

## Parameter estimation with a spinning multimode waveform model

Chinmay Kalaghatgi<sup>✉</sup>, Mark Hannam, and Vivien Raymond

*School of Physics and Astronomy, Cardiff University,  
Queens Buildings, Cardiff, CF24 3AA, United Kingdom*

 (Received 4 October 2019; accepted 6 April 2020; published 1 May 2020)

Gravitational waves from compact binary coalescence sources can be decomposed into spherical-harmonic multipoles, the dominant being the quadrupole ( $(l, |m|) = (2, 2)$ ) modes. The contribution of subdominant modes toward total signal power increases with increasing binary mass ratio and source inclination to the detector. It is well known that in these cases neglecting higher modes could lead to measurement biases, but these have not yet been quantified with a higher-mode model that includes spin effects. In this study, we use the multimode aligned-spin phenomenological waveform model IMRPhenomHM [1] to investigate the effects of including multimode content in estimating source parameters and contrast the results with using a quadrupole-only model (IMRPhenomD). We use as sources IMRPhenomHM and hybrid effective-one-body–numerical-relativity waveforms with zero spin over a range of mass-ratio and inclination combinations, and recover the parameters with IMRPhenomHM and IMRPhenomD. These allow us to quantify the accuracy of parameter measurements using a multimode model, the biases incurred when using a quadrupole-only model to recover full (multimode) signals, and the systematic errors in the IMRPhenomHM model. We see that the parameters recovered by multimode templates are more precise for all nonzero inclinations as compared to quadrupole templates. For multimode injections, IMRPhenomD recovers biased parameters for nonzero inclinations with lower likelihood while IMRPhenomHM-recovered parameters are accurate for most cases, and if a bias exists, it can be explained as a combined effect of observational priors and (in the case of hybrid-NR signals) waveform inaccuracies. However, for cases where IMRPhenomHM recovers biased parameters, the bias is always significantly smaller than the corresponding IMRPhenomD recovery, and we conclude that IMRPhenomHM will be sufficiently accurate to allow unbiased measurements for most gravitational wave observations.

DOI: [10.1103/PhysRevD.101.103004](https://doi.org/10.1103/PhysRevD.101.103004)

### I. INTRODUCTION

The first gravitational-wave (GW) detection, from a binary black hole (BBH) merger, was achieved on the September 14, 2015 [2] by the two advanced LIGO (aLIGO) detectors at Hanford and Livingston [3]. Two observing runs have since been completed by the aLIGO detectors, and from the second half of 2017, Advanced Virgo (AdV) [4] joined the GW detector network, facilitating the first three-detector observation of a BBH source [5]. During the first two observation runs of aLIGO and AdV, a total of ten BBH mergers and one binary neutron star merger were detected [6,7]. Most signal measurements were performed using waveform models that included only the dominant quadrupole harmonic, although one signal (GW170729) showed evidence for a binary with unequal masses, for which models that include higher harmonics allow for improved measurements [8]. The goal of this work is to quantify the measurement accuracy possible with higher-multipole models.

Any GW  $h(\theta, \phi, \vec{\lambda}, t)$  can be decomposed in terms of spherical harmonics with spin-weight  $-2$ ,  ${}^{-2}Y_{lm}(\theta, \phi)$ ,

$$h(\theta, \phi, \vec{\lambda}, t) = \sum_l \sum_{m=-l}^{m=l} {}^{-2}Y_{lm}(\theta, \phi) h_{lm}(\vec{\lambda}, t), \quad (1)$$

where  $h_{lm}(\vec{\lambda}, t)$  are the GW modes, and  $\vec{\lambda}$  are the intrinsic parameters of the source, i.e., the black-hole masses and spins. For coalescing binary systems with aligned spins, or in the coprocessing frame of precessing systems [9–11], the quadrupole modes  $(l, |m|) = (2, 2)$  are the strongest. Relative to the corresponding quadrupole mode, the subdominant modes ( $l = 3, 4, 5, \dots$ ;  $|m| \in [0, l] \forall l$ ) are weakest for equal-mass systems, their strength increasing with increasing mass ratio. In addition, for a given system, as the binary's inclination to the detector is increased from face-on ( $\theta = 0^\circ$ ) to edge-on ( $\theta = 90^\circ$ ), the contribution of the dominant modes decreases, as does the overall signal power, and the relative importance of subdominant modes grows.

Black-hole binaries in noneccentric orbits are characterized by the two black-hole masses,  $m_1$  and  $m_2$ , and the black-hole spins  $\mathbf{S}_1$  and  $\mathbf{S}_2$ . The inspiral rate of the binary (and phasing of the GW signal, which is crucial to measuring the properties of the binary) is effected most

strongly by combinations of these parameters: the chirp mass,  $\mathcal{M}_c = (m_1 m_2)^{3/5} / (m_1 + m_2)^{1/5}$  (during the inspiral), and the total mass,  $M = m_1 + m_2$  (during the merger and ringdown); the mass ratio  $q = m_1/m_2$ , or alternatively symmetric mass ratio  $\eta = m_1 m_2 / M^2$ ; and the weighted sum of the spin components parallel to the orbital angular momentum  $\hat{\mathbf{L}}$ ,  $\chi_{\text{eff}} = (m_1 \chi_1 + m_2 \chi_2) / M$ , where  $\chi_i = \mathbf{S}_i \cdot \hat{\mathbf{L}} / m_i^2$  [12–15]. The overall strength of the GW signal scales with  $M/d_L$ , where  $d_L$  is the distance from the source to the detector, and is also affected by the binary inclination  $\theta_{\text{JN}}$ , which is the angle between the total angular momentum  $\hat{\mathbf{J}}$  (or equivalently  $\hat{\mathbf{L}}$  for aligned-spin binaries) and the line of sight,  $\hat{\mathbf{N}}$ . If there are spin components perpendicular to  $\hat{\mathbf{L}}$ , then the binary’s orbital plane and spins will precess, leading to modulations of the GW signal. In general, precession has little effect on the overall GW phase [16,17], and so might not strongly affect the measurement of other parameters [18,19]. Precession has not yet been measured in GW observations [6], and for these reasons we focus on aligned-spin binaries in this study.

Waveform models describing the inspiral-merger-ringdown (IMR) stages of BBH mergers are available for nonspinning, aligned-spin and precessing configurations. Many of the aligned-spin waveform models model only the dominant quadrupole mode, but several multimode models now exist. For nonspinning systems, there are effective-one-body (EOB)–numerical-relativity (NR) multimode models, EOBNRv2HM [20] and TEOBiResumMultipoles [21], and “Phenom” models [22,23]. The EOBNR model was recently extended to aligned-spin systems (SEOBNRv4HM) [24] and models the  $(l, |m|) = (2, 2), (2, 1), (3, 3), (4, 4)$ , and  $(5, 5)$  modes. In this study, we will use the phenomenological aligned-spin multimode model IMRPhenomHM [1], which models the  $l = (2, 3, 4)$  and  $|m| = (l, l - 1)$  modes; IMRPhenomHM is described in more detail in Sec. II A. Higher modes are also included in a series of surrogate models constructed from NR waveforms [25–28].

