal signal SNR that any waveform model can recover—with the only loss being caused by intrinsic inaccuracies of the model itself. We find that (i) both SEOBNRv2 and IMRPhenomD recover more than 99.5% of the optimal SNR over most of the mass and spin parameter space, except when both BHs have large aligned spins, where the inaccuracies of SEOBNRv2 lead to a drop in SNR recovery to 97% of its optimal value; and (ii) both IMRPhenomC and SEOBNRv1 compensate for their intrinsic inaccuracy with maximization of SNR over waveform parameters, recovering $>98\%$ of the optimal SNR over most (not all) of the parameter space considered within their domain of applicability. This is a manifestation of the efficient utilization of the intrinsic mass and spin degeneracy of gravitational waveforms [49,50], allowing IMRPhenomC to be a fairly *effectual* model despite being unable to reproduce NR waveforms with identical masses and spins. On the other hand, that SEOBNRv2 has low fitting factors for large aligned component spins, despite maximizing over intrinsic binary parameters, hints that the model might benefit from a different parametrization of the transition regime from late-inspiral to ringdown. Overall, we conclude that both SEOBNRv2 and IMRPhenomD are viable for modeling waveforms in aLIGO searches aimed at comparable mass-ratio high-mass BBHs. This validates the use of SEOBNRv2 by current and future aLIGO searches. We note that due to the high computational cost of evaluating the SEOBNRv2 model, aLIGO data analyses use its reduced-order model [51] which mitigates this drawback.

Our third study concerns BBH parameter estimation from GW signals, which, when accurately done, will provide unique insight into astrophysical processes involving stellar evolution, compact binary formation and evolution [52–65]. Full Bayesian analyses of GW signals require models that faithfully reproduce real GWs in order to map them back to the properties of their source binaries. Model inaccuracies manifest themselves as biases in the recovered values of the mass and spin parameters of BBHs. Therefore, we investigate the level of systematic biases that using different (aforementioned) inspiral-merger-ringdown (IMR) waveform models will incur. We find that (i) binary chirp mass is best recovered by IMRPhenomD (within ± 2–5%), especially for spin-aligned systems. For systems with antialigned spins, the systematic bias in chirp mass is similar for both IMRPhenomD and SEOBNRv2, rising above 5% at the higher end of the sampled binary mass range. (ii) Total mass is recovered with similar accuracy (2–5%) by both SEOBNRv2 and IMRPhenomD, although not as well as both recover $\mathcal{M}_c$. The older SEOBNRv1 and IMRPhenomC models, while furnishing larger biases overall, recover M better than $\mathcal{M}_c$. (iii) Binary mass ratio is also best recovered by IMRPhenomD (within 10–15%), with SEOBNRv2 systematically underestimating mass ratios for binaries with antialigned spins, and overestimating for positive-aligned spins (by up to $\pm 20\%$). (iv) We test the recovery of the PN effective-spin combination $\chi_{\rm eff}$ that appears at leading order in inspiral phasing. As with the mass parameters, we find that IMRPhenomD recovers $\chi_{\rm eff}$ best (within ± 0.1), especially for strongly spin-aligned binaries. While SEOBNRv2 shows marginally higher spin biases (up to ± 0.15) for high-mass binaries with $M \gtrsim 100 M_\odot$, both SEOBNRv1 and IMRPhenomC models incur higher biases in spin recovery (up to ± 0.25) over different regions of the parameter space. Overall, we find that both SEOBNRv2 and IMRPhenomD have comparable accuracy in terms of parameter recovery, with IMRPhenomD performing better of the two for binaries with large aligned $\chi_{\rm eff}$ and/or high masses.

We note that a recent study [50] shows that the biases we find for SEOBNRv2 will become comparable to statistical uncertainty in spin recovery at SNRs ≈ 20–30. However, a more detailed Markov-chain Monte Carlo (MCMC) analysis will be needed to (i) determine the same for highly spinning binaries, where SEOBNRv2 deviates significantly from NR, and (ii) to compare the statistical biases for the IMRPhenomD model with its (much smaller) systematic biases that we report here. We also recall that the present study applies to high-mass BBHs, with total masses $\gtrsim 50 M_\odot$. At lower binary masses, the NR waveforms no longer cover the entire aLIGO frequency band, and one needs either longer NR simulations or one needs to hybridize the existing simulations with PN inspiral waveforms. We also note that we plan to follow up on the interesting patterns seen in the high-spin/high-spin corner of the BBH parameter space in the future in order to better understand the accuracy of analytical models there.

The remainder of the paper is organized as follows. Section II summarizes the salient features of the new catalog of NR simulations used in this analysis, describes different measures of waveform-model accuracy, and summarizes the different waveform models analyzed in this paper. In Sec. III we present overlap comparisons of different waveform models with our NR waveforms. In Sec. IV we measure the efficacy of different waveform models as detection filters. In Sec. V we analyze the systematic biases in the recovery of binary mass and spin parameters, associated with the different waveform models we consider in this paper. Finally, in Sec. VI we summarize and discuss our results.

II. METHODOLOGY

A. Numerical relativity simulations

The BBH simulations considered here were performed with the Spectral Einstein Code (SpEC) [44], and were presented in [45]. Initial data are constructed with the pseudospectral elliptic solver described in [66], using the extended conformal thin-sandwich method [67] with quasiequilibrium boundary conditions [68]. Evolutions use a first-order representation of the generalized harmonic system [69–72] with a damped-harmonic gauge [73]. The computational grid is adaptively refined [74], and the excision boundaries are dynamically adjusted to follow the apparent horizons [73,75,76]. Interdomain boundary conditions are enforced with a penalty method [77,78], and constraint-preserving outgoing-wave conditions [79–81] are imposed at the outer boundary.

Our simulations consist of 84 configurations at mass ratios $q = m_1/m_2 = \{1, 2, 3\}$. All simulations are nonprecessing; i.e. the dimensionless spin $\vec{\chi}_{1,2}$ of each hole is either aligned or antialigned with the direction of the orbital angular momentum $\hat{L}$. The parameters of all simulations are plotted in Fig. 1. A total of 22 simulations have only one hole spinning, 32 have both holes spinning with equal spin magnitudes, and the remaining 30 have both holes spinning with unequal spin magnitudes. The spin components along $\hat{L}$ range over $-0.9 \leq \chi_{1,2} \leq 0.9$. All evolutions have initial orbital eccentricity $e < 10^{-4}$. The evolutions include an average of 24 orbits, with the shortest having 21.5 orbits and the longest having 32 orbits. BBH waveforms can be rescaled to any total mass $M = m_1 + m_2$. Figure 1 also indicates the lowest total mass $M_{\rm low}$ for each configuration, such that the rescaled waveform covers the aLIGO frequency range for $f \geq f_{\rm low} = 15$ Hz.

B. Accuracy measures

We can define an inner product between two waveforms h_1 and h_2 as

$$\langle h_1, h_2 \rangle \equiv \int_{f_{\rm low}}^{f_{\rm high}} \frac{\tilde{h}_1(f)\tilde{h}_2^*(f)}{S_n(|f|)} \mathrm{d}f, \quad (1)$$

where $\tilde{h}(f)$ represents the Fourier transform of h, the superscript * represents complex conjugation, and $S_n(|f|)$ is the power spectral density of detector noise. We integrate the inner product over the frequency interval $[f_{\rm low}, f_{\rm high}]$, which spans the sensitive band of the GW detector. In this paper we use $f_{\rm low} = 15$ Hz, $f_{\rm high} = 4096$ Hz, and the zero-detuning high-power noise curve [48] to model aLIGO at design sensitivity. This inner product is sensitive to an arbitrary phase and time shift between the two waveforms. Since both of these are extrinsic parameters and of little astrophysical interest, we maximize the inner product over them to define the maximized overlap $\mathcal{O}$,

$$\mathcal{O}(h_1, h_2) = \max_{\phi_0, t_0} \frac{\langle h_1, h_2 \rangle}{\sqrt{\langle h_1, h_1 \rangle \langle h_2, h_2 \rangle}}. \quad (2)$$

This overlap measures the correlation between any two given waveforms. We use the overlap to measure the accuracy of analytical waveform families by comparing to NR waveforms with identical physical parameters. This assumes that the latter closely reproduce *true* waveforms in nature. The error analysis in [45] shows that numerical errors of the NR waveforms cause mismatches $1 - \mathcal{O} < 5 \times 10^{-4}$, with a median value of $1 - \mathcal{O} \sim 3 \times 10^{-4}$. Therefore, we expect that overlaps computed here to be influenced by NR errors only for $\mathcal{O} > 0.9995$.

GW detection searches use a discrete set of waveforms, called a "template bank," to filter detector data. This bank spans the range of mass and spin parameters considered in the search, and can be visualized as a multidimensional lattice. There are two sources of SNR loss from using template banks. First, the density of templates in the parameter space. This is a free parameter which trades the loss of SNR with the number of templates to be searched. Customarily, a 3% loss in SNR is viewed as acceptable. The second source of error—the focus of this paper—is the accuracy of the underlying analytical waveform family that is used to generate the templates. The second source is somewhat compensated for by the freedom of maximizing the recovered SNR over intrinsic binary parameters; i.e., it does not matter *which* template waveform fits a given signal in a detection search. To investigate the SNR loss due to the second factor alone, we compute the *fitting factors* of different waveform models as follows. For each combination p of $(M = m_1 + m_2, q = m_1/m_2, \chi_1, \chi_2)$ that we rescale our NR waveforms to, we sample a set $\mathcal{S}_p$ of 8,000,000 points in the vicinity of the true parameters (p) and compute the overlaps between the NR waveform $h^{\rm NR}(p)$ and model waveforms $h^{\rm M}(i)$ for all points $i \in \mathcal{S}_p$. Finally, the fitting factor FF of model M for signal parameters p is given by

$$\mathrm{FF}^{\rm M}(p) = \max_{i \in \mathcal{S}_p} \mathcal{O}(h^{\rm NR}(p), h^{\rm M}(i)). \quad (3)$$

FF is therefore the maximum fraction of the optimal SNR that a waveform model can recover for a GW signal with parameters p. The deviation of the fitting factor from unity quantifies the loss in SNR due to model inaccuracy alone, and is in addition to any loss incurred due to the discreteness of the actual template bank used in a GW search.

C. Waveform models

In this paper we investigate the following waveform models for aligned-spin binary black holes.

1. Effective one-body

Buonanno and Damour [82] developed an effective-one-body (EOB) approach to the two-body problem in general relativity. Over the past decade parametrized EOB models capable of describing the complete binary coalescence process have been developed and calibrated using information from NR simulations [42,46,82–90,90–94]. In the spin EOB framework, the dynamics of two compact objects of masses m_1 and m_2 and spins $\vec{\chi}_1$ and $\vec{\chi}_2$ is mapped onto the dynamics of an effective particle of mass $\mu = m_1 m_2/(m_1 + m_2)$ and spin $\vec{\chi}_*$ moving in a deformed-Kerr background with mass $M = m_1 + m_2$ and spin $\vec{\chi}_{\rm Kerr}$. The parametrized spin mapping $\{\vec{\chi}_1, \vec{\chi}_2\} \to \vec{\chi}_*$ and the deformation of the background from Kerr is chosen to ensure that the inspiral dynamics of the test particle reproduce the PN-expanded dynamics of the original two-body system. Free parameters are introduced into the models that represent unknown, higher-order PN terms, or additional physical effects like corrections due to noncircularity. Such free parameters are calibrated with NR simulations. With the EOB system specified, a Hamiltonian $H_{\rm EOB}$ to describe its conservative dynamics can be written [42,46]. The nonconservative dynamics is contained in a parametrized radiation-reaction term that is inserted in the equations of motion. This term sums over the outgoing GW modes and is calibrated to reproduce NR simulations. The combination of these two pieces describes the binary inspiral through to merger, at which point a ringdown waveform is stitched on to the inspiral-merger waveform. This BH ringdown waveform is constructed as a linear superposition of the dominant quasinormal modes (QNMs) of the Kerr BH formed at merger [90,95], with amplitude and phase of each QNM mode determined by the stitching process.

In this paper we focus on two recent aligned-spin EOB models which are calibrated to NR: SEOBNRv1 and SEOBNRv2 [42,46]. The SEOBNRv1 model has been calibrated to five nonspinning simulations with $q = m_1/m_2 = \{1, 2, 3, 4, 6\}$ and two equal-mass equal-spin simulations [46]. It models binaries with nonprecessing BH spins in the range $-1 \le \chi_{1,2} \le +0.6$. The improved SEOBNRv2 model has been calibrated to a significantly larger set of NR simulations, including eight nonspinning simulations with $q \le 8$ and 30 spinning, nonprecessing simulations [42]. This model is capable of modeling binaries with nonprecessing component spins over the range $-1 \le \chi_{1,2} \le +1$. We refer the reader to [42,46] for a comprehensive summary of the technical details of these two models. We note that due to the high computational cost of evaluating these models, we and both current LIGO searches use a reduced-order model of SEOBNRv2 [51] for search templates.

2. Phenomenological

Offline GW searches and parameter estimation efforts aimed at binary black holes involve filtering the detector data with modeled waveforms in the *frequency* domain. One way to minimize their computational cost is to use frequency-domain closed-form GW models as search filters. Past LIGO-Virgo searches used the TaylorF2 model (see, e.g., [96]), although with the significant limitation that TaylorF2 describes only the inspiral phase. A phenomenological model (IMRPhenomC) based on it has been developed to also capture the plunge and merger phase waveforms [47]. This model uses TaylorF2 phasing and amplitude prescriptions during the early inspiral, and stitches on an analytic *Ansatz* for GW phasing and amplitude during the late-inspiral, plunge and merger phases. These *Ansätze* are written as polynomials in $f^{1/3}$, where f is the instantaneous gravitational-wave frequency, and the associated coefficients are treated as free parameters. In the ringdown regime, IMRPhenomC models binary phasing as a linear function in f, capturing the effect of the leading QNM with a Lorentzian. The model is calibrated to reproduce accurate NR waveforms for nonprecessing binaries with mass ratios $q \le 4$ and BH spins between $[-0.75, +0.83]$, produced by different groups [97–101]. The free parameters are interpolated over the binary mass and spin parameter space as polynomials in the symmetric mass ratio η and mass-weighted spin $\chi_{\rm mw}$,

$$\chi_{\rm mw} \coloneqq \frac{m_1}{m_1 + m_2}\chi_1 + \frac{m_2}{m_1 + m_2}\chi_2, \tag{4}$$

to obtain IMRPhenomC inspiral-merger-ringdown waveforms at arbitrary binary masses and spins. We refer the reader to Ref. [47] for a complete description of this model.