Previous studies have investigated the effect of employing higher-order mode models for gravitational wave searches [29–32] and provided an estimate of the systematic errors that could be incurred from neglecting higher-order modes in the template waveforms [33–35]. In [36,37], the authors performed a full Bayesian analysis of the effects of including and neglecting higher-order modes in template waveforms nonspinning systems. We summarize some notable results relevant to the current study.

In Ref. [35], the authors used multimode Post-Newtonian - Numerical Relativity (PN-NR) hybrids as signals and computed the expected statistical and systematic errors from using quadrupole-only templates to estimate source parameters over a range of total mass and mass-ratio values, for a signal sky-averaged signal-to-noise ratio (SNR) of 8. The statistical errors are estimated from the Fisher information matrix, which is the noise-weighted

inner product between partial derivatives of the waveform. The authors also estimate the systematic errors by calculating the fitting factor, which is the noise-weighted inner product between the signal and model waveforms, maximized over the model parameters. The effective systematic error is proportional to the difference between the best fit and true parameters. In this study, the authors found that noninclusion of the subdominant modes in templates will lead to  $\sim 10\%$  loss in detection rate for  $q \geq 6$  and  $M \geq 100 M_\odot$  and will lead to systematic errors larger than statistical errors for  $q \geq 4$  and  $M \geq 150 M_\odot$ . The results obtained from a Fisher information matrix approximation are valid for high SNR events. To study the waveform errors for low or moderate SNRs and for a realistic assessment of the model’s measurement capabilities, a full Bayesian analysis is required.

In a Bayesian analysis, the physical parameters of the source are estimated by matching the detector data with model waveforms. Given detector data  $d(t)$  and a waveform model  $h(t)$ , the posterior over  $\vec{\lambda}$  is

$$p(\vec{\lambda}|d(t), h(t, \vec{\lambda})) = \frac{p(\vec{\lambda}|h(t, \vec{\lambda}))p(d(t)|\vec{\lambda}, h(t, \vec{\lambda}))}{p(d(t)|h(t, \vec{\lambda}))}, \quad (2)$$

where  $\vec{\lambda}$  is the vector of intrinsic and extrinsic parameters.  $p(\vec{\lambda}|h(t, \vec{\lambda}))$ ,  $p(d(t)|\vec{\lambda}, h(t, \vec{\lambda}))$ , and  $p(d(t)|h(t, \vec{\lambda}))$  are the prior, likelihood, and evidence, respectively, where the likelihood is

$$p(d|\vec{\lambda}, h(\vec{\lambda})) \propto e^{-\frac{1}{2}\langle d-h(\vec{\lambda})|d-h(\vec{\lambda}) \rangle}, \quad (3)$$

and the quantity  $\langle a|b \rangle$  gives the noise-weighted inner product between the two functions; this is the *match* if  $a$  and  $b$  are descriptions of waveforms from systems with the same physical parameters, optimized over a relative time and phase shift. See Ref. [38] for more details on the techniques and algorithms employed for GW parameter estimation. Accuracy of the inferred parameters depends on the accuracy of the waveform model used to simulate the real signal and the noise content of the detector data.

In Ref. [36], the authors injected multimode nonspinning NR waveforms in zero noise at different mass ratios with fixed inclination of  $60^\circ$  and compared the systematic and statistical errors of the posteriors recovered by nonspinning quadrupole-only (EOBNRv2) and multimode (EOBNRv2HM) waveform models for nonspinning systems. They found that up to  $q = 6$  and for SNRs  $\leq 50$ , the systematic errors from EOBNRv2HM were smaller than or comparable to the statistical errors. The fractional systematic error (defined as the ratio between systematic bias and statistical error) for the intrinsic parameters is consistently lower for EOBNRv2HM than EOBNRv2. Also, the posteriors were recovered at an overall higher likelihood by EOBNRv2HM than EOBNRv2 (see Fig. 2 of Ref. [36]).

In Ref. [37], the authors performed a comprehensive study of the effects of using EOBNRv2HM and EOBNRv2 templates to recover EOBNRv2HM signals across a range of total mass values ( $50 M_{\odot} \leq M_{\text{total}} \leq 500 M_{\odot}$ ) and SNRs ( $6 \leq \rho \leq 18$ ) for  $q = 1.25$  and  $q = 4$  systems at two inclinations ( $\theta_{\text{JN}} = 0^\circ, 60^\circ$ ). Consistent with Ref. [36], the posteriors are recovered at an overall larger total evidence by EOBNRv2HM compared to EOBNRv2 for inclined systems; see Fig. 5 of Ref. [37]. These differences increase with increasing inclination, mass ratio, and total mass. The posteriors are better constrained by the multimode model than a quadrupole-only model (see Fig. 7 of Ref. [37]) with lower systematic bias for inclined systems. They found that the multimode model constrains the inclination angle better than quadrupole-only model, which in turn leads to better constraints on the distance.

In the previous studies, the authors used nonspinning multimode and quadrupole-only waveforms for the Bayesian analysis, and hence, were restricted in the  $(m_1, m_2)$  space for intrinsic parameters. In this study, we will use a multimode aligned-spin waveform template (IMRPhenomHM) and increase the dimensionality of the problem by one, i.e., covering the  $(m_1, m_2$  and  $\chi_{\text{eff}})$  space of intrinsic parameters. Of course, the extrinsic parameter space remains the same.

One of the aims of this study is to explore the effects of using a multimode waveform template (IMRPhenomHM) on inferring source parameters from a multimode signal and to contrast it with a quadrupole-only model (IMRPhenomD). For this, we perform a set of injections at three different mass ratios and three inclinations in zero noise using the IMRPhenomHM model. This allows us to quantify parameter errors from not including subdominant modes in templates and accuracy improvements when the subdominant modes are included.

IMRPhenomHM is an approximate model of the subdominant modes and does not model all of the higher harmonics (all modes with  $l \geq 5$ , all  $m = 0$  modes, and the (3,1), (4,2), (4,1) modes). The subdominant modes of the model are not

tuned to NR simulations and mode-mixing effects are not modeled. With that in mind, the other aim of the study is to determine the ability of IMRPhenomHM to recover parameters of real physical signals. For that, we perform the same set of injections as for IMRPhenomHM injections but with multimode EOB-NR hybrid waveforms and compare the parameters recovered by IMRPhenomD and IMRPhenomHM. Hybridization is required to include the contributions to the signals below the starting frequency of the NR waveforms.