The very recent IMRPhenomD model [43] improves upon IMRPhenomC in several crucial aspects: (i) use of both component spins to model the inspiral phasing; (ii) use of the spin parameter $\chi_{\rm eff}$ [102],

$$\chi_{\rm eff} \coloneqq \chi_{\rm mw} - \frac{38\eta}{113}(\chi_1 + \chi_2) \tag{5}$$

[with symmetric mass-ratio $\eta = m_1 m_2/(m_1 + m_2)^2$], to capture the late-inspiral/plunge phase; (iii) use of

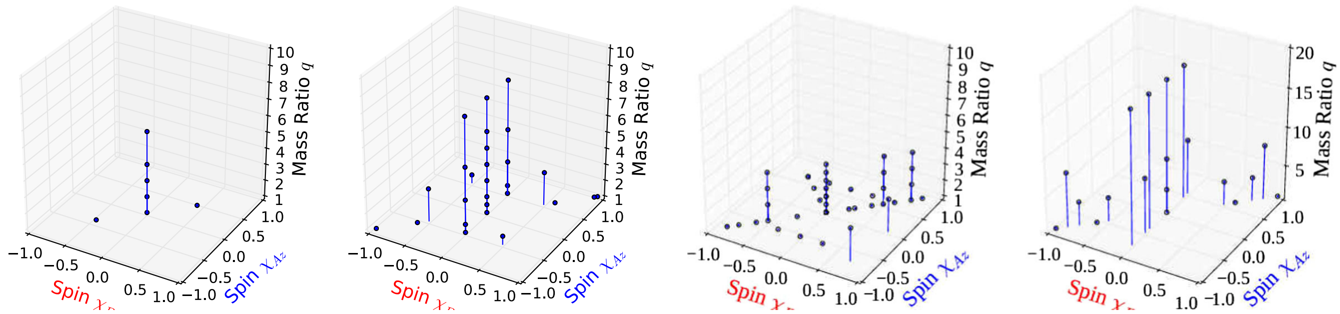

FIG. 2. Parameters of numerical-relativity simulations used to calibrate the various inspiral-merger-ringdown models that we investigate in this paper, i.e. (left to right) SEOBNRv1, SEOBNRv2, IMRPhenomC and IMRPhenomD.

(uncalibrated) EOB + NR hybrid waveforms to constrain free parameters; and (iv) use of several high mass-ratio NR simulations to extend the range of validity of the model. The simulations used to calibrate IMRPhenomD sample component spins more densely than the set used for IMRPhenomC, and cover mass ratios up to $q = 18$. We refer the reader to Refs. [43,103] for further details of IMRPhenomD.

III. FAITHFULNESS ANALYSIS

We now proceed to a comparison of the NR waveforms introduced in Sec. II A with the analytical waveform models introduced in Sec. II C, beginning with an analysis of their faithfulness [cf. Eq. (2)]. We rescale the NR waveforms to a range of total masses, and compute overlaps with model waveforms with identical BH parameters.

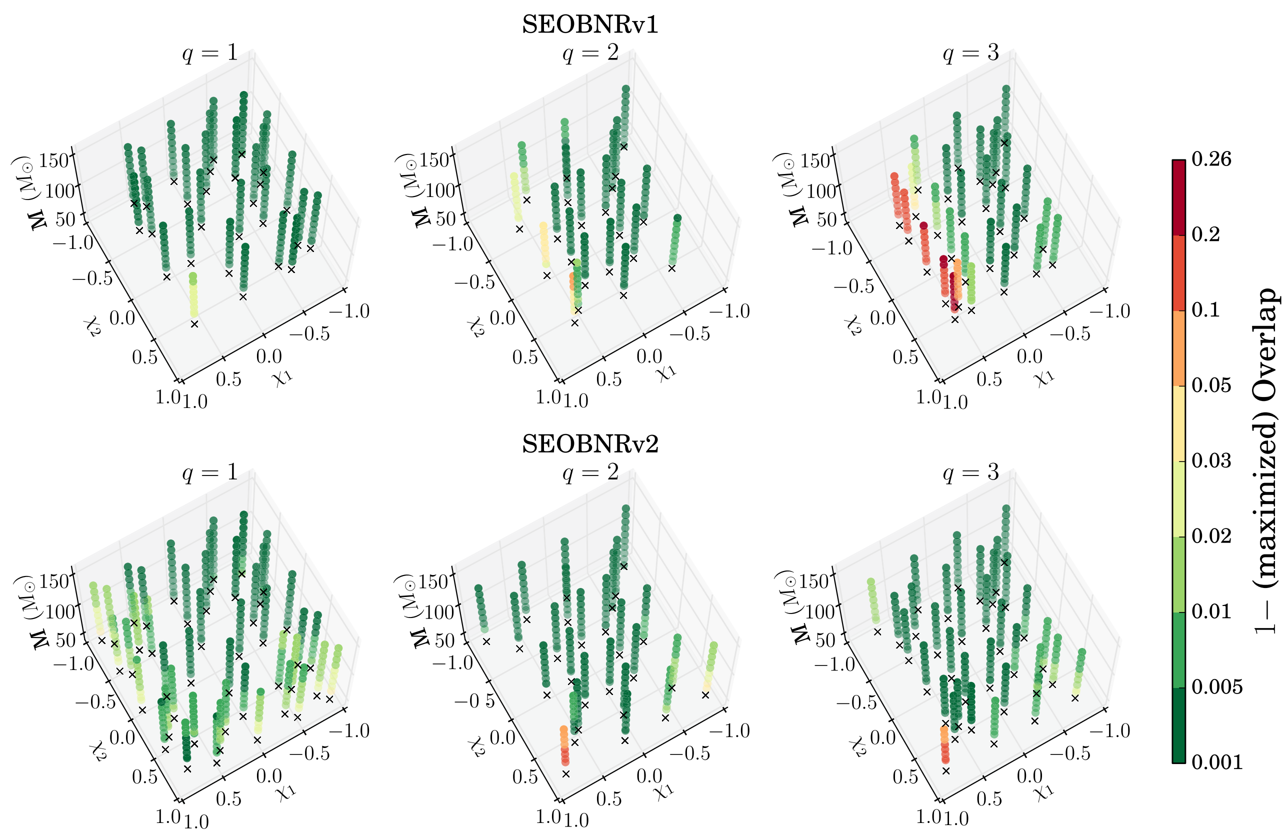

FIG. 3. Unfaithfulness between SEOBNR and NR waveforms as a function of mass ratio $q = m_1/m_2$, component spins χ_1, χ_2, and total mass M. SEOBNRv1 (top panel) reproduces NR well when the spin on the bigger BH does not exceed $+0.5$, with inaccuracies increasing with mass ratio. SEOBNRv2 (bottom panel) significantly improves over SEOBNRv1 with overlaps against NR higher than 98% over most of the parameter space considered. However, when spins on both component BHs are large and positive aligned, SEOBNRv2 fails to produce accurate waveforms ($\mathcal{O} \simeq 0.80$). We note that both models are accurate within their respective calibration range, but become inaccurate outside this range. Therefore it is crucial to test waveform models before using them in aLIGO analyses.

These overlaps are maximized over the extrinsic parameters however, i.e. over the time and phase at coalescence. They measure the accuracy of the models at specific points in the parameter space $(m_1, m_2, \chi_1, \chi_2)$.

In Fig. 3 we show the unfaithfulness (i.e. $1 - \mathcal{O}$) of the two EOB models, SEOBNRv1 and SEOBNRv2. In each row, the three panels correspond to mass ratios $q = \{1, 2, 3\}$. In each panel, the three axes correspond to component spins and total mass with the color showing the unfaithfulness. Note that the total masses probed are restricted to $M \gtrsim 50 M_\odot$ (cf. Fig. 1).

For SEOBNRv1, we find that its unfaithfulness increases with binary mass ratio as well as with the more massive component's spin, with little dependence on the binary's total mass. From the top left panel in Fig. 3 we note that for the smallest mass ratio, $q = 1$, SEOBNRv1 reproduces the NR waveforms well with unfaithfulness below 0.5% over most of the spin parameter space, except when the spins on both holes are close to the maximum value that the model supports (i.e. +0.6), where its unfaithfulness rises above 2%. As we increase the mass ratio to $q = 2$ (top middle panel of the same figure) SEOBNRv1's faithfulness further drops below 95% in the high-aligned-spin region. Furthermore, we also find that the unfaithfulness of the model reaches 1–2% when the smaller hole carries large antialigned spin. Further increasing the mass ratio to $q = 3$ increases the differences of the model with NR further, with overlaps falling below 90% when the larger black hole's spin $\to +0.6$. Overall, we find that the model performs better when the more massive hole has antialigned spins rather than aligned.

Turning to the more recent SEOBNRv2 model, we find that it significantly improves over SEOBNRv1: for equal-mass binaries, we find from the bottom left panel of Fig. 3 that the unfaithfulness of SEOBNRv2 is generally better than 1% except for mixed aligned/antialigned spin directions of large spin-magnitudes, where its unfaithfulness reaches 3%. For higher mass ratios $q = \{2, 3\}$, the slight increase of unfaithfulness towards the aligned/antialigned spin corner persists. For instance, $1 - \mathcal{O} \simeq 0.97$ for $q = 2$, $\chi_1 = -0.85, \chi_2 = +0.85$. However, the most significant deviation between SEOBNRv2 and NR occurs for both spins aligned with large magnitudes. For $\chi_1 = \chi_2 = +0.85$, the unfaithfulness rises above 10% for mass ratios $q = \{2, 3\}$. We explore these differences between SEOBNRv2 and NR further. In Fig. 4, we compare the model and NR waveforms for $q = \{2, 3\}, \chi_1 = \chi_2 = +0.85$. In both panels, the waveform pairs are aligned near the start of the NR waveform. We find that the SEOBNRv2 phase evolution agrees with NR during most of the inspiral phase, but its frequency rises faster during the plunge phase than that found with NR, resulting in an artificially accelerated merger. This evidence hints that the calibration of the merger portion and ringdown attachment of SEOBNRv2 will need further tuning.

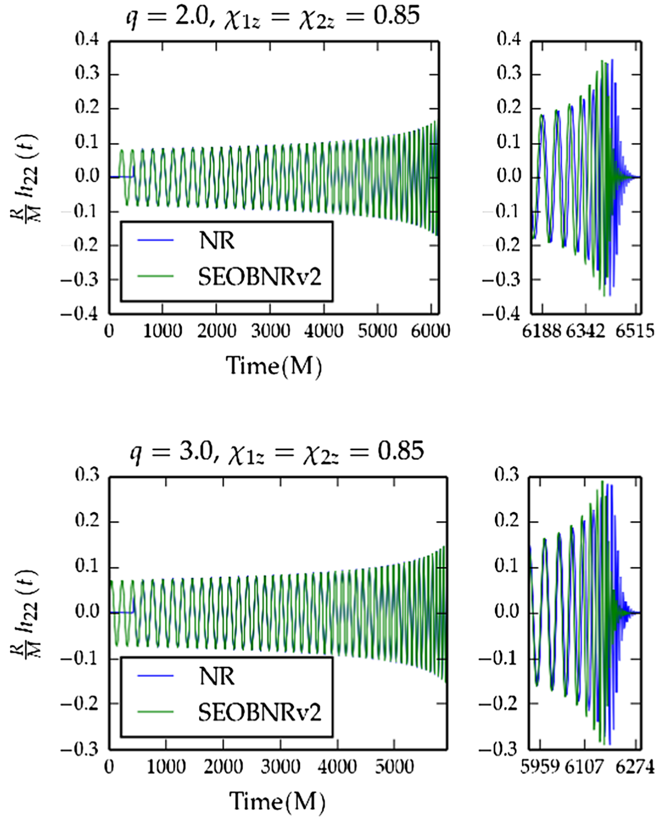

FIG. 4. SEOBNRv2 and NR waveforms for the problematic cases identified in the high spin corner of Fig. 3. Top: $q = 2, \chi_1 = \chi_2 = +0.85$. Bottom: $q = 3, \chi_1 = \chi_2 = +0.85$. Waveforms are aligned during their first few inspiral cycles.

We now turn our attention to the phenomenological models IMRPhenomC/D. The unfaithfulness of IMRPhenomC and IMRPhenomD with respect to NR, shown in Fig. 5, displays patterns distinct from the SEOBNR models. We find that IMRPhenomC shows poorer agreement with NR than either of the SEOBNR models, with unfaithfulness *increasing* rapidly with *mass ratio*, *spin magnitudes*, and with *decreasing binary masses*. The top panels of Fig. 5 show that this disagreement rises to 10–15% unfaithfulness, especially as the spin magnitude of the smaller BH grows. We notice disagreement between PhenomC and NR for large antialigned spins, which increases to 10–15% unfaithfulness over most of the spin parameter space as we go from $q = 1$ to $q = \{2, 3\}$. This disagreement increases, also, as more of the NR waveform is integrated over, i.e. at lower masses. In stark contrast, the newest model considered, IMRPhenomD, shows better agreement with NR than either of the SEOBNR models, with faithfulness above 99% over most of the analyzed parameter space, as seen in the bottom panels of Fig. 5. The only region where we see somewhat smaller overlaps is for $q \neq 1$ mixed-aligned spins with large positive spin on the larger hole.