Note that this study is not exhaustive enough to make quantitative statements of the bias across the whole mass-spin parameter space as we choose to consider systems at a fixed total mass, coalescence phase, polarization, sky position, and zero spins. As was shown in [39], varying polarization can change recovered posteriors for inclined systems. From Fig. 1, it can be seen that the coalescence phase has a large effect on the matches and as was shown in [33], the same can cause varying biases on recovered parameters. The point of this study is to explore the general trends of recovered parameters across the mass ratio and inclination space with multimode models and make approximate statements of the validity of IMRPhenomHM in those regions. A more detailed systematic exploration of the parameter space is beyond the scope of this paper.

Section II will provide a short summary of the template waveform models IMRPhenomHM and IMRPhenomD construction of the multimode hybrids and details of the setup for parameter estimation. The results will be introduced in general in Sec. III, and the specific results for intrinsic and extrinsic parameters will be given in Secs. IV and V.

## II. METHODS

### A. Summary of waveform models

IMRPhenomD [40,41] is a quadrupole-only frequency-domain phenomenological waveform model describing inspiral-merger-ringdown (IMR) stages of aligned-spin BBH systems. IMRPhenomD is calibrated to NR simulations

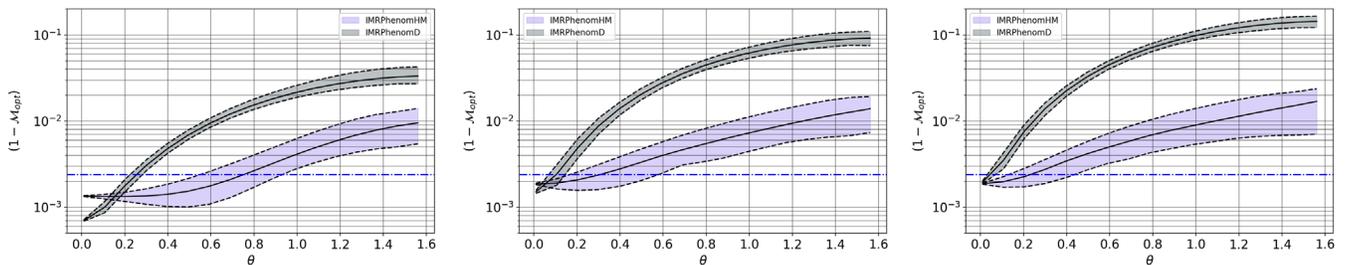


FIG. 1. Mismatches between hybrid-NR and IMRPhenomHM waveforms (blue) and IMRPhenomD waveforms (gray). The left, central, and right panels give the mismatches for  $q = (2, 4, 8)$  systems. The match is computed over a range of signal inclination, phase and polarization values with the match optimized over template phase and polarization. We quote the minimum (dashed-lower bound), average (central black line), and maximum value (dashed-upper bound) of the match at each signal inclination. The dashed blue line shows the minimum match required for the obtaining minimal PE bias (at 90% credible intervals) for an injected SNR of 25, by using the relation  $\rho_c \leq \sqrt{\chi_k^2(1-p)/2(1-\mathcal{M})}$  [15] for  $k = 3$ . For the match, the total mass is set to  $100 M_{\odot}$  with a lower-frequency bound of  $f_{\text{min}} = 30$  Hz. This choice of lower frequency cutoff for the matches was made to keep these results consistent with those in Fig 3 of [1].

with  $1 \leq q \leq 18$  and spins  $|\chi_i| \lesssim 0.85$ . The agreement between IMRPhenomD and other quadrupole-only models is extremely good in the region of parameter space that we consider in this study [41,42], and so we expect that results obtained with this model will be indicative of the performance of any accurate quadrupole-only model.

IMRPhenomHM [1] is a frequency-domain aligned-spin phenomenological waveform model, wherein results from post-Newtonian theory (for inspiral) and BH perturbation theory (for ringdown) are used to map the dominant quadrupole mode to the subdominant modes. So, given the quadrupole frequency ( $f_{22}$ ), amplitude ( $A_{22}$ ) and phase ( $\varphi_{22}$ ), the frequency ( $f_{lm}$ ), amplitude ( $A_{lm}$ ), and phase ( $\phi_{lm}$ ) of other modes are computed via

$$\tilde{h}_{lm}(f) = A_{lm}(f) \times \exp\{i\varphi_{lm}(f)\} \quad (4)$$

$$\approx |\beta_{lm}(f)| A_{22}(f_{22}^A) \times \exp\{i[\kappa(f)\varphi_{22}(f_{22}^A) + \Delta_{lm}(f)]\}, \quad (5)$$

where  $|\beta_{lm}(f)|$  and  $\kappa(f)$  are amplitude and phase scaling functions and  $\Delta_{lm}(f)$  are phase shifts to be computed separately for each mode.

IMRPhenomHM uses IMRPhenomD for the quadrupole mode information to obtain the subdominant modes. Though the subdominant modes of IMRPhenomHM are not calibrated to NR simulations and are thus an *approximation*, IMRPhenomHM has much better matches to NR waveforms than a quadrupole-only model; see Fig. 1. As discussed in Ref. [1], the higher mismatch of IMRPhenomHM as compared to IMRPhenomD for face-on  $q = 2$  waveforms is due to inaccuracy in the modeling of the  $(l, |m|) = (3, 2)$  modes by the former. We refer the reader to Ref. [1] for further information on the construction and validity of the model.

## B. Construction and validation of multimode hybrids

The discussion below closely follows that of Sec. VI of Ref. [43].

Two perfectly accurate gravitational waveforms [ $h^a(t)$  and  $h^b(t)$ ] computed with different methods or with

differing conventions can be mapped between each other with a time and an orbital phase shift, along with a shift in polarization to account for differing conventions. So, we can write  $h^a(t)$  as

$$h^a(t, \theta, \phi) = h^b(t + \tau, \theta, \phi + \phi_0) e^{i\psi_0}. \quad (6)$$

This allows us to relate the modes of the two waveforms with each other as

$$h_{lm}^a(t, \theta, \phi) = h_{lm}^b(t + \tau, \theta) e^{i(\psi_0 + m\phi_0)}. \quad (7)$$

So, given a PN inspiral waveform and an NR waveform for the same physical configuration, if the NR waveform is long enough, then there would be a common region where both the waveforms are accurate and agree with each other. We can then construct a hybrid waveform by *stitching* the two within the overlapping regions. The symmetry requirements of nonprecessing systems [see Eq. (11)] restrict  $\psi_0$  to be either 0 or  $\pi$ ; a more detailed discussion of which is given in Ref. [43].

The problem then reduces to finding appropriate time and orbital phase shifts between the two waveforms. To obtain the time shift ( $\tau$ ), we minimize the quantity  $\Delta(\tau, t_0, dt)$  with respect to  $\tau$ , where  $\Delta(\tau, t_0, dt)$  is given as

$$\Delta(\tau, t_0, dt) = \int_{t_0}^{t_0+dt} (\omega^{\text{PN}}(t) - \omega^{\text{NR}}(t - \tau))^2 dt. \quad (8)$$

Using this time shift, we construct the phase integral  $\Phi(\phi_0)$ ,

$$\Phi(\phi_0) = \int_{t_0}^{t_0+dt} (\phi^{\text{NR}}(t - \tau) - \phi^{\text{PN}}(t) + \phi_0)^2 dt. \quad (9)$$

The orbital phase shift used for constructing the hybrid would be the one that minimizes  $\Phi(\phi_0)$ .

Once the quantities  $(\tau, \phi_0, \psi_0)$  are computed, the hybrid waveform is obtained by stitching the PN and NR waveforms as

$$h_{lm}(t) = \begin{cases} e^{i(m\phi_0 + \psi_0)} h^{\text{PN}}(t + \tau) & t < t_0 - \tau \\ w^-(t) e^{i(m\phi_0 + \psi_0)} h^{\text{PN}}(t + \tau) + w^+(t) h^{\text{NR}}(t) & t_0 - \tau < t < t_0 - \tau + dt \\ h^{\text{NR}}(t) & t > t_0 - \tau + dt, \end{cases} \quad (10)$$

where  $w^-(t)$  and  $w^+(t)$  are the blending functions which go from [1,0] and [0,1], respectively, between  $t_0 - \tau < t < t_0 - \tau + dt$ . We use Planck taper windowing function [44] for blending the PN and NR waveforms.

To construct the hybrids for injections, we use an EOB code to generate the inspiral and hybridize it with the

corresponding public SXS NR waveform [45] following the procedure summarized above. The reason for using the SXS waveforms is that these include more inspiral cycles than any other currently available set. The EOB code used to obtain the inspiral modes is based on the method described in Refs. [46,47] with the fits to the parameters

as published in Ref. [48]. Since our purpose is to construct hybrids, we only use the inspiral contribution to these EOB waveforms. We construct the hybrid for (2,2), (2,1), (3,3), (3,2), (3,1), (4,4), (4,3), and (4,2) modes and the negative “ $m$ ” modes are related to the positive  $m$  modes by the relation

$$h_{l,m}(t, \vec{\lambda}) = (-1)^l h_{l,-m}^*(t, \vec{\lambda}). \quad (11)$$

The EOB and NR waveforms are matched at the time when the (2,2) mode has frequencies  $M\omega_{2,2} = (0.072, 0.066, 0.044)$  for mass ratios  $q = (2, 4, 8)$  with a hybridization window of  $dt = 200M$ . The hybrid-NR waveforms are validated by computing the match between the hybrid waveform and corresponding SXS waveform (which all have a starting frequency of  $M\omega_{2,2} \sim 0.04$ ), over the  $(\theta, \phi)$  space, where the match is maximized over the phase of the hybrid waveform. We find that the match is never less than 0.9999.

### C. Setup

For this study, we use hybrid-NR waveforms and IMRPhenomHM waveforms for injections. We create EOB-NR multimode hybrids for nonspinning  $q = (2, 4, 8)$  systems. The waveforms are injected at a constant SNR of 25, total mass of  $100 M_\odot$ , and at inclinations of  $0^\circ$ ,  $60^\circ$ , and  $90^\circ$ . We choose a polarization value ( $\psi$ ) of 1.4 and the gps time is set to 1186741623. For the recovery PSD, we compute the median detector PSD via BayesWave [49,50] with gps time set to be near to the trigger time for GW170814 [51]. This is so that the recovery PSD is close to the final O2 sensitivity. The accuracy of the recovered extrinsic parameters will depend strongly on the total detector response. We choose the sky position (for a given polarization and gps time) such that the total detector response for Hanford and Livingston is of comparable value. The right ascension and declination values used are 0.2897 and 1.4323 rad, respectively. All waveforms are injected in zero noise. The lower frequency cutoff for both injected signal and for the parameter estimation (PE) runs is set to 20 Hz. See Ref. [52] for details regarding the frame transformations performed during the injection and definitions of the above parameters.

For this study, we performed a total of 36 PE runs. For each of the three mass-ratio and inclination combinations, we perform both hybrid-NR and IMRPhenomHM injections. For each injection, the parameters were recovered using both IMRPhenomD and IMRPhenomHM. All signals have an SNR of 25; this is at the high end of SNRs we would expect in aLIGO and AdV observations (at best roughly only 1 in 15 observations will have a higher SNR, assuming a detection threshold at SNR 10 [53]), so provides an indication of the best measurement precision we could achieve, as well as the worst impacts of measurement biases and systematic errors.

The PyCBC [54] hardware injection function (PYCBC\_GENERATE\_HWINJ) is used to generate the injection frames. We use the LALInferenceNest [38] pipeline to obtain the posterior samples. All runs are performed with 1024 live points.

### III. RESULTS

It has already been shown in a number of previous studies [36,37,55–57] that recovery using multimode models can improve the measurement of intrinsic and extrinsic parameters (depending on total mass of the system). Here we present the first results that quantify these effects when recovery is performed with an aligned-spin multimode IMR model. Given that there is a well-known partial degeneracy between the binary’s mass ratio and the black-hole spins [15], we expect the inclusion of spin in the recovery template to significantly affect the precision of the parameter recovery, even for signals from nonspinning binaries. Furthermore, the effects of precession (which are not considered here) in general are driven by the in-plane spins [58,59] and precession measurement approximately decouples from aligned-spin parameters [19,60] and hence, we expect that our results will in many cases carry over to recovery using generic-binary models.

We first discuss the recovery of the intrinsic parameters (black-hole masses and spins), and then consider the extrinsic parameters, i.e., the distance ( $d_L$ ) and inclination ( $\theta_{\text{JN}}$ ). As we are considering only nonspinning or aligned-spin binaries,  $\vec{L} \parallel \vec{J}$  and so,  $\theta_{\text{JN}} = \theta_{LN}$ .