We conclude that both SEOBNRv2 and IMRPhenomD models describe well binaries with *low to moderate spins*,

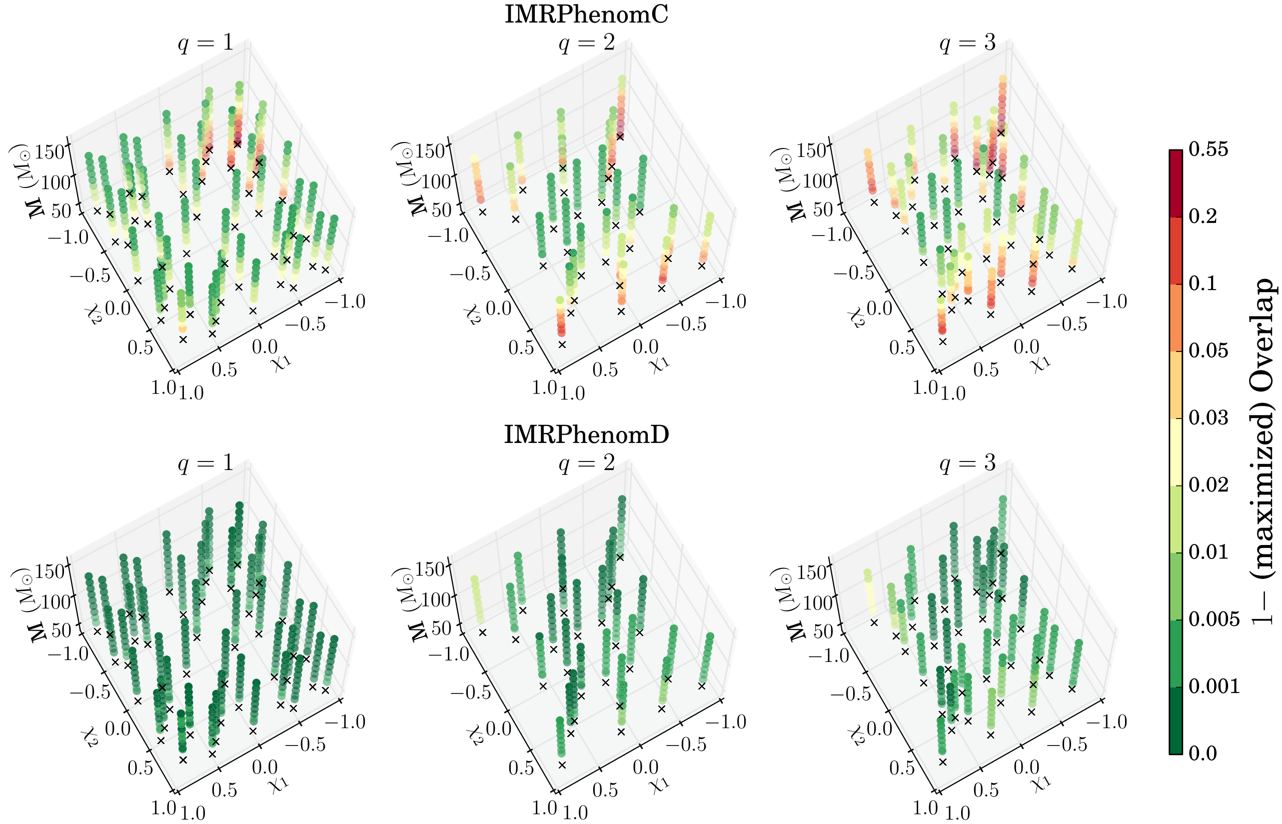

FIG. 5. This figure is similar to Fig. 3 with the difference that the models considered here are IMRPhenomC and IMRPhenomD (top and bottom panels, respectively). We note that both of the phenomenological models have been calibrated over most of the mass-ratio and spin range probed here. While IMRPhenomC shows significant deviation from NR as soon as we increase the mass ratio above $q = 1$, and/or spin magnitudes above ≈0.5, we find that IMRPhenomD reproduces NR remarkably well with overlaps above 99% everywhere (above 99.5% over most of the space).

and even *high antialigned spins*, with the latter also representing well *high-aligned-spins* binaries. The accuracy of both degrades somewhat with increasing mass ratio in the high-aligned/aligned spin and high-aligned/antialigned spin corners of the parameter space respectively. We also find that both of these models outperform their earlier counterparts significantly.

Further, we ask the question: how loud does a GW signal have to be for modeling errors to degrade scientific conclusions derived from it? To answer that, we use the sufficient criterion $(\delta h|\delta h) < 1$, where $\delta h = h^{\mathrm{true}} - h^{\mathrm{modeled}}$, to calculate the SNR threshold ρ_{eff} below which the true and modeled waveforms will not be distinguishable by aLIGO [104], i.e.

$$\rho_{\mathrm{eff}} = \frac{1}{\sqrt{2(1 - \mathcal{O}(h^{\mathrm{NR}}, h^{\mathrm{modeled}}))}}. \tag{6}$$

ρ_{eff} is the threshold value of the GW SNR, such that for $\rho \le \rho_{\mathrm{eff}}$ the statistical errors in mass and spin estimation will dominate over any systematic biases due to model inaccuracies, and therefore our scientific conclusions will not be degraded by model choice. The condition $\rho \le \rho_{\mathrm{eff}}$ is necessary, but not sufficient; i.e. it is not necessarily true that for all $\rho \ge \rho_{\mathrm{eff}}$ modeling inaccuracies will actually dominate [104]. With this caveat, we show in Fig. 6 the SNR threshold ρ_{eff} for the SEOBNRv2 and IMRPhenomD models, as a function of binary mass ratio, total mass, and component spins. From the top row of the figure, we find that SEOBNRv2 is sufficiently accurate for all aLIGO measurement purposes when concerned with moderately spinning binaries at SNRs up to ≈15–20. However, (i) for equal-mass binaries with *large mixed-aligned* spins, and (ii) for unequal-mass binaries with *large aligned* spins, using SEOBNRv2 waveforms may lead to a loss in information at fairly low aLIGO SNRs.

Turning to IMRPhenomD (lower panels of Fig. 6), we observe that this model is particularly accurate for equal-mass and/or equal-spin binaries and will be indistinguishable from NR for SNRs up to ≈30, or possibly even higher.[1] The SNR threshold falls to ≈15 for unequal-mass unequal-spin systems. Overall, we find that IMRPhenomD is best suited for aLIGO parameter estimation efforts aimed at

[1]The agreement between IMRPhenomD and the NR waveforms is so good that NR error estimates are of comparable order.

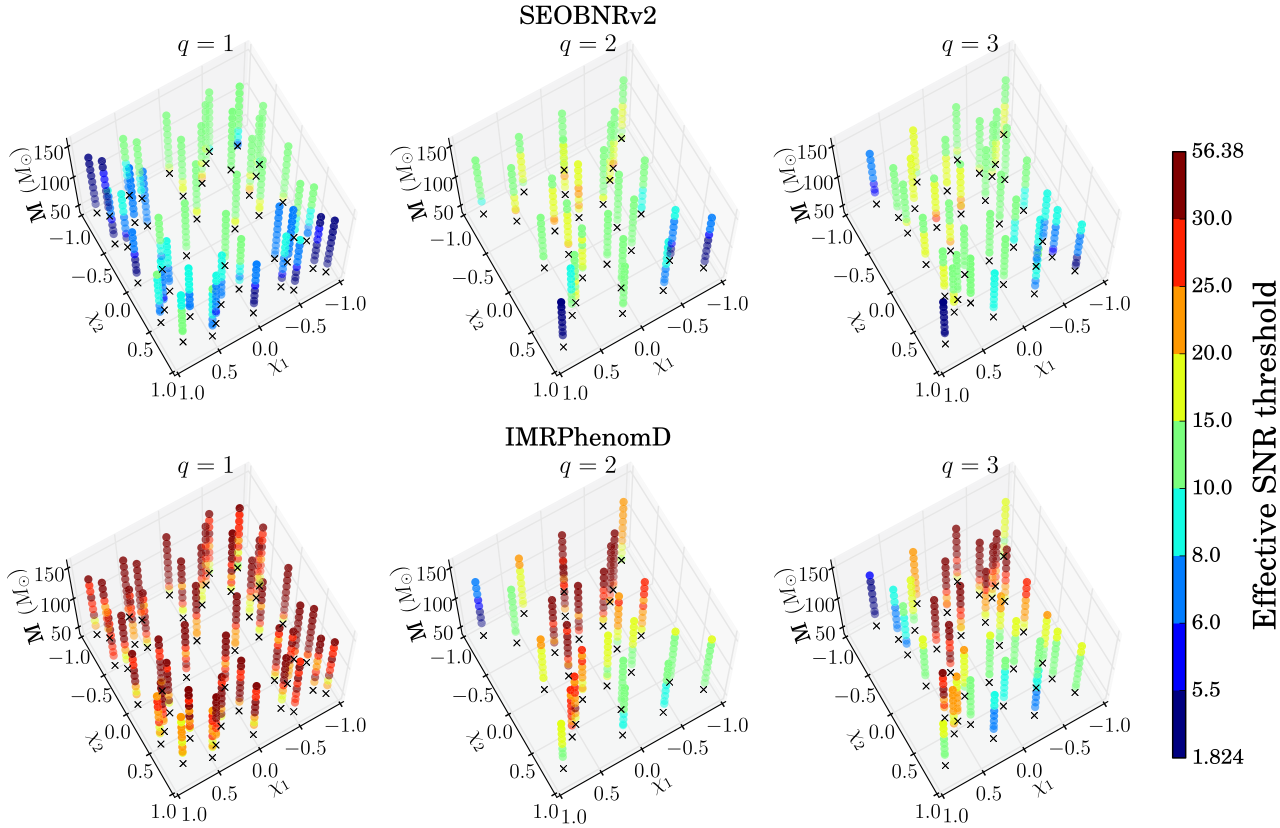

FIG. 6. We show the effective SNR level at which the SEOBNRv2 and IMRPhenomD models become distinguishable from NR waveforms with the Advanced LIGO instruments. Here we use the indistinguishability criterion proposed in Ref. [104].

comparable mass-ratio aligned-spin binaries of high total mass ($M \geq M_{\min} \gtrsim 50M_\odot$).

IV. EFFECTUALNESS

Matched-filtering-based GW searches use modeled waveforms as waveforms to filter detector data and recover signals that are otherwise buried in instrument noise. In such a search, the recovered SNR for a given signal is optimized over a discrete grid of binary mass and spin parameters that describe the waveforms, and is the highest when the filter waveform matches the signal exactly. In any real search, some fraction of the optimal SNR is lost due to two reasons: (i) the discreteness of the set of filter waveforms, and (ii) inaccuracies in the modeled waveforms. In this section we investigate the second factor for different waveform models from the perspective of aLIGO detection searches, focusing on nonprecessing BBHs. We use an overdense sampling of the waveform parameter space to mitigate any SNR losses due to reason (i). For each analytical waveform family, we compute overlaps between waveforms at all of the sampled points and with each of our NR waveforms. For each NR waveform, the highest overlap yields the fraction of optimal SNR recoverable by each waveform model.

This calculation involves a maximization over physical parameters of the model waveforms, and is therefore computationally more expensive than the faithfulness comparisons of Sec. III. The results of this effectualness study are summarized in Fig. 7. This figure shows the ineffectualness $\mathcal{M} := 1 - \mathrm{FF}$ [cf. Eq. (3)] of all IMR models considered here. From top to bottom, different rows correspond to SEOBNRv1, SEOBNRv2, IMRPhenomC and IMRPhenomD, respectively. In each row, different panels correspond to different mass ratios, and each panel spans the 3D subspace of binary total mass + component spins. From the top row, we immediately notice that even though SEOBNRv1 has support only for binaries with $\chi_{1,2} \leq +0.6$, it recovers $\geq 99.5\%$ of the optimal SNR for most of the parameter space where *either* $\chi_1 \geq +0.6$ *or* $\chi_2 \geq +0.6$. However, when both spins are large and aligned, its SNR recovery deteriorates to 93–95%. From the second row we notice that SEOBNRv2 performs significantly better with $\mathcal{E}^{\mathrm{SEOBNRv2}} \geq 99.5\%$ over most of the parameter space for all mass ratios considered. The recovered SNR by SEOBNRv2 drops, however, when both holes have large aligned spins. For $\chi_1 = \chi_2 = +0.85$, only 97% of optimal SNR are recovered, with worse performance at higher mass ratios. From the third row, we observe that IMRPhenomC achieves better than 98% SNR recovery over the parameter space considered. When the magnitude of the spins on *both* BHs is large and they are

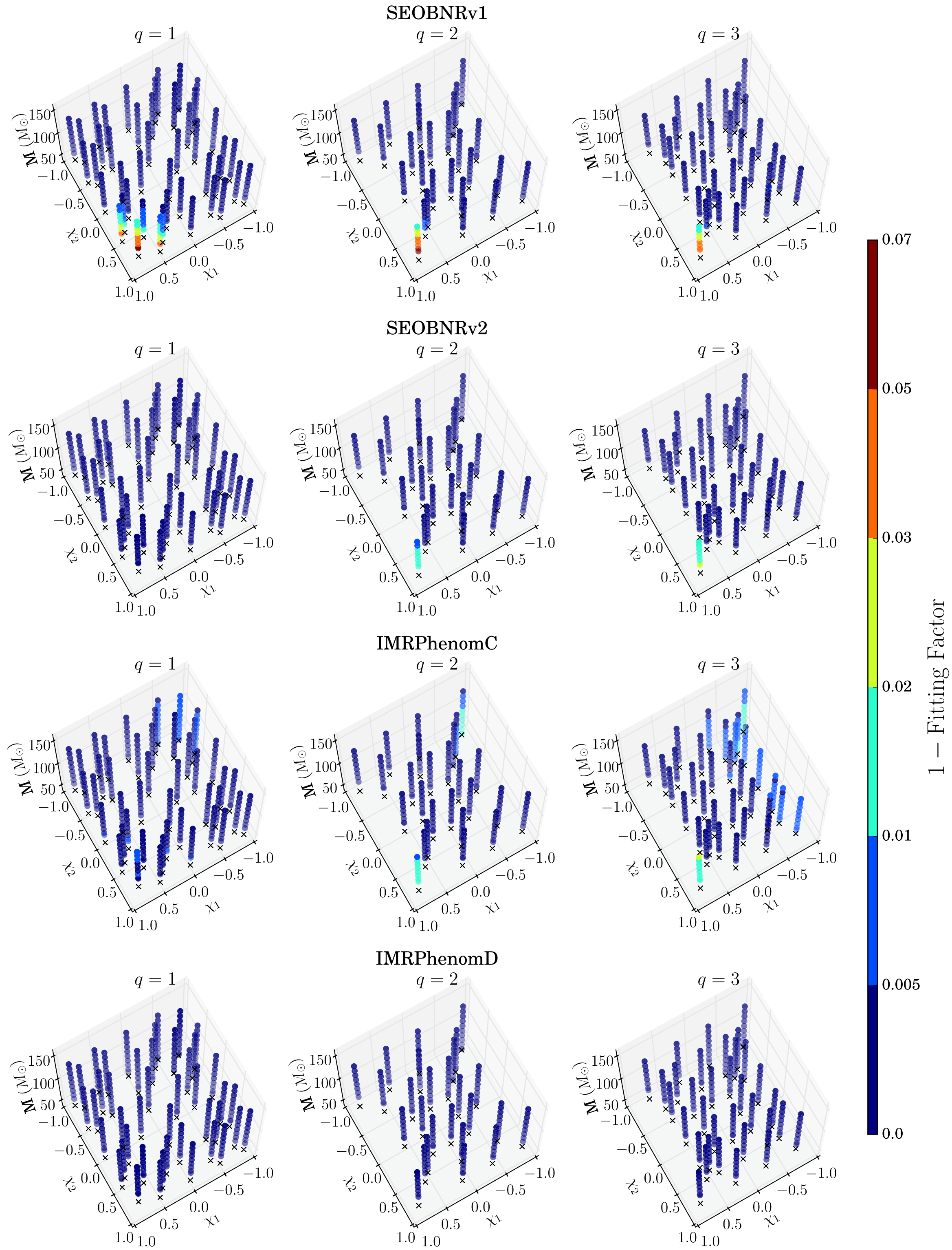

FIG. 7. Effectualness of the four waveform models considered. Plotted is the fractional loss in recovered SNR. Rows correspond to different models, and within each row, the data are plotted as a function of mass ratio q, BH spins χ_1, χ_2, and total mass M. The black crosses denote the values of component spins in the $x - y$ plane. We note that SEOBNRv1 does not model binaries with component spins higher than $+0.6$. We find that the more recent SEOBNRv2 and IMRPhenomD models supersede their earlier counterparts, SEOBNRv1 and IMRPhenomC, respectively, with FFs over 99.5% over most of the spin and mass parameter space probed. However, we do find that for binaries with high spins on both BHs, IMRPhenomD clearly outperforms all others with FFs > 99.5%, while SEOBNRv2's FFs against NR deteriorate to 97%.

parallel (i.e. either both spins aligned or both antialigned), the SNR loss increases 2% with increasing mass ratio. By comparing with the top row of Fig. 5 we see a clear demonstration of how well IMRPhenomC exploits the degeneracies of the binary parameter space through its use of an effective spin parameter. These results are consistent with the understanding that it was constructed with the aim of being an *effectual* model, and calibrated in the region of the parameter space which we probe here [47]. The bottom row of Fig. 7, finally, shows results for IMRPhenomD. As expected from its faithfulness measurements stated in the previous section, this model recovers $\geq 99.5\%$ of the optimal SNR in all of the parameter space which we probe here. Note that this includes all high-spin/high-spin corners, which were problematic with the other IMR waveform models.