IMRPhenomHM is an approximate model for the subdominant modes and in particular is not tuned to fully general-relativistic NR results through the merger and ringdown. Systematic errors due to these approximations can be tested by using IMRPhenomHM to recover injections of hybrid-NR waveforms. Also, our hybrid-NR waveforms contain extra mode content, namely, the  $(l, m) = (3, 1)$  and  $(4, 2)$  modes, but given that these modes contribute less than 10% to total signal power (even for edge-on configurations), we expect that the dominant source of systematic errors will be amplitude and phase errors in the modes that are present in the model.

The results from the IMRPhenomHM injections quantify the accuracy of parameter recovery using a multimode model, and any biases that may be incurred by using a quadrupole-only model. We expect these to be largely independent of the choice of model; if IMRPhenomD and IMRPhenomHM were replaced by some other (accurate) quadrupole-only and multimode models, the results would show similar qualitative behaviors. In contrast, the results of the hybrid-NR injections indicate the systematic errors that might be incurred when estimating the parameters of real data using the approximate IMRPhenomHM model. With that in mind, we split the results of intrinsic parameter recovered for IMRPhenomHM and hybrid-NR injections.

Our main results are in the form of the posterior distributions for each of the 18 parameter-estimation runs (three mass ratios, three inclinations, two recovery models). The posteriors are truncated at the 90% confidence interval, indicating the uncertainty in each parameter measurement. Comparison against the true parameters indicates whether the measurement is unbiased (the true value lies within the 90% CI), or the level of bias. We also use the opacity of the distributions to indicate the relative log-likelihood ( $\Delta \log \mathcal{L}$ ), which tells us how well the model agrees with the signal. In our zero-noise injections, the log-likelihood is proportional to  $-|h_S(\lambda_0) - h_M(\lambda)|$ , where  $h_S$  is the injected signal, with the specific parameters  $\lambda_0$ , and  $h_M$  is the template model evaluated with parameters  $\lambda$ , and the magnitude of the difference between the signal and model is calculated using an inner product weighted with the detector's spectral noise density. If the model is able to exactly reproduce the signal (as in the case of IMRPhenomHM injection and recovery), then the maximum log-likelihood will be  $\log \mathcal{L} \approx 0$  and less than zero for all other cases. In our figures, an opaque posterior distribution indicates excellent agreement between the signal and model, while a more transparent posterior distribution indicates that the model parameters that provide the best agreement with the signal are nonetheless a poor representation of it.

Another way to quantify the difference between the performance of two models that we report is the Bayes factor, which measures the probabilistic support of one model over another, as opposed to the maximum likelihood, which provides the goodness of fit. We can quantify the support of IMRPhenomHM over IMRPhenomD for any injection with the logarithm of the Bayes factor between the two models,  $\log(B_{PhnD}^{PhnHM})$ . For instance, a log Bayes factor of  $\log(B_{PhnD}^{PhnHM}) = 5$  means that IMRPhenomHM is  $e^5$  more likely than IMRPhenomD. Note that those probabilities are given by comparing the models IMRPhenomHM-plus-Gaussian-noise versus IMRPhenomD-plus-Gaussian-noise. While in this study we use zero noise as the noise realization, in practice Gaussian noise is only an approximation for detector noise and the Bayes factor is used with an empirically set threshold.

#### IV. RECOVERY OF INTRINSIC PARAMETERS

In this section, we discuss the differences between the source mass and spin parameters ( $\mathcal{M}_c, q, \chi_{\text{eff}}, M_{\text{total}}$ ) when recovered by both IMRPhenomD and IMRPhenomHM.

Overall, we see that the IMRPhenomD recovered parameters are consistently biased for inclined systems for both IMRPhenomHM and hybrid-NR injections, with the bias increasing with increasing mass ratio of the system. Parameters recovered by IMRPhenomHM also show a bias for large ( $q, \theta_{\text{IN}}$ ) hybrid-NR injections, but this bias is always smaller than the corresponding IMRPhenomD recovery. If a bias exists for IMRPhenomHM injection-IMRPhenomHM

recovery, it can be explained by marginalization and prior effects (see Sec. IV A).

For a given (mass ratio, inclination) configuration, the improved constraints on the inclination and distance parameters by using multimode templates might lead to better constraints on the intrinsic parameters. To check for that, let us define  $d_{\lambda_i}^{\text{model}} = C_{\lambda_i}^{\text{upper}} - C_{\lambda_i}^{\text{lower}}$ , where  $C_{\lambda_i}^{\text{upper}}$  and  $C_{\lambda_i}^{\text{lower}}$  are the upper and lower bounds of the 90% CI for a given parameter  $\lambda_i$ . Hence,  $d_{\lambda_i}^{\text{model}}$  would provide a measure of the posterior width. Using this, we define the relative percentage difference between the credible interval widths for a given configuration ( $\Delta_{\lambda_i}$ ) as

$$\Delta_{\lambda_i} = 100 \left( \frac{d_{\lambda_i}^{\text{IMRPhenomD}} - d_{\lambda_i}^{\text{IMRPhenomHM}}}{d_{\lambda_i}^{\text{IMRPhenomD}}} \right). \quad (12)$$

For a given intrinsic parameter,  $\Delta_{\lambda_i}$  would quantify the improvements on the parameter constraints from using multimode templates.

For the following plots, the posterior over a parameter from each run is clipped within its 90% credible intervals and we plot them as a violin plot. For each parameter, the y axis shows the value of the recovered posterior and the x axis gives the injected inclination-recovery waveform combination. For example, if the recovery is for an edge-on injection by IMRPhenomD, it is labeled as  $\iota = 90^\circ \text{PhnD}$ . Posteriors for  $q = 2, 4, 8$  systems are shown in blue, gray, and orange, respectively. We will first discuss the results of IMRPhenomHM injections.

#### A. IMRPhenomHM injections

The results of recovered intrinsic parameters for the IMRPhenomHM injections are given in Fig. 2. The percentage improvement in parameter measurements,  $\Delta_{\lambda_i}$ , is shown for  $\lambda_i \in (\mathcal{M}_c, M_{\text{total}}, q, \chi_{\text{eff}})$  in Fig. 3.

We consider the face-on systems first, which are the left-hand columns in each of the panels in Fig. 2 and the upper panel in Fig. 3. For all mass ratios, at face-on inclination the posteriors recovered by IMRPhenomHM and IMRPhenomD are consistent with each other, show no bias, and are recovered at very similar maximum likelihood. This is expected due to the *almost zero* contributions ( $<10\%$ ) of higher-order modes to total signal power at face-on inclination and due to the underlying quadrupole model for IMRPhenomHM being IMRPhenomD.