To summarize, we find here that IMRPhenomD is the most effectual for BH binaries with $1 \leq q \leq 3$, $-0.85 \leq \chi_{1,2} \leq +0.85$ and total masses greater than those shown in Fig. 1. SEOBNRv2 also shows a comparable fitting factor, except for a slight drop in SNR recovery in the high-spin/high-spin corner of the nonprecessing BBH space.

V. SYSTEMATIC PARAMETER BIASES

Bayesian parameter estimation of BH masses and spins uses (semi)analytical waveform models. Its efficacy, therefore, depends critically on the accuracy of the waveform model used [105]. Modeling inaccuracies introduce systematic biases in the inferred parameter values. In this section, we quantify these systematic parameter biases for the four waveform models considered in this work. To avoid the complete MCMC procedure, we shall approximate the parameter bias of a waveform model as the difference in parameters between those parameters that maximize overlap with a NR waveform, and the parameters of the NR waveform. We repeat this calculation for every NR waveform. The broad features of the resulting parameter bias data are dependent most strongly on the effective spin parameter $\chi_{\rm eff}$, and therefore, we will present results as a function of it. Because we project two spins χ_1, χ_2 onto the one effective spin, the plotted data will not be single valued. Configurations with different χ_1, χ_2, but the same $\chi_{\rm eff}$ yield in general different biases, which are plotted at their $\chi_{\rm eff}$ values.

First, in the left column of Fig. 8 we show the fractional systematic bias in the recovery of binary chirp mass $\mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}}$ that is intrinsic to different waveform models, as a function of the effective spin $\chi_{\rm eff}$ of the NR waveforms. In the right column of the same figure we show the fractional biases in the recovery of binary total mass M. In the top row, we show results for SEOBNRv1. The magnitude of the systematic biases for this model increases rapidly with (i) increasing magnitude of $\chi_{\rm eff}$, and (ii) increasing mass ratio. For example, we see that the recovered chirp mass can be biased by up to 15% when the effective spin is antialigned, while the total mass bias does not exceed 5%. On the other hand, the increasing trend of systematic biases at high $\chi_{\rm eff}$ is to be expected since SEOBNRv1 does not support spins $\chi_{1,2} \geq +0.6$ [46]. In the third row, we show the intrinsic bias of IMRPhenomC in recovering binary's chirp and total masses. Focusing at the plot markers in both panels, we observe that the systematic biases stay below $\sim 3\%$ for binaries with masses at the lower end of the mass range probed here. However at higher masses, as with SEOBNRv1, both the recovered chirp mass and total mass can be shifted by 15% if the binary's $\chi_{\rm eff} \leq 0$. Relatively, the total mass is recovered better by this model. In comparison with SEOBNRv1, IMRPhenomC allows for less accurate parameter recovery. Next, we consider the more recent SEOBNRv2 model (second row). This waveform model is of interest, in part, because its reduced-order model [51] is being used in BBH searches being run for the presently ongoing aLIGO observing run O1. Focusing on the plot markers we find that the systematic biases in $\mathcal{M}_c$ recovery stay below $\sim$1–2% of the true $\mathcal{M}_c$ value, for binaries with masses $\lesssim 80 M_\odot$. For higher masses (100–150$M_\odot$), biases go up to 5%, but are still smaller than the statistical uncertainty in $\mathcal{M}_c$ measurement at high masses [106,107]. In comparison, SEOBNRv2 recovers binary total mass *less* accurately with systematically larger fractional biases than for $\mathcal{M}_c$. We also observe that the bias in M has the same sign as the $\chi_{\rm eff}$ of the binary. Finally, in the bottom right panel, we show the results for the most recently published IMRPhenomD model. Performing better than SEOBNRv2, IMRPhenomD furnishes biases in the recovery of $\mathcal{M}_c$ which rarely exceed 2%. For $\chi_{\rm eff} \in [-0.6, +0.6]$ the total mass recovery does rise to 5%, which is worse than the model's $\mathcal{M}_c$ bias for the same signals. We also highlight the aligned-spin/aligned-spin corner, where SEOBNRv2's mass-recovery biases rise up to 5–10%, while they stay within 2–5% for IMRPhenomD. This is to be expected given the disagreement between SEOBNRv2 and NR in the same region of parameter space, as shown in Sec. III. For all models, as illustrated in Fig. 11, we note that the highest parameter biases for chirp mass correspond to the upper edge of the total mass range probed here, i.e. the edge of the "error bars" corresponds to $M \sim 150 M_\odot$. In summary, for $M \leq 100 M_\odot$, both SEOBNRv2 and IMRPhenomD are likely to yield similarly accurate estimates of chirp mass, while for higher masses we find IMRPhenomD to be relatively more suited to parameter estimation studies.

Further, Fig. 9 shows the recovered value of binary mass ratio $q = m_1/m_2$, for different waveform models, as a function of the mass ratio and effective spin $\chi_{\rm eff}$ of the NR waveforms. As before, the plot markers correspond to a fixed total mass $M = 80 M_\odot$, while the error bars show the entire range of y-values for the mass range that we probe here (i.e. $M \in [M_{\rm min}, 150 M_\odot]$). In the spin range supported

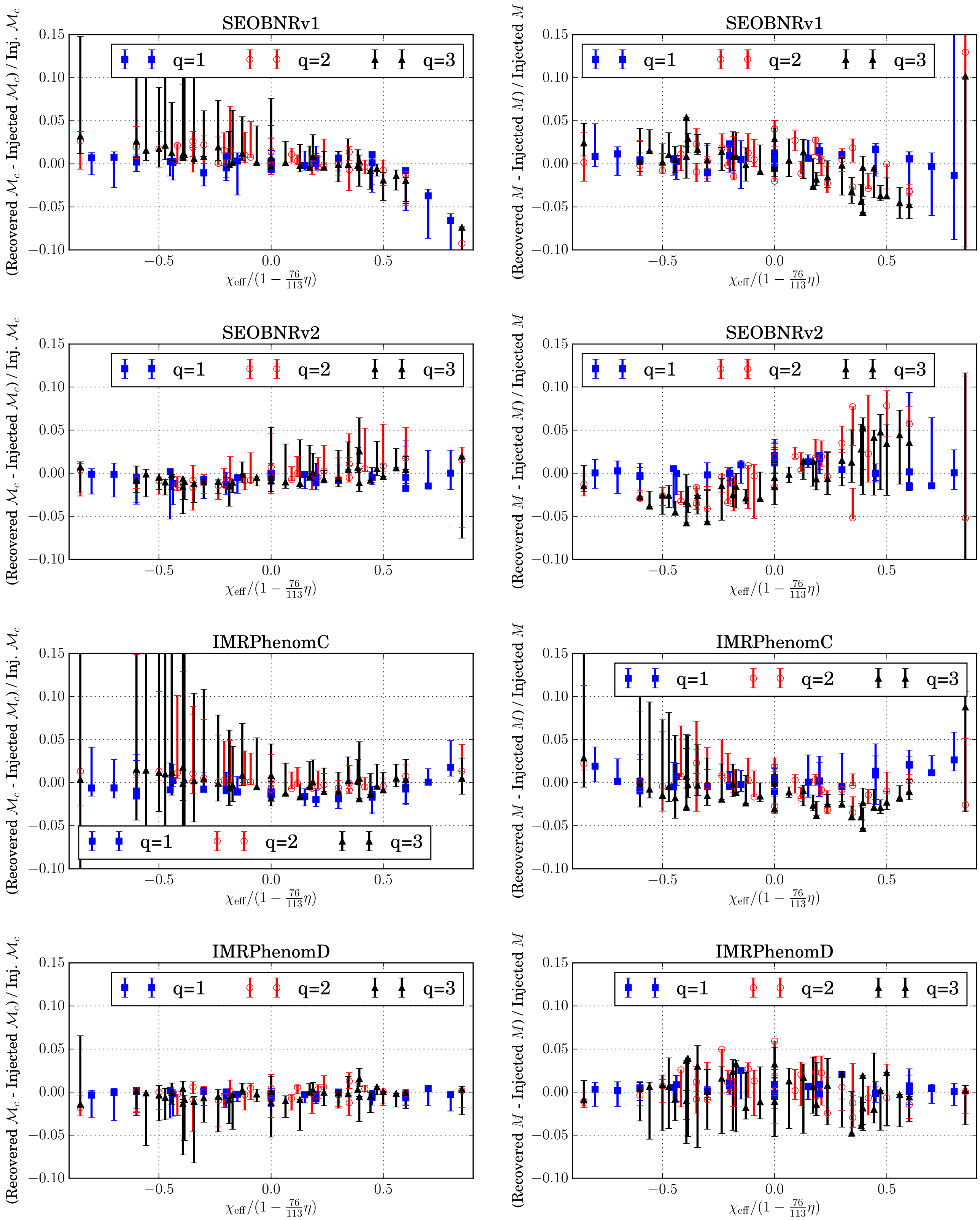

FIG. 8. Systematic bias in the recovery of chirp mass $\mathcal{M}_c$ (left column) and total mass M (right column) for different waveform models (rows). In each panel, the respective bias is shown as a function of the normalized effective spin of the NR waveforms. The plot markers show the bias for a binary with total mass fixed at $M = 80M_\odot$. The "error bars" show the range of biases for total masses between the minimum allowed mass and $150M_\odot$.

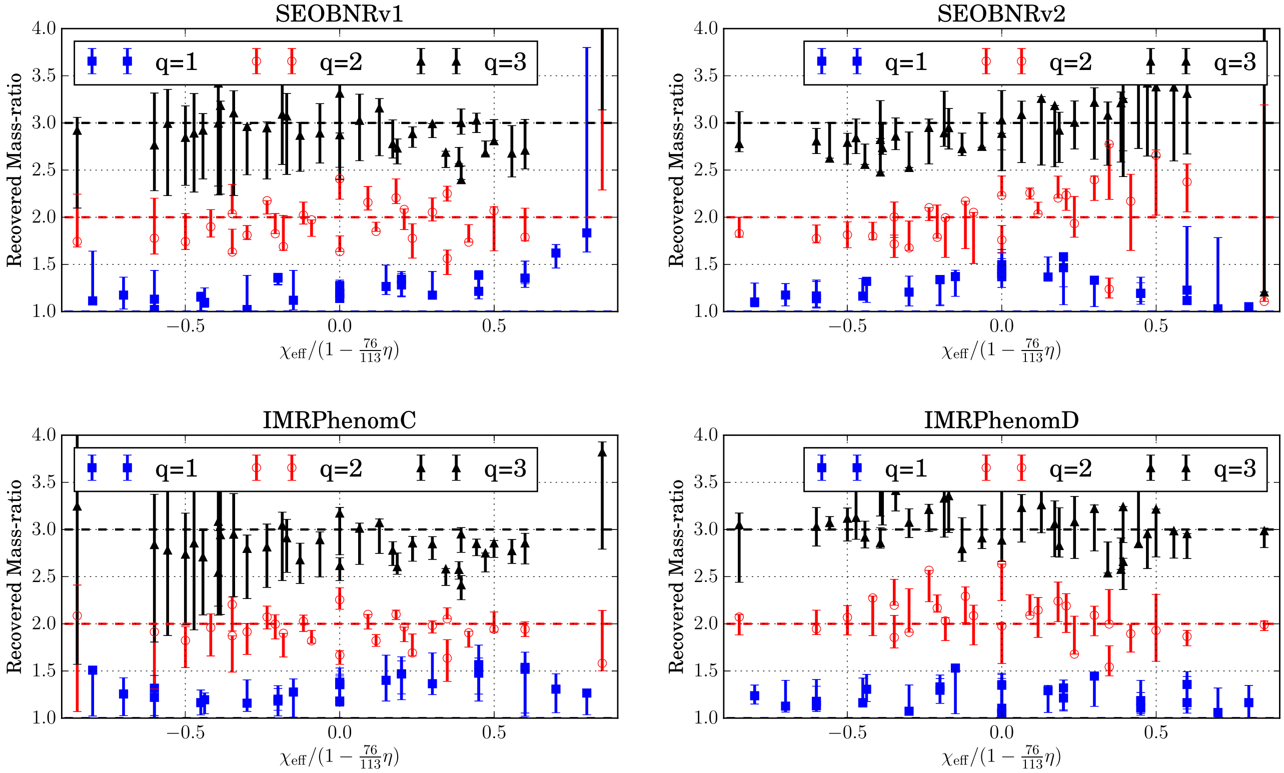

FIG. 9. Systematic bias in the recovery of the binary mass ratio $q \coloneqq m_1/m_2$, as a function of the normalized effective spin of the NR waveforms. Different mass ratios are shown with different color, with horizontal dashed lines of the same color drawn to guide the eye. The plot markers show the recovered q for a binary with total mass fixed at $80M_{\odot}$, while the "error bars" show the range spanned by the recovered q as the injected binary mass is varied between its lowest allowed value and $150M_{\odot}$.

by SEOBNRv1, we find that it exhibits up to 15% systematic bias in the recovery of q, with biases increasing as $|\chi_{\rm eff}| \to 1$, i.e. for highly spinning binaries, including at the lower end of the mass range probed here. SEOBNRv2, on the other hand, shows a systematic trend with $\chi_{\rm eff}$. We find that the difference between the recovered and true mass ratios increases with $\chi_{\rm eff}$. For negative $\chi_{\rm eff}$, q tends to be underestimated, whereas for positive $\chi_{\rm eff}$, q tends to be overestimated. Therefore, if the radiating source has $\chi_{\rm eff} < 0$ then SEOBNRv2 will give an artificially *lower* mass-ratio value as the maximum-likelihood parameter estimate, and vice versa if $\chi_{\rm eff} > 0$. At the high-aligned-spin end, the mass ratio can be overestimated by more than 15% by SEOBNRv2. Turning to IMRPhenomC, we find that its associated q bias stays within 20% at the lower mass end, and is much larger for high binary masses. This is particularly true for large antialigned $\chi_{\rm eff}$. IMRPhenomD, on the other hand, shows little dependency of its intrinsic mass-ratio bias on effective spin, except that it gives slightly elevated q-bias close to $\chi_{\rm eff} = 0$, i.e. for mixed-aligned binaries. Overall, we note that all models recover q worse as the total mass of the system increases. IMRPhenomD gives a relatively better estimate of the mass ratio than the other models considered here.