The log Bayes factor between IMRPhenomHM and IMRPhenomD  $\log(B_{PhnD}^{PhnHM})$  for all face-on systems is  $\leq 4$ . At  $q = 2$ , the confidence intervals are almost identical between IMRPhenomHM and IMRPhenomD recovery, although the mass-ratio recovery shows a greater preference for lower  $q$ . Any improvements in the measurement precision are difficult to detect in the posterior plots, but are clear in Fig. 3. We see that  $q$  is always recovered slightly better by IMRPhenomHM even for the  $q = 2$  system, with an

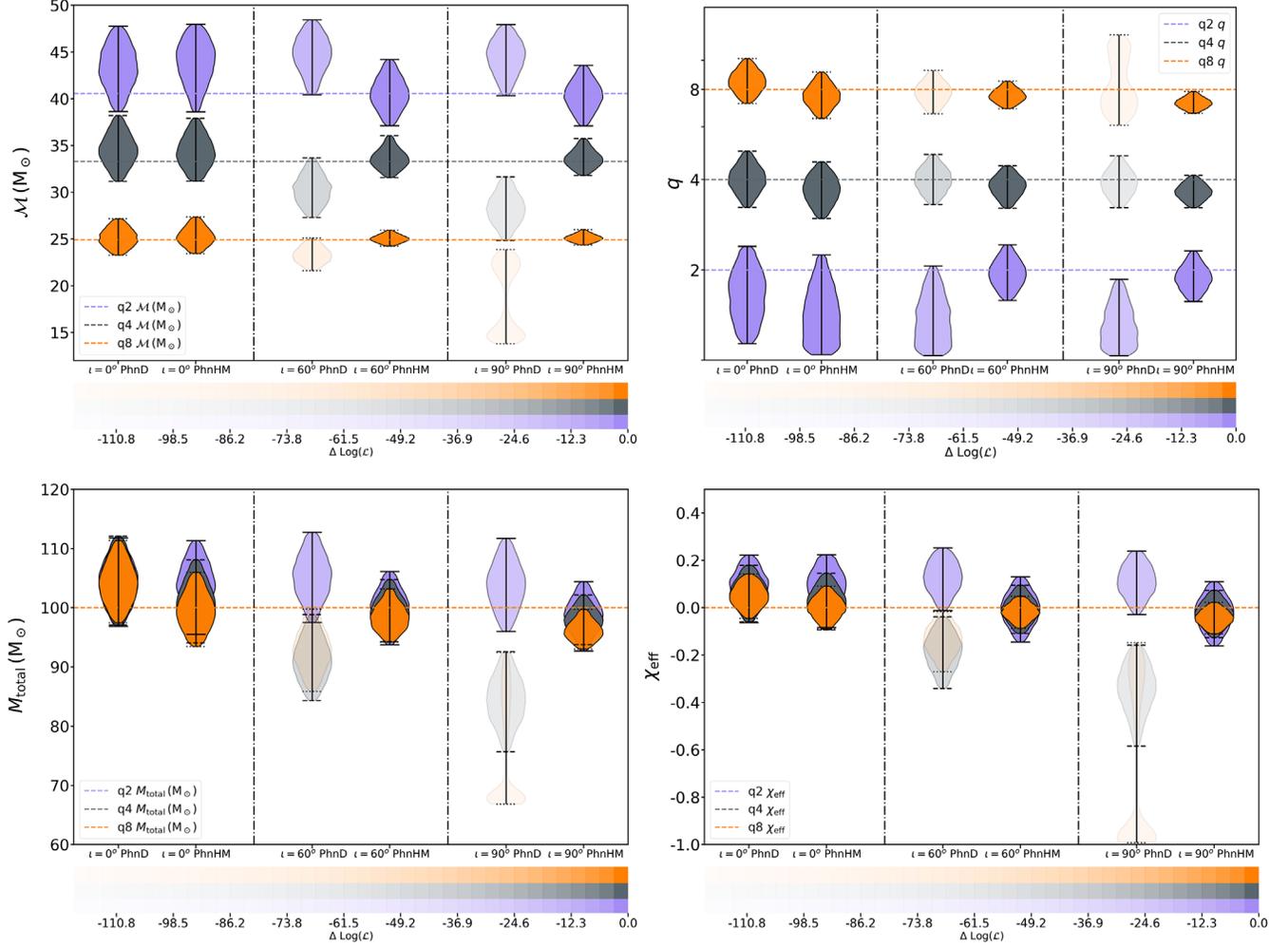


FIG. 2. Posteriors of intrinsic parameters ( $\mathcal{M}_c$ ,  $q$ ,  $M_{\text{total}}(M_{\odot})$ ,  $\chi_{\text{eff}}$ ) for IMRPhenomHM waveform injected at  $q = 2$ ,  $q = 4$ ,  $q = 8$  with  $\theta_{\text{JN}} = 0^\circ, 60^\circ, 90^\circ$ . Posteriors for  $q = 2$  ( $q = 4$ ) [ $q = 8$ ] are shown in blue (gray) (orange) with the opacity of each determined from the maximum likelihood value of that run. The variation of opacity over the likelihood values is shown at the bottom of each graph. See Sec. IV A for a description of these results, specifically, the bimodal posteriors recovered by IMRPhenomD at  $q = 8$ ,  $\theta_{\text{JN}} = 90^\circ$ .

improvement of  $\sim 5\%$ , with the recovery of the other parameters slightly *less* constrained at  $q = 2$ . Although the effect is small, for face-on signals, the higher-mode content *does* lead to a slight improvement at higher mass ratios, where we see that the total mass and effective spin are recovered more accurately, and the effective spin measurement is also more precise.

Now consider systems with inclination  $60^\circ$ . The effects of higher modes are perhaps the most relevant for these cases, because this is where we statistically expect to observe greater number of signals; this is clear from, for example, the IMRPhenomD inclination recoveries in Fig. 4, which predominantly recover the observational prior. In the middle column of each Fig. 2 panel, we see that the IMRPhenomD recovery starts to show a bias away from the true values and the parameters are recovered at comparatively lower likelihood. These effects are stronger with increasing subdominant mode contribution to total signal

power. At  $q = 2$ , the IMRPhenomD recovery of  $\mathcal{M}_c$  and  $M_{\text{total}}$  are slightly biased away from the true value toward overall higher total mass and more equal mass. For a given  $q$ , the waveform length decreases (increases) at higher (lower) total mass or more negative (positive)  $\chi_{\text{eff}}$ . For  $q = 2$  system, the shift to an overall higher total mass is compensated by higher  $\chi_{\text{eff}}$  recovery. We might expect that the increase in total mass puts more power into the signal, to mimic the extra power that is there from the higher harmonics, but we see the opposite trend for mass ratios  $q = 4$  and  $q = 8$ . Regardless of the parameter shifts in the quadrupole-only model to find the best agreement with the higher-mode signal, it is clear that best matching IMRPhenomD signals does not agree especially well with the true signal, as indicated by the drop in log-likelihood for signals with increasing mass ratio.