Finally, in Fig. 10 we show the bias in the recovery of the effective-spin combination, $\chi_{\rm eff}$, as a function of the $\chi_{\rm eff}$ of the NR waveforms. $\chi_{\rm eff}$ is the leading-order spin combination that enters the binary's inspiral phasing, and therefore the matched filter is expected to be most sensitive to this combination of the component spins [102]. Overall, we find that $\chi_{\rm eff}$ is well constrained, within ± 0.2 of its true value, by all the waveform models considered. From the left column, we can compare the spin recovery of the two older models, SEOBNRv1 and IMRPhenomC. Both of these models exhibit strong dependence of the accuracy of spin recovery on $\chi_{\rm eff}$. For SEOBNRv1, we find that its associated $\chi_{\rm eff}$ bias is constrained within ± 0.1 of the true value, when the source binary's $\chi_{\rm eff} \leq +0.4$. When the binary's $\chi_{\rm eff}$ exceeds $+0.4$, the model gives rapidly increasing systematic biases in its spin recovery, with $\chi_{\rm eff}$ being underestimated by up to 0.25. This trend arises because SEOBNRv1 is restricted to component spins $\chi_{1,2} \leq +0.6$, so higher NR spin must—by construction—be recovered by $\chi_{1,2}$ within SEOBNRv1's range. IMRPhenomC exhibits a similar trend at the negative side of the spin range: it recovers $\chi_{\rm eff}$ within ± 0.1 when the source's $\chi_{\rm eff} > -0.5$, with the bias increasing sharply for more antialigned spins. In the top right panel of Fig. 10, we show the spin recovery by the SEOBNRv2 model. Primarily, we note that SEOBNRv2 recovers $\chi_{\rm eff}$ very well, with a systematic bias that stays below ± 0.1 in dimensionless spin magnitude (with rare excursions up to

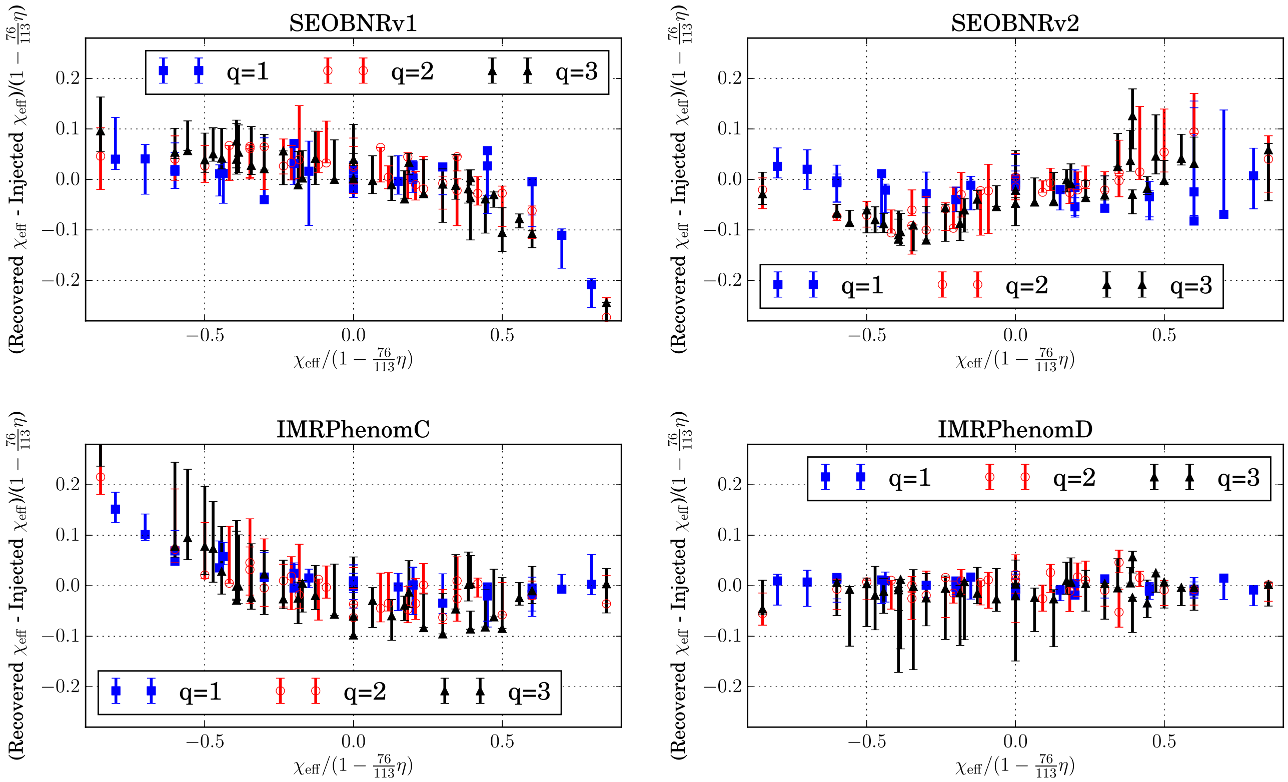

FIG. 10. Systematic bias in the recovery of the effective spin parameter $\chi_{\rm eff}$, as a function of the normalized effective spin of the NR waveforms. The plot markers show the recovered $\chi_{\rm eff}$ for a binary with total mass fixed at $80M_\odot$, while the "error bars" show the range spanned by the recovered q as the injected binary mass is varied between its lowest allowed value and $150M_\odot$.

± 0.2 for large aligned spins). In addition, we note a (minor, but interesting) pattern: the bias in spin recovery increases almost linearly with $\chi_{\rm eff}$ between $-0.5 \le \chi_{\rm eff} \le +0.6$, going from -0.1 for $\chi_{\rm eff} = -0.5$ to $+0.1$ for $\chi_{\rm eff} = +0.6$. Finally, the bottom right panel of Fig. 10 shows the $\chi_{\rm eff}$ bias for IMRPhenomD. We find the systematic bias in $\chi_{\rm eff}$ associated with this waveform model stays between ± 0.1 (as for SEOBNRv2), with best recovery for aligned spins and low total masses. We also note that this bias shows little dependence on $\chi_{\rm eff}$ itself; however it does increase systematically with mass ratio q for the higher binary masses. Overall, we find that of all the waveform models considered, both SEOBNRv2 and IMRPhenomD recover $\chi_{\rm eff}$ within ± 0.1, with IMRPhenomD performing markedly more consistently for binaries with large aligned spins.

For all models, we note that the highest parameter biases for mass and spin parameters correspond to the upper edge of the total-mass range probed; i.e. the edge of the error bars corresponds to $M \sim 150M_\odot$. We present detailed results showing the dependence of systematic parameter biases on signal parameters in Appendixes A and B.

From the results presented in this and the previous section, we find that both IMRPhenomD and SEOBNRv2 outperform their earlier incarnations in the recovery of various mass and spin combinations probed here, with IMRPhenomD performing systematically better (i) at recovering binary's chirp mass, and (ii) for parameter recovery, in general, for systems with high-aligned spins.

A more detailed MCMC analysis is necessary to measure the statistical uncertainties in parameter recovery from different models in order to determine the GW SNRs at which modeling inaccuracies will *actually* begin to dominate. Figure 6 only gives a lower limit on this SNR, and we may well find that statistical uncertainties remain dominant for even louder signals. We do, however, recommend based on this study that aLIGO parameter estimation efforts use either of the two waveform models to model filters.

VI. CONCLUSIONS

LIGO and other ground-based gravitational-wave detectors rely on waveform models for detection of compact object binaries as well as for parameter estimation of the candidate events. Accurate aveform models are therefore necessary to ensure high detection efficiency and to avoid systematic biases in parameter estimation.

Past studies focused on evaluating the accuracy of waveform models have either used model precision as a proxy for accuracy (i.e. used model/model discrepancy as a proxy for model/true-signal discrepancy) [26–31], or have used NR simulations with zero/low-to-moderate component spins as benchmarks [32–41,108,109]. In this paper we investigate the accuracy of four inspiral-merger-ringdown waveform models for binary black holes. Our analysis improved in several ways over earlier work: first, we compare with numerical relativity waveforms, rather than using the difference between analytical models as a proxy for their error [26–31]. Second, the NR waveforms are independent of the investigated waveform models, in the sense that *none* of them was used in calibrating these waveform models. Furthermore, a companion paper [45] establishes the accuracy of the NR waveforms. Third, we consider two recently published models, IMRPhenomD [43] and SEOBNRv2 [42], the accuracy of which has not been investigated independently (except for neutron-star black-hole binaries [41]). Finally, our set of reference waveforms comprehensively samples the component-spin parameter space up to $\chi_1, \chi_2 = 0.9$ for $q = m_1/m_2 = 1$ and 0.85 for $q = \{2, 3\}$, extending the spin coverage beyond the spins used in calibrating the waveform models.

First, we investigate the modeling accuracy of different waveform models by computing their overlaps against our NR reference waveforms. We rescale the NR waveforms to a range of total mass values, from the lowest permissible (and still ensuring that it starts at 15 Hz; see Fig. 1) up to $m_1 + m_2 = 150M_\odot$. From Fig. 3, we find that (i) SEOBNRv1 has overlaps above 99% against NR waveforms for binaries where the more massive black hole has spin $\chi_1 < 0.5$, which drop to 80% for larger χ_1, and (ii) SEOBNRv2 performs better with overlaps above 98% across the parameter space except when both $\chi_{1,2}$ are large and aligned. From Fig. 5, we find (iii) IMRPhenomC is faithful *only* to NR for very mildly spinning binaries, with overlaps falling below 90% when $|\chi_2| \geq +0.3$, and (iv) IMRPhenomD is superior to other waveform models with overlaps (against our reference NR waveforms) above 99% over the entire spin and mass parameter space considered. For the two most faithful models (SEOBNRv2 and IMRPhenomD), we evaluate the indistinguishability criterion, to find the SNR below which modeling errors do *not* significantly bias parameter estimation. From Fig. 6, we find that, except for binaries with large aligned spins on *at least one BH*, SEOBNRv2 remains indistinguishable from real GW signals with SNRs up to 15 or higher. IMRPhenomD will be indistinguishable from real GW signals with SNRs of 30 and above for equal-mass, equal-spin binaries, and for SNRs $\gtrsim 15$ over most of the remaining parameter space. These SNR ranges are very likely to be conservative, due to the overly strict nature of the distinguishability criterion used [104].

Second, we investigate the effectualness of different waveform models (including two additional PN-based ones) for use as aLIGO BBH detection filters. Detection searches have an additional degree of freedom: the recovered SNR is maximized over the mass and spin parameters that characterize model waveforms. We compute the fitting factors [110] of different waveform models against our NR waveforms, to measure the SNR loss due to modeling inaccuracies in isolation. As shown in Fig. 7, we find that (i) SEOBNRv1 is effectual over the entire parameter range it supports, i.e. for $\chi_{1,2} \leq +0.6$, with fitting factors higher than 99.5%; (ii) SEOBNRv2 has fitting factors above 99.5% across the considered region of the parameter space, except for the high-spin/high-spin corner, where its fitting factors fall to 97%; (iii) IMRPhenomC recovers more than 99% of the SNR over most of the parameter space, except when both holes have either large aligned or large anti-aligned spins, in which cases it still recovers more than 98% of the optimal SNR; and (iv) IMRPhenomD outperforms all other waveform models with fitting factors above 99.5% over the *entire* parameter range probed. We note that the frequency domain IMRPhenomC model makes good use of the intrinsic degeneracy in the waveform parameter space, and is therefore well suited to detection searches. SEOBNRv2, on the other hand, does not compensate for its inaccuracy in the high-spin/high-spin corner of the parameter space with modified intrinsic parameters, and will likely need to be recalibrated there.

Third, we investigate the systematic biases in parameter recovery caused by intrinsic model inaccuracies. We find that (i) both IMRPhenomD and SEOBNRv2 recover binary *chirp mass* to within ±2% for $M \gtrsim 70M_\odot$, and ±5–7% for $M \gtrsim 110M_\odot$, with IMRPhenomD systematically more accurate for aligned spins. (ii) Binary total mass is recovered with somewhat larger systematic biases across the mass range, spanning ±5% for binaries for which the chirp mass is recovered within ±2%. (iii) SEOBNRv2 and IMRPhenomD recover the binary *mass ratio* with comparable accuracy (within ±10–15%), with IMRPhenomD showing the smallest biases for aligned spin binaries. Finally, (iv) the leading-order PN spin combination $\chi_{\rm eff}$ is the best recovered with IMRPhenomD (within ±0.1), followed closely by SEOBNRv2. The remaining two models show larger biases for all intrinsic parameters (see Fig. 14).

In summary, we find that the more recently published SEOBNRv2 and IMRPhenomD models reproduce NR waveforms with identical parameters more accurately than their earlier counterparts, and have very good SNR recovery. For both models this is explained by improvements in their spin parametrizations, and extensive merger-ringdown calibration against new NR simulations (few tens in number, and mass ratios up to $q = 8$–18). This manifest systematic improvement of models over time highlights the tactical synergy between numerical relativity and waveform modeling. Further on, we also find that the frequency-domain IMRPhenomC model is effectual enough for aLIGO detection searches aimed at comparable-mass aligned-spin high-mass BBHs, making good use of intrinsic degeneracies of the waveform parameter space. Overall, we recommend that aLIGO parameter estimation efforts prefer IMRPhenomD or SEOBNRv2 as the waveform model of choice, in favor of other currently available frequency and time domain waveform models.

As noted previously, the parameter biases estimated here need to be comprehensively compared with the statistical errors in parameter recovery from detailed MCMC analyses, in order to determine the actual GW SNRs where modeling errors begin to dominate over other error sources of uncertainty. A recent study [50] indicates that such might be the case for SNRs ≈20–30 and higher (for SEOBNRv2). We also note that in order to thoroughly sample the spin parameter space, we have restricted ourselves to small mass ratios, i.e. $q = \{1, 2, 3\}$. The results presented here are therefore applicable to comparable-mass BBHs with total masses $M \gtrsim 50 M_\odot$, and will be extended to higher mass ratios and lower total masses in the future, as longer and higher q simulations become less computationally expensive with advances in NR technology [111]. Finally, we use the dominant quadrupolar multipoles here of the reference NR waveforms, and leave a study of the subdominant modes for future work. We expect their effect to be limited to the highest masses and mass ratios considered here [112], although a more rigorous treatment is needed to reaffirm this conclusion.