Biases in IMRPhenomD recovery increase with inclination and are most extreme for edge-on cases. Note that there are

also biases in the IMRPhenomHM recovery, but these are caused by the prior. The inclination prior has very low support from edge-on inclinations, and hence, the recovered  $\theta_{\text{JN}}$  posterior tends to have more support from non-edge-on inclinations. This leads to an overestimation of the distance. The amplitude ( $\mathcal{A}$ ) of a BBH source is  $\mathcal{A} \propto \mathcal{M}_c^{5/6}/d_L = M^{5/6}\sqrt{\eta}/d_L$ . At higher masses,  $\mathcal{M}_c$  and  $M_{\text{total}}$  are the better constrained mass parameters. Hence, overestimating  $d_L$  (with good constraints on  $\mathcal{M}_c$  and  $M_{\text{total}}$ ) would lead to a higher value of  $\eta$  or, equivalently, a lower  $q$ . This effect is what causes the slight bias on the IMRPhenomHM recovered  $q$  for edge-on  $q = 4$  and  $q = 8$  IMRPhenomHM injections.

Returning to the IMRPhenomD recovery biases, the most extreme case we see is that of the IMRPhenomD recovery of the  $q = 8$  edge-on IMRPhenomHM injection. The recovered posteriors show a bimodal distribution. For this injection, IMRPhenomD sees the signal as two completely different systems with parameters  $[(M_{\text{total}}, q, \chi_{\text{eff}}) \sim (85, 7, -0.25), (63, 11, -1.)]$ , with comparable (but overall very low) likelihood. Including the prior difference, the posterior values around those two areas in the parameter space are very similar, leading to the bimodality. As an additional test, we checked that the bimodal distribution is not due to sampling error by performing two additional nested-sampling PE runs with 2048 and 4096 live points and another MCMC run with 16 parallel chains and 5000 effective samples, but the bimodal distribution persists. Two additional PE runs were then done where the sky position of the signal was randomized while keeping the polarization fixed and vice versa. The bimodality seen by IMRPhenomD for the run in Fig. 2 is lost for these runs, but the corresponding parameters recovered by IMRPhenomD were (1) highly biased and (2) recovered with similar low maximum likelihood ( $\Delta\text{Log}(\mathcal{L}) \sim -95$ ). The bimodality of recovered parameters in Fig. 2 is a consequence of IMRPhenomD seeing the signal as from two different but equally likely systems, which is then lost when the signal morphology changes with changing sky position and polarization values. But, for all sky-position and polarization combinations, parameters recovered by IMRPhenomD for  $q = 8$  show a consistent bias toward lower total mass and negative  $\chi_{\text{eff}}$ . Also,  $\log(B_{\text{PhinD}}^{\text{PhnHM}})$  for the bimodal run is  $\approx 94$ , which implies that the signal as seen by IMRPhenomD is highly unlikely as compared to IMRPhenomHM. All this suggests that the observed bimodality is a combined effect of the priors over the physical parameters and the inaccuracy of IMRPhenomD toward recreating the true multimode signal.

$\mathcal{M}_c$  posteriors recovered by IMRPhenomHM are accurate for all the cases. At face-on  $q = 2$ , recovered  $q$  has large support from near-equal mass systems, but this behaviour is lost at higher inclinations. At edge-on  $q = 4$  and  $q = 8$ , mass ratio and  $M_{\text{total}}$  are slightly biased toward lower values, with accurate recovery of the mass-spin parameters for all other cases.

Where quadrupole models tend to recover a biased  $\chi_{\text{eff}}$  at higher inclinations, IMRPhenomHM recovery does not. For the  $q = 2$  injections with  $\theta_{\text{JN}} = 60^\circ$  and  $90^\circ$ , the recovered  $\chi_{\text{eff}}$  posteriors have almost the same width for IMRPhenomD and IMRPhenomHM templates ( $\Delta\chi_{\text{eff}} \sim 0$ ). Although the spread of these posteriors is similar,  $\chi_{\text{eff}}$  recovery with IMRPhenomHM is accurate, whereas IMRPhenomD recovery is biased.

We now consider the relative improvement in parameter precision (i.e., the widths of the posteriors, irrespective of any bias from the true injected values), as shown in Fig. 3.

At inclinations of  $60^\circ$  and  $90^\circ$ , the mass parameters recovered by IMRPhenomHM are always better constrained than corresponding IMRPhenomD recoveries (see middle and bottom panels of Fig. 3), i.e.,  $\Delta\lambda_i > 0$ . For a given inclination-parameter combination,  $\Delta\lambda_i$  increases with increasing  $q$ . For, e.g.,  $\theta_{\text{JN}} = 60^\circ$ ,  $\Delta\mathcal{M}_c \sim 20\%, 30\%, 50\%$  for  $q = 2, 4, \text{ and } 8$ , respectively. The comparatively high  $\Delta\lambda_i$  values for the edge-on  $q = 8$  configuration is due to the bimodality of the IMRPhenomD recovered posteriors. At  $q = 4$  and  $q = 8$ ,  $\Delta\chi_{\text{eff}} \geq 0$  for all inclinations, and IMRPhenomD recovers biased  $\chi_{\text{eff}}$  posteriors for non-face-on inclinations whereas IMRPhenomHM recovery is accurate for all configurations. Overall, we observe better constraints on the mass parameters for inclined system across the mass-ratio space.

The results in Fig. 3 illustrate that in addition to more accurate parameter recovery when using a multimode model, we also find improved precision in the parameter measurements. We see that other than for face-on configurations, the recovered mass and spin parameters are better constrained ( $\Delta\lambda_i > 0$ ), with the constraints improving with increasing mass ratio and/or inclination.

## B. Hybrid-NR injections

We now consider injections of the same physical systems, but using hybrid-NR waveforms instead of IMRPhenomHM. The purpose of this is to assess systematic errors in the IMRPhenomHM model. If the hybrid-NR waveforms and the corresponding IMRPhenomHM waveforms were almost identical, then the results from hybrid-NR injections would be nearly identical to those in the previous section. This would require not only that the IMRPhenomHM model accurately capture all of the features of the NR waveforms, but that the numerical errors in the higher-mode content of the NR waveforms be insignificant, along with the differences between the hybrids' EOB inspiral and the IMRPhenomHM inspiral. We do not expect any of these requirements to hold, and so with hybrid-NR injections we can determine which of the previous results is robust against uncertainties in the IMRPhenomHM model, and for which parameters and regions of the parameter space systematic errors may contaminate measurements.