ACKNOWLEDGMENTS

We thank Kipp Cannon, Adam Lewis, Eric Poisson and Aaron Zimmerman for helpful discussions. We are grateful to Ofek Birnholtz, Sebastian Khan, Lionel London, Frank Ohme and Michael Pürrer, for providing access to the IMRPhenomD code. Simulations used in this work were performed with SpEC [44]. We gratefully acknowledge support for this research at CITA from NSERC of Canada, the Ontario Early Researcher Awards Program, the Canada Research Chairs Program, and the Canadian Institute for Advanced Research; at Caltech from the Sherman Fairchild Foundation and NSF Grants No. PHY-1404569 and No. AST-1333520; at Cornell from the Sherman Fairchild Foundation and NSF Grants No. PHY-1306125 and No. AST-1333129; and at Princeton from NSF Grant No. PHY-1305682 and the Simons Foundation. Calculations were performed at the GPC supercomputer at the SciNet HPC Consortium [113]; SciNet is funded by the Canada Foundation for Innovation (CFI) under the auspices of Compute Canada; the Government of Ontario; Ontario Research Fund (ORF)—Research Excellence; and the University of Toronto. Further calculations were performed on the Briarée cluster at Sherbrooke University, managed by Calcul Québec and Compute Canada and with operation funded by the Canada Foundation for Innovation (CFI), Ministére de l'Économie, de l'Innovation et des Exportations du Quebec (MEIE), RMGA and the Fonds de recherche du Québec—Nature et Technologies (FRQ-NT); on the Zwicky cluster at Caltech, which is supported by the Sherman Fairchild Foundation and by NSF Grant No. PHY-0960291; on the NSF XSEDE network under Grant No. TG-PHY990007N; on the NSF/NCSA Blue Waters at the University of Illinois with allocation jr6 under NSF PRAC Grant No. ACI-1440083. H. P. and P. K. thank the Albert-Einstein Institute, Potsdam, for hospitality during part of the time where this research was completed.

APPENDIX A: BIAS IN MASS COMBINATIONS

In this appendix, we present additional information about the parameter estimation mass recovery. Figure 11 shows the chirp-mass recovery as a function of both component spins, expanding on the left column of Fig. 8 in the main text. Figure 12 shows the total-mass recovery, similarly expanding on the right column of Fig. 8. Figure 13 plots the recovery of symmetric mass ratio η (cf. Fig. 9).

From Fig. 11, we find that (i) for the *least* massive binaries considered, both SEOBNRv2 and IMRPhenomD introduce less than 2% systematic biases in the recovery of binary chirp mass; with the same rising to 10% for the *most* massive binaries. (ii) The chirp-mass bias measured for SEOBNRv1 closely follows that of SEOBNRv2, except when both black holes carry large spins (both aligned and antialigned), where the bias exceeds 10%. (iii) IMRPhenomC, on the other hand, has intrinsic chirp-mass biases that remain below 5% over the considered parameter space, except when the more massive hole has large antialigned spin—for which the biases exceed 10% for binary mass $M \gtrsim 100 M_\odot$. From Fig. 12, we find that (i) both SEOBNRv1 and IMRPhenomC incur smaller systematic biases in M recovery than they do for $\mathcal{M}_c$ recovery, especially for large antialigned and aligned spins. (ii) SEOBNRv2 shows the opposite pattern; i.e. it recovers M with *more* accuracy than $\mathcal{M}_c$, especially for larger mass ratios and larger spins on the bigger black hole. Finally, (iii) IMRPhenomD recovers both mass combinations with the relatively highest accuracy.

Further onto η recovery, the first thing we note from Fig. 13 is that all four models recover η well (within 2%) for equal-mass binaries, and this fidelity decreases as we go from $q = 1 \to 3$. The only exception is SEOBNRv1 at spins outside the range of the model (i.e. $\chi_{1,2} > +0.6$). For $q = 2$, we find that (i) the biases intrinsic to SEOBNRv2 are higher than SEOBNRv1, reaching 15–20% and 10–15%, respectively, for both. SEOBNRv2 also gives a systematic underestimation of η by -15% when both holes have large positive-aligned spins. (ii) IMRPhenomC, in contrast, performs better with biases staying below 10%, even at the highest binary masses. And, (iii) IMRPhenomD shows the highest fidelity (with η biases below 5%). Increasing the mass ratio to $q = 3$, we find that (i) all three models other than IMRPhenomD manifest larger than 10% systematic biases in η recovery. (ii) For SEOBNRv1 the η bias increases as the spin on the smaller hole becomes increasingly antialigned, while SEOBNRv2 overestimates

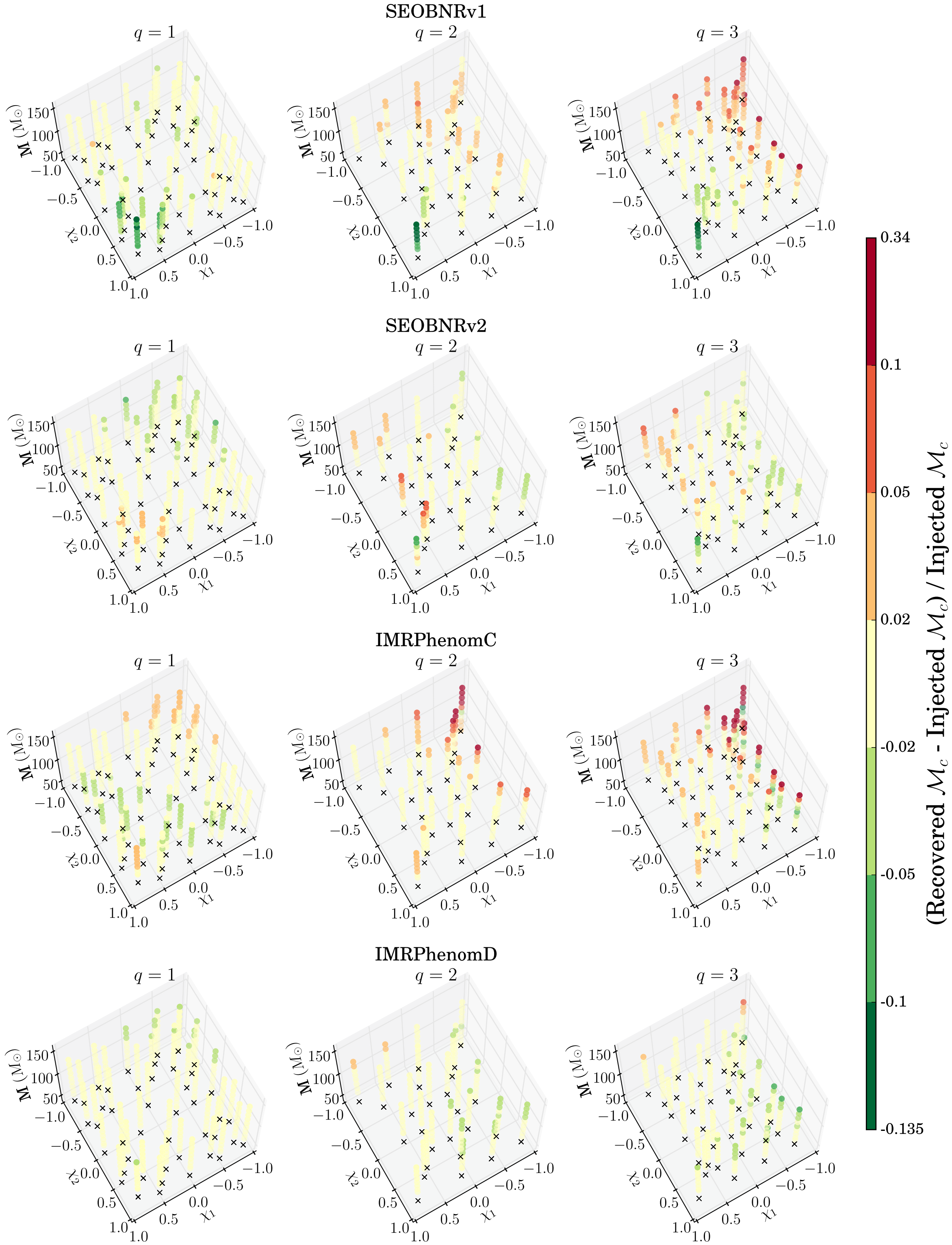

FIG. 11. Systematic bias in the recovered chirp mass $\mathcal{M}_c$ for each waveform model; compare to Fig. 8. As in Fig. 7, the black crosses denote the values of component spins in the $x - y$ plane. Biases below 2% are shown nearly transparently, to emphasize regions with larger biases.

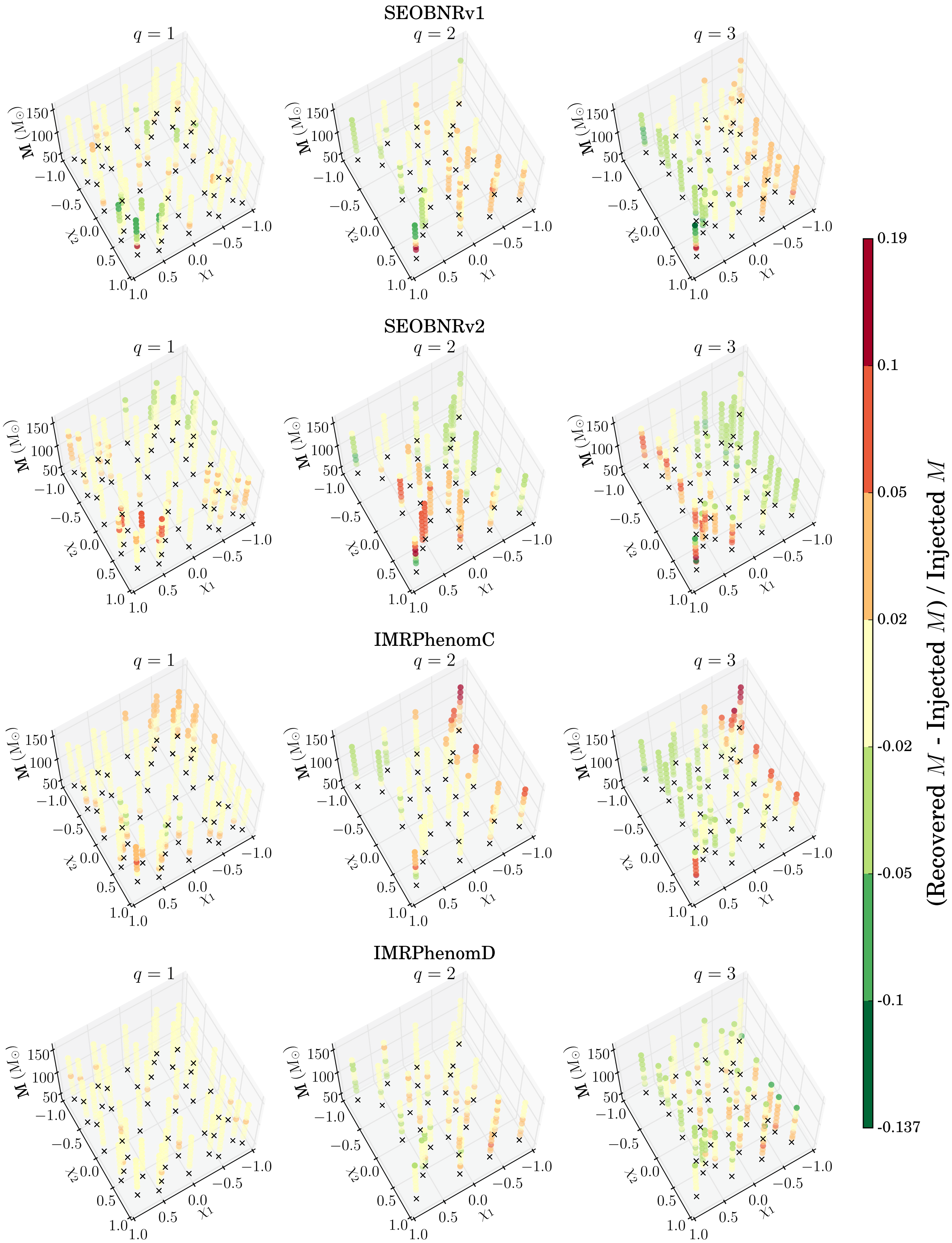

FIG. 12. Systematic bias in the recovered total mass M for each waveform model; compare to Fig. 8. As in Fig. 7, the black crosses denote the values of component spins in the $x - y$ plane. Biases below 2% are shown nearly transparently, to emphasize regions with larger biases.

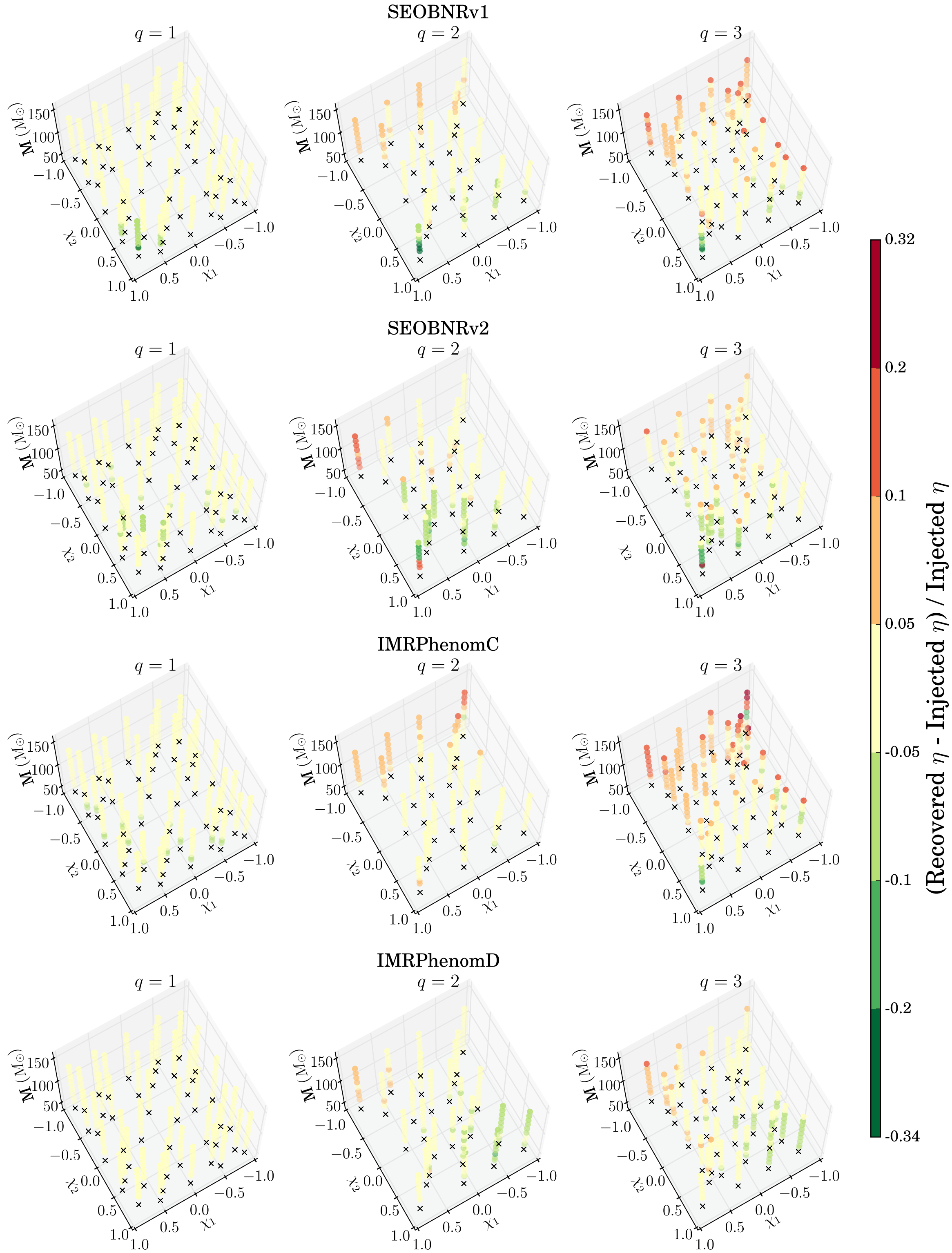

FIG. 13. Systematic bias in the recovered symmetric mass ratio η for the considered waveform models. As in Fig. 7, the black crosses denote the values of component spins in the $x-y$ plane.

η by 5–10% for aligned BH spins (this trend was already apparent in Fig. 9). (iii) IMRPhenomC shows the relatively worst η recovery of the four with biases ranging from −15% to 20%. IMRPhenomD confirms our earlier results and is found to perform best at η recovery, significantly improving upon its predecessor IMRPhenomC.

APPENDIX B: BIASES IN RECOVERED SPINS

In this appendix, we present additional information about the parameter estimation spin recovery. Figures 14 and 15 show the bias in the recovery of two different spin parameters, the effective spin $\chi_{\rm eff}$ [cf. Eq. (5)] and the mass-weighted spin $\chi_{\rm mw}$ [cf. Eq. (4)].

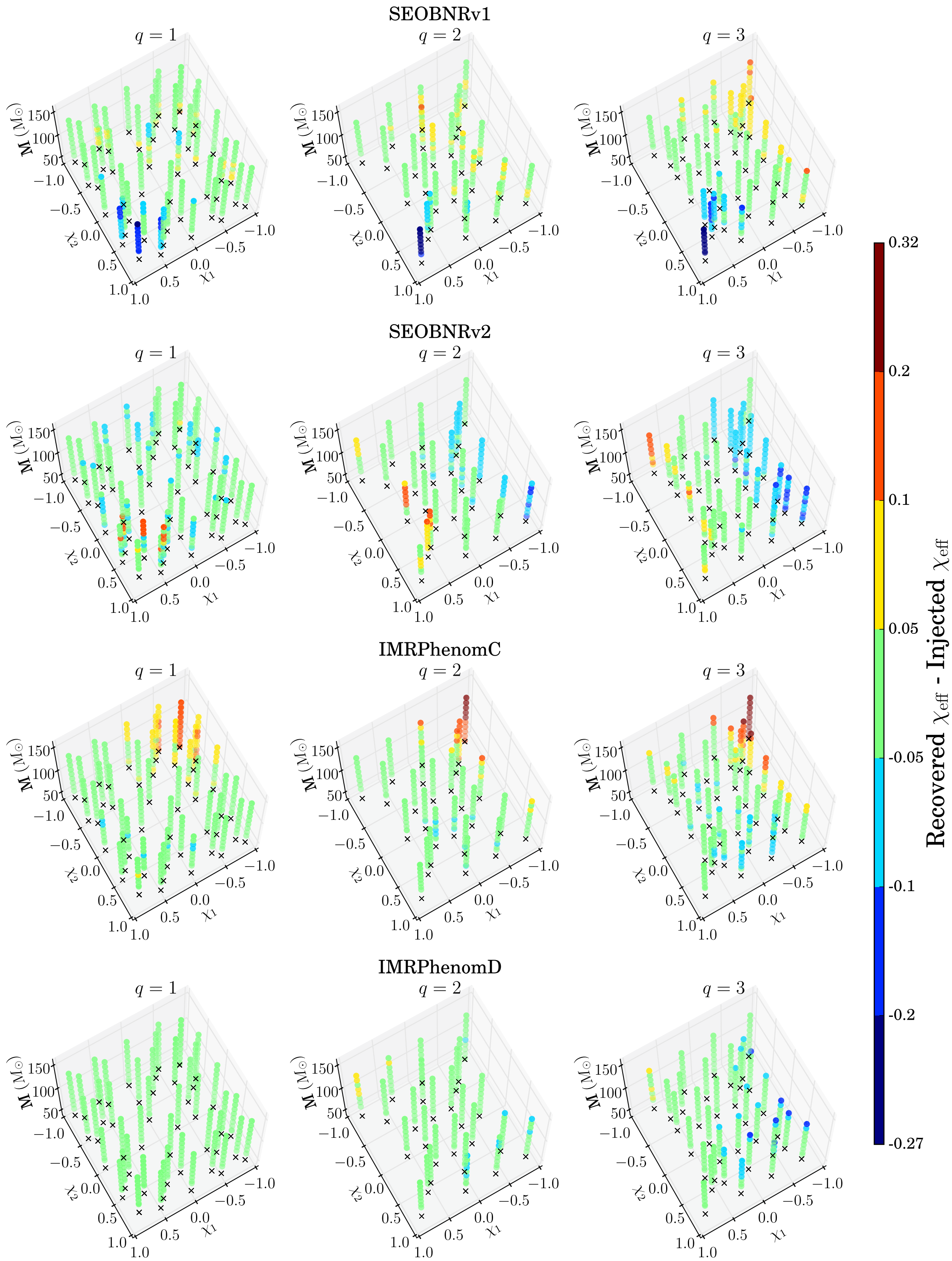

FIG. 14. Systematic bias in the recovered values of the 1.5PN effective spin $\chi_{\rm eff}$, for the SEOBNRv1, SEOBNRv2, IMRPhenomC, and IMRPhenomD models (from top to bottom).

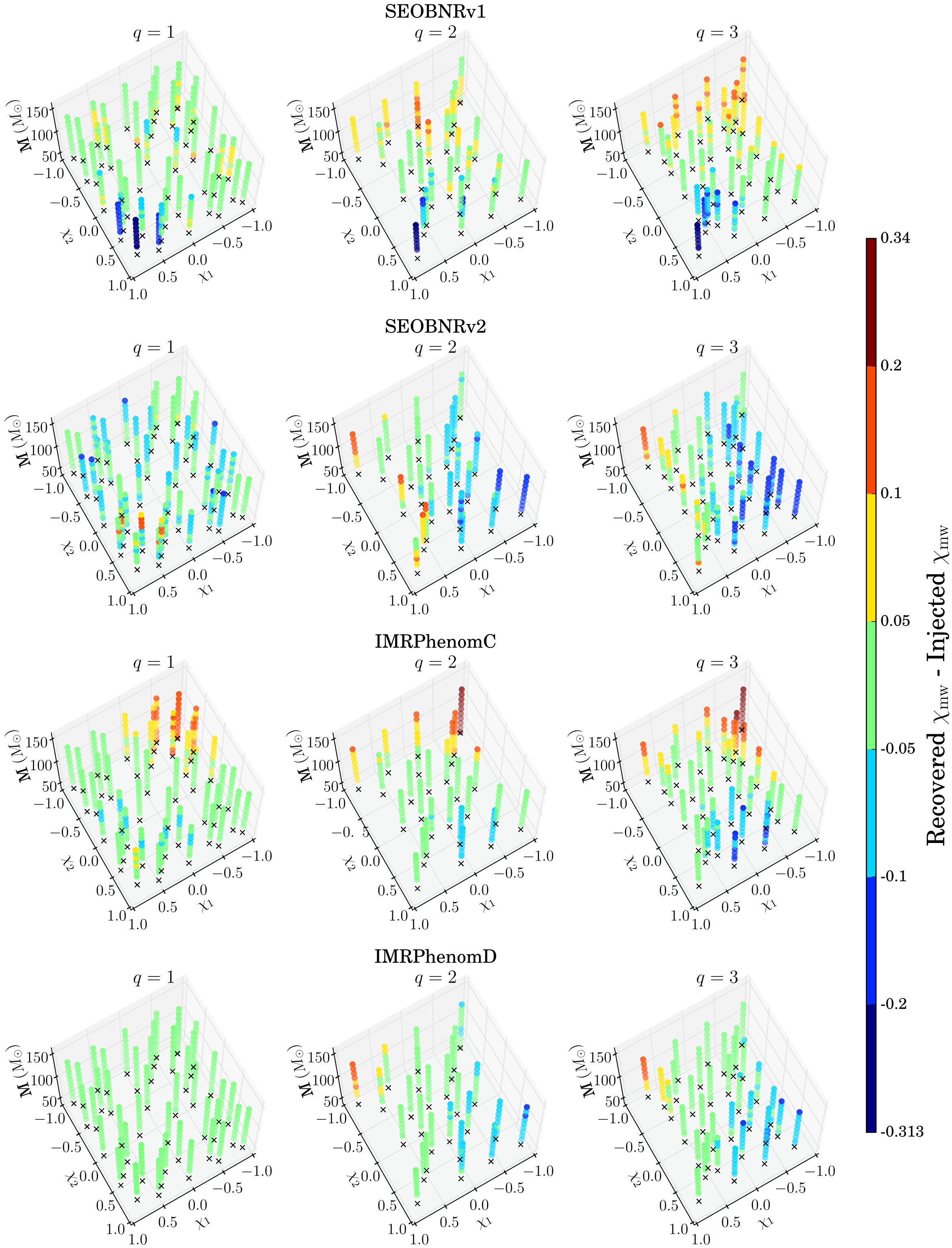

FIG. 15. Systematic bias in the recovered values of the mass-weighted effective spin $\chi_{\rm mw}$, for the SEOBNRv1, SEOBNRv2, IMRPhenomC, and IMRPhenomD models (from top to bottom).

The overall trends are similar for $\chi_{\rm eff}$ and $\chi_{\rm mw}$: all four models recover $\chi_{\rm eff}$ well, with absolute systematic biases between ± 0.2. Of the four, IMRPhenomD stands out by recovering $\chi_{\rm eff}$ to within ± 0.1 of the true value. SEOBNRv2 follows very closely, with $\chi_{\rm eff}$ biases rising higher than 0.1 only for very massive binaries (with $M \gtrsim 100 M_\odot$) with large spins (magnitude) on at least one hole. Both of the two remaining models show a strong correlation between the $\chi_{\rm eff}$ bias and the $\chi_{\rm eff}$ of the binary itself. While SEOBNRv1 underestimates $\chi_{\rm eff}$ by up to 0.25 when both holes have large *aligned* spins, IMRPhenomC overestimates $\chi_{\rm eff}$ when both holes have large *antialigned* spins. Next, we focus on $\chi_{\rm mw}$. As for $\chi_{\rm eff}$, IMRPhenomD was found to recover $\chi_{\rm mw}$ with the smallest biases, which only exceed ± 0.1 for unequal-mass binaries with aligned (antialigned) spin on the larger (smaller) hole. As this is the spin mapping used by IMRPhenomC to capture component spin effects on phasing, we notice from Fig. 14 that it also recovers $\chi_{\rm mw}$ very well—except when both components have large antialigned spins, in which case it overestimates $\chi_{\rm mw}$ by up to 0.3 dimensionless units. Of the two EOB models, SEOBNRv2 recovers $\chi_{\rm mw}$ better with systematic biases increasing with mass ratio q, but not exceeding ± 0.2. SEOBNRv1, on the other hand, shows the inverse pattern of IMRPhenomC, giving large systematic biases in $\chi_{\rm mw}$ for binaries with $\chi_{1,2} \geq 0.6$, which is expected by construction from the model as it does not support these component spins.

Overall, we find both IMRPhenomD and SEOBNRv2 models viable for aLIGO parameter estimation studies aimed at high-mass binary black holes with nonprecessing spins.

[1] G. M. Harry (LIGO Scientific Collaboration), Classical Quantum Gravity **27**, 084006 (2010).
[2] J. Aasi *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **32**, 074001 (2015).
[3] J. Aasi *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Living Rev. Relativity **19**, 1 (2016).
[4] J. Aasi *et al.* (VIRGO, LIGO Scientific Collaboration), Classical Quantum Gravity **32**, 115012 (2015).
[5] F. Acernese *et al.* (VIRGO Collaboration), Classical Quantum Gravity **32**, 024001 (2015).
[6] K. Somiya (KAGRA Collaboration), Classical Quantum Gravity **29**, 124007 (2012).
[7] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto (KAGRA Collaboration), Phys. Rev. D **88**, 043007 (2013).
[8] C. S. Unnikrishnan, Int. J. Mod. Phys. D **22**, 1341010 (2013).
[9] J. Abadie *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **27**, 173001 (2010).
[10] J. M. Silverman and A. V. Filippenko, Astrophys. J. Lett. **678**, L17 (2008).
[11] K. Belczynski, T. Bulik, C. L. Fryer, A. Ruiter, F. Valsecchi, J. S. Vink, and J. R. Hurley, Astrophys. J. **714**, 1217 (2010).
[12] L. Gou, J. E. McClintock, M. J. Reid, J. A. Orosz, J. F. Steiner, R. Narayan, J. Xiang, R. A. Remillard, K. A. Arnaud, and S. W. Davis, Astrophys. J. **742**, 85 (2011).
[13] A. C. Fabian, D. R. Wilkins, J. M. Miller, R. C. Reis, C. S. Reynolds, E. M. Cackett, M. A. Nowak, G. G. Pooley, K. Pottschmidt, J. S. Sanders, R. R. Ross, and J. Wilms, Mon. Not. R. Astron. Soc. **424**, 217 (2012).
[14] L. Gou, J. E. McClintock, R. A. Remillard, J. F. Steiner, M. J. Reid, J. A. Orosz, R. Narayan, M. Hanke, and J. García, Astrophys. J. **790**, 29 (2014).
[15] J. M. Miller, C. S. Reynolds, A. C. Fabian, G. Miniutti, and L. C. Gallo, Astrophys. J. **697**, 900 (2009).
[16] J. E. McClintock, R. Shafee, R. Narayan, R. A. Remillard, S. W. Davis, and L.-X. Li, Astrophys. J. **652**, 518 (2006).
[17] J. Miller, C. Reynolds, A. Fabian, G. Miniutti, and L. Gallo, Astrophys. J. **697**, 900 (2009).
[18] J. E. McClintock, R. Narayan, and J. F. Steiner, Space Sci. Rev. **183**, 295 (2014).
[19] J. Abadie *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. D **85**, 082002 (2012).
[20] J. Abadie *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. D **82**, 102001 (2010).
[21] B. P. Abbott *et al.* (LIGO Scientific), Phys. Rev. D **80**, 047101 (2009).
[22] B. P. Abbott *et al.* (LIGO Scientific), Phys. Rev. D **79**, 122001 (2009).
[23] B. Abbott *et al.* (LIGO Scientific), Phys. Rev. D **78**, 042002 (2008).
[24] I. W. Harry, A. H. Nitz, D. A. Brown, A. P. Lundgren, E. Ochsner, and D. Keppel, Phys. Rev. D **89**, 024010 (2014).
[25] D. A. Brown, P. Kumar, and A. H. Nitz, Phys. Rev. D **87**, 082004 (2013).
[26] T. Damour, B. R. Iyer, and B. S. Sathyaprakash, Phys. Rev. D **63**, 044023 (2001).
[27] T. Damour, B. R. Iyer, and B. S. Sathyaprakash, Phys. Rev. D **63**, 044023 (2001); **72**, 029902(E) (2005).
[28] T. Damour, B. R. Iyer, and B. S. Sathyaprakash, Phys. Rev. D **66**, 027502 (2002).
[29] T. Damour, B. R. Iyer, and B. S. Sathyaprakash, Phys. Rev. D **66**, 027502 (2002); **72**, 029901(E) (2005).
[30] A. Gopakumar, M. Hannam, S. Husa, and B. Brügmann, Phys. Rev. D **78**, 064026 (2008).
[31] A. Buonanno, B. R. Iyer, E. Ochsner, Y. Pan, and B. S. Sathyaprakash, Phys. Rev. D **80**, 084043 (2009).
[32] M. Boyle, D. A. Brown, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, G. B. Cook, and S. A. Teukolsky, Phys. Rev. D **76**, 124038 (2007).
[33] M. Boyle, D. A. Brown, and L. Pekowsky, Classical Quantum Gravity **26**, 114006 (2009).
[34] M. Boyle, A. Buonanno, L. E. Kidder, A. H. Mroué, Y. Pan, H. P. Pfeiffer, and M. A. Scheel, Phys. Rev. D **78**, 104020 (2008).
[35] Y. Pan, A. Buonanno, J. G. Baker, J. Centrella, B. J. Kelly, S. T. McWilliams, F. Pretorius, and J. R. van Meter, Phys. Rev. D **77**, 024014 (2008).

[36] M. Hannam, S. Husa, B. Brügmann, and A. Gopakumar, Phys. Rev. D **78**, 104007 (2008).
[37] M. Hannam, S. Husa, F. Ohme, D. Muller, and B. Brugmann, Phys. Rev. D **82**, 124008 (2010).
[38] M. Hannam, S. Husa, F. Ohme, D. Müller, and B. Brügmann, Phys. Rev. D **82**, 124008 (2010).
[39] I. MacDonald, A. H. Mroué, H. P. Pfeiffer, M. Boyle, L. E. Kidder, M. A. Scheel, B. Szilágyi, and N. W. Taylor, Phys. Rev. D **87**, 024009 (2013).
[40] I. Hinder *et al.* (NRAR Collaboration), Classical Quantum Gravity **31**, 025012 (2014).
[41] P. Kumar, K. Barkett, S. Bhagwat, N. Afshari, D. A. Brown, G. Lovelace, M. A. Scheel, and B. Szilagyi, Phys. Rev. D **92**, 102001 (2015).
[42] A. Taracchini, A. Buonanno, Y. Pan, T. Hinderer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, A. H. Mroue, H. P. Pfeiffer, M. A. Scheel, B. Szilagyi, N. W. Taylor, and A. Zenginoglu, Phys. Rev. D **89**, 061502 (2014).
[43] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Phys. Rev. D **93**, 044007 (2016).
[44] http://www.black-holes.org/SpEC.html.
[45] T. Chu, H. Fong, P. Kumar, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, M. A. Scheel, and B. Szilagyi, arXiv:1512.06800.
[46] A. Taracchini, Y. Pan, A. Buonanno, E. Barausse, M. Boyle, T. Chu, G. Lovelace, H. P. Pfeiffer, and M. A. Scheel, Phys. Rev. D **86**, 024011 (2012).
[47] L. Santamaría, F. Ohme, P. Ajith, B. Brügmann, N. Dorband, M. Hannam, S. Husa, P. Mösta, D. Pollney, C. Reisswig, E. L. Robinson, J. Seiler, and B. Krishnan, Phys. Rev. D **82**, 064016 (2010).
[48] D. Shoemaker (LIGO Collaboration), Advanced LIGO anticipated sensitivity curves, LIGO Document T0900288-v3, 2010.
[49] E. Baird, S. Fairhurst, M. Hannam, and P. Murphy, Phys. Rev. D **87**, 024035 (2013).
[50] M. Pürrer, M. Hannam, and F. Ohme, arXiv:1512.04955.
[51] M. Pürrer, Phys. Rev. D **93**, 064041 (2016).
[52] J. S. Read, C. Markakis, M. Shibata, K. Uryū, J. D. E. Creighton, and J. L. Friedman, Phys. Rev. D **79**, 124033 (2009).
[53] L. Kreidberg, C. D. Bailyn, W. M. Farr, and V. Kalogera, Astrophys. J. **757**, 36 (2012).
[54] S. Fairhurst, New J. Phys. **11**, 123006 (2009).
[55] S. Nissanke, J. Sievers, N. Dalal, and D. Holz, Astrophys. J. **739**, 99 (2011).
[56] J. Abadie, B. P. Abbott, R. Abbott, T. D. Abbott, M. Abernathy, T. Accadia, F. Acernese, C. Adams *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Astron. Astrophys. **539**, A124 (2012).
[57] J. Abadie, B. P. Abbott, R. Abbott, T. D. Abbott, M. Abernathy, T. Accadia, F. Acernese, C. Adams, R. Adhikari, C. Affeldt *et al.*, Astron. Astrophys. **541**, A155 (2012).
[58] P. A. Evans, J. K. Fridriksson, N. Gehrels, J. Homan, J. P. Osborne, M. Siegel, A. Beardmore, P. Handbauer, J. Gelbord, J. A. Kennea *et al.*, Astrophys. J. Suppl. Ser. **203**, 28 (2012).
[59] S. Nissanke, M. Kasliwal, and A. Georgieva, Astrophys. J. **767**, 124 (2013).
[60] B. D. Metzger and E. Berger, Astrophys. J. **746**, 48 (2012).
[61] I. Mandel, L. Z. Kelley, and E. Ramirez-Ruiz, in *IAU Symposium*, Vol. 285, edited by E. Griffin, R. Hanisch, and R. Seaman (Cambridge University Press, Oxford, 2012), pp. 358–360.
[62] T. G. F. Li, W. Del Pozzo, S. Vitale, C. Van Den Broeck, M. Agathos, J. Veitch, K. Grover, T. Sidery, R. Sturani, and A. Vecchio, Phys. Rev. D **85**, 082003 (2012).
[63] T. G. F. Li, W. Del Pozzo, S Vitale, C. Van Den Broeck, M. Agathos, J. Veitch, K. Grover, T. Sidery, R. Sturani, and A. Vecchio, J. Phys. Conf. Ser. **363**, 012028 (2012).
[64] T. B. Littenberg, B. Farr, S. Coughlin, V. Kalogera, and D. E. Holz, Astrophys. J. **807**, L24 (2015).
[65] I. Mandel, C.-J. Haster, M. Dominik, and K. Belczynsk, Mon. Not. R. Astron. Soc. **450**, L85 (2015).
[66] H. P. Pfeiffer, L. E. Kidder, M. A. Scheel, and S. A. Teukolsky, Comput. Phys. Commun. **152**, 253 (2003).
[67] H.-J. Yo, J. N. Cook, S. L. Shapiro, and T. W. Baumgarte, Phys. Rev. D **70**, 084033(E) (2004); **70**, 089904(E) (2004).
[68] G. B. Cook and H. P. Pfeiffer, Phys. Rev. D **70**, 104016 (2004).
[69] H. Friedrich, Commun. Math. Phys. **100**, 525 (1985).
[70] D. Garfinkle, Phys. Rev. D **65**, 044029 (2002).
[71] F. Pretorius, Classical Quantum Gravity **22**, 425 (2005).
[72] L. Lindblom, M. A. Scheel, L. E. Kidder, R. Owen, and O. Rinne, Classical Quantum Gravity **23**, S447 (2006).
[73] B. Szilágyi, L. Lindblom, and M. A. Scheel, Phys. Rev. D **80**, 124010 (2009).
[74] B. Szilágyi, Int. J. Mod. Phys. D **23**, 1430014 (2014).
[75] M. A. Scheel, M. Boyle, T. Chu, L. E. Kidder, K. D. Matthews, and H. P. Pfeiffer, Phys. Rev. D **79**, 024003 (2009).
[76] D. A. Hemberger, M. A. Scheel, L. E. Kidder, B. Szilágyi, G. Lovelace, N. W. Taylor, and S. A. Teukolsky, Classical Quantum Gravity **30**, 115001 (2013).
[77] D. Gottlieb and J. S. Hesthaven, J. Comput. Appl. Math. **128**, 83 (2001).
[78] J. S. Hesthaven, Applied Numerical Mathematics **33**, 23 (2000).
[79] L. Lindblom, M. A. Scheel, L. E. Kidder, R. Owen, and O. Rinne, Classical Quantum Gravity **23**, S447 (2006).
[80] O. Rinne, Classical Quantum Gravity **23**, 6275 (2006).
[81] O. Rinne, L. Lindblom, and M. A. Scheel, Classical Quantum Gravity **24**, 4053 (2007).
[82] A. Buonanno and T. Damour, Phys. Rev. D **59**, 084006 (1999).
[83] T. Damour, B. R. Iyer, P. Jaranowski, and B. S. Sathyaprakash, Phys. Rev. D **67**, 064028 (2003).
[84] T. Damour and A. Nagar, Phys. Rev. D **76**, 064028 (2007).
[85] T. Damour and A. Nagar, Phys. Rev. D **77**, 024043 (2008).
[86] T. Damour, A. Nagar, E. N. Dorband, D. Pollney, and L. Rezzolla, Phys. Rev. D **77**, 084017 (2008).
[87] T. Damour and A. Nagar, Phys. Rev. D **76**, 044003 (2007).
[88] A. Buonanno, Y. Pan, J. G. Baker, J. Centrella, B. J. Kelly, S. T. McWilliams, and J. R. van Meter, Phys. Rev. D **76**, 104049 (2007).
[89] T. Damour, A. Nagar, M. Hannam, S. Husa, and B. Brugmann, Phys. Rev. D **78**, 044039 (2008).
[90] A. Buonanno, Y. Pan, H. P. Pfeiffer, M. A. Scheel, L. T. Buchman, and L. E. Kidder, Phys. Rev. D **79**, 124028 (2009).

[91] Y. Pan, A. Buonanno, L. T. Buchman, T. Chu, L. E. Kidder, H. P. Pfeiffer, and M. A. Scheel, Phys. Rev. D **81**, 084041 (2010).
[92] E. Barausse, A. Buonanno, and A. Le Tiec, Phys. Rev. D **85**, 064010 (2012).
[93] Y. Pan, A. Buonanno, M. Boyle, L. T. Buchman, L. E. Kidder, H. P. Pfeiffer, and M. A. Scheel, Phys. Rev. D **84**, 124052 (2011).
[94] T. Damour and A. Nagar, Phys. Rev. D **90**, 044018 (2014).
[95] E. Berti, V. Cardoso, and A. O. Starinets, Classical Quantum Gravity **26**, 163001 (2009).
[96] J. Abadie *et al.* (LIGO Collaboration, Virgo Collaboration), Phys. Rev. D **85**, 082002 (2012).
[97] B. Brügmann, J. A. González, M. Hannam, S. Husa, U. Sperhake, and W. Tichy, Phys. Rev. D **77**, 024027 (2008).
[98] S. Husa, M. Hannam, J. A. González, U. Sperhake, and B. Brügmann, Phys. Rev. D **77**, 044037 (2008).
[99] D. Pollney, C. Reisswig, L. Rezzolla, B. Szilágyi, M. Ansorg, B. Deris, P. Diener, E. N. Dorband, M. Koppitz, A. Nagar, and E. Schnetter, Phys. Rev. D **76**, 124002 (2007).
[100] D. Pollney, C. Reisswig, E. Schnetter, N. Dorband, and P. Diener, Phys. Rev. D **83**, 044045 (2011).
[101] M. A. Scheel, H. P. Pfeiffer, L. Lindblom, L. E. Kidder, O. Rinne, and S. A. Teukolsky, Phys. Rev. D **74**, 104006 (2006).
[102] P. Ajith, Phys. Rev. D **84**, 084037 (2011).
[103] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Phys. Rev. D **93**, 044006 (2016).
[104] L. Lindblom, B. J. Owen, and D. A. Brown, Phys. Rev. D **78**, 124020 (2008).
[105] J. Aasi *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. D **88**, 062001 (2013).
[106] J. Veitch, M. Purrer, and I. Mandel, Phys. Rev. Lett. **115**, 141101 (2015).
[107] C.-J. Haster, Z. Wang, C. P. L. Berry, S. Stevenson, J. Veitch, and I. Mandel, arXiv:1511.01431.
[108] T. B. Littenberg, J. G. Baker, A. Buonanno, and B. J. Kelly, Phys. Rev. D **87**, 104003 (2013).
[109] M. Pürrer, M. Hannam, P. Ajith, and S. Husa, Phys. Rev. D **88**, 064007 (2013).
[110] T. A. Apostolatos, Phys. Rev. D **52**, 605 (1995).
[111] B. Szilágyi, J. Blackman, A. Buonanno, A. Taracchini, H. P. Pfeiffer, M. A. Scheel, T. Chu, L. E. Kidder, and Y. Pan, Phys. Rev. Lett. **115**, 031102 (2015).
[112] P. B. Graff, A. Buonanno, and B. S. Sathyaprakash, Phys. Rev. D **92**, 022002 (2015).
[113] C. Loken, D. Gruner, L. Groer, R. Peltier, N. Bunn, M. Craig, T. Henriques, J. Dempsey, C.-H. Yu, J. Chen, L. J. Dursi, J. Chong, S. Northrup, J. Pinto, N. Knecht, and R. V. Zon, J. Phys. Conf. Ser. **256**, 012026 (2010).