The results of recovered intrinsic parameters for hybrid-NR waveform injections are given in Fig. 5. For these injections, we will not plot  $\Delta\lambda_i$ s, but will discuss them

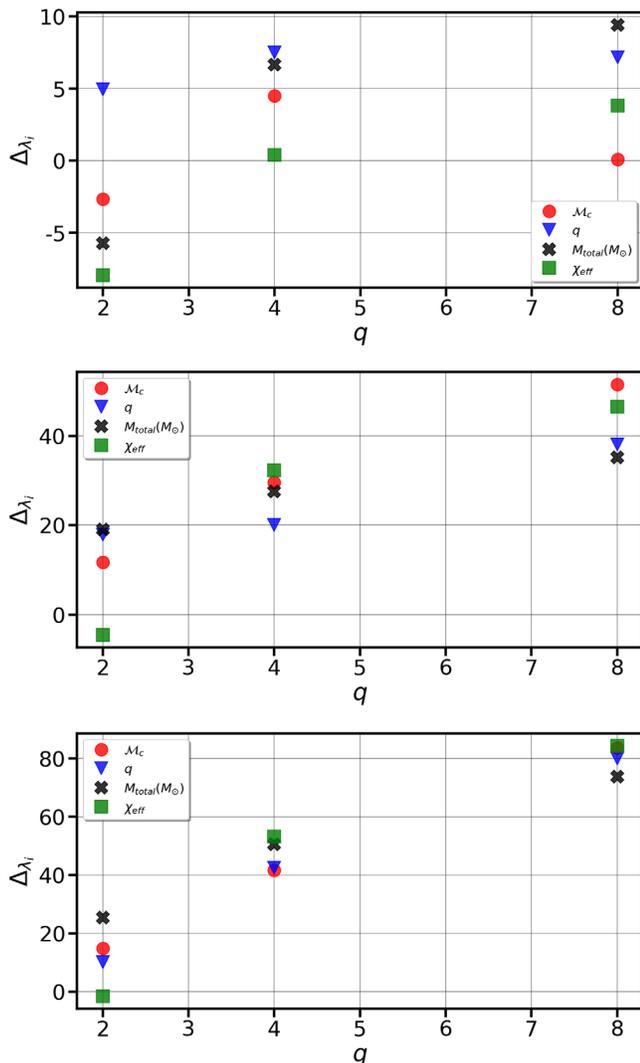


FIG. 3. Plot of  $\Delta\lambda_i$  for all the IMRPhenomHM injections. Results for face-on,  $60^\circ$ , and  $90^\circ$  inclination injections are shown in the top, middle, and bottom panels, respectively.  $\Delta\lambda_i$  for the parameters ( $\mathcal{M}_c$ ,  $M_{\text{total}}(M_\odot)$ ,  $q$ ,  $\chi_{\text{eff}}$ ) are shown with red circle, black cross, blue lower triangle, and green square, respectively.

when relevant. For  $q = 2$ ,  $q = 4$ , and  $q = 8$ , at face-on inclinations, posteriors recovered by IMRPhenomHM and IMRPhenomD follow the same behaviour that we saw for IMRPhenomHM injections. Parameters recovered by both models are accurate, but the posteriors recovered by IMRPhenomHM show slightly improved constraints on the mass and spin parameters ( $\Delta\lambda_i \geq 0$ ). Given the weak higher-mode content in face-on injections, it is not surprising that the results do not depend strongly on whether IMRPhenomHM or hybrid-NR waveforms were injected.

The same is true for  $q = 2$  injections at all inclinations. In some cases, the posteriors are wider for the hybrid-NR injections, the most notable example being the IMRPhenomHM recovery of  $M_{\text{total}}$  at  $\theta_{\text{JN}} = 60^\circ$ . However, it is not too surprising that recovery is more accurate and

precise when the injection was an instance of the waveform model used for recovery, as with the IMRPhenomHM injections. The lack of any notable impact of systematic errors at  $q = 2$  means that the bias in the quadrupole-only recovery of  $\chi_{\text{eff}}$  for the  $60^\circ$  injection is a robust result.

For  $q = 4$  and  $q = 8$  injections, we do see some clear differences between the results from IMRPhenomHM and hybrid-NR injections. To ease the interpretation of a large number of results, we first focus on the recovery using IMRPhenomHM, which is most relevant to future GW observations. Here we see that for most parameters the recovery again does not change significantly between the two classes of injections. There are two exceptions. One is the recovery of the mass ratio for  $q = 8$  edge-on systems, where we see that the prior effect, which leads to an underestimation of  $q$ , becomes yet more pronounced with the hybrid-NR injection, and the true value is outside the 90% CI. The other is the measurement of the total mass. The posteriors are broader in the  $q = 2$  case for both  $60^\circ$  and edge-on configurations, and for  $q = 4$  and  $q = 8$  the mass is clearly biased. However, we also note that these are cases where the quadrupole-only IMRPhenomD shows extremely large biases, and the IMRPhenomHM recovery shows a clear improvement.

As with the biases on mass parameters, biases on IMRPhenomD recovered  $\chi_{\text{eff}}$  for the hybrid-NR injections follow the same behavior as for IMRPhenomHM injections for  $q = 2$  and 4, but the bias for  $q = 8$  is in the opposite direction. However, IMRPhenomHM is able to measure the true value of  $\chi_{\text{eff}}$  within its 90% CIs for all the configurations.

If we now look at the quadrupole-only recovery, we see that there are many differences between the IMRPhenomHM and hybrid-NR injections. Some of these are counterintuitive: for example, for  $q = 8$ ,  $60^\circ$  injections, the bias in the chirp mass recovery is in opposite directions for the two classes of injections. However, we note that in all of these examples, the log-likelihood for the IMRPhenomD recovery is low, and so it is possible for the parameters of best agreement between IMRPhenomD and the injected waveform to show greater variation; and note also that these will also be sensitive to all of the intrinsic and extrinsic parameters.

Overall, we conclude that only in the most extreme cases ( $q = 8$  and edge-on) do we see a risk of biases when using IMRPhenomHM for parameter recovery, and in all cases it shows an improvement, and often a dramatic improvement, over a quadrupole-only model.

## V. RECOVERY OF EXTRINSIC PARAMETERS

We saw in Ref. [1] that one of the most significant impacts of using a higher-multipole model for parameter measurement is in the recovery of the binary's inclination ( $\theta_{\text{JN}}$ ) and distance ( $d_L$ ). In a quadrupole-only model, the only effect of changing the inclination is to change the overall amplitude, which is degenerate with a change in distance. The strength of the (2,2) mode varies by a factor of 2 between face-on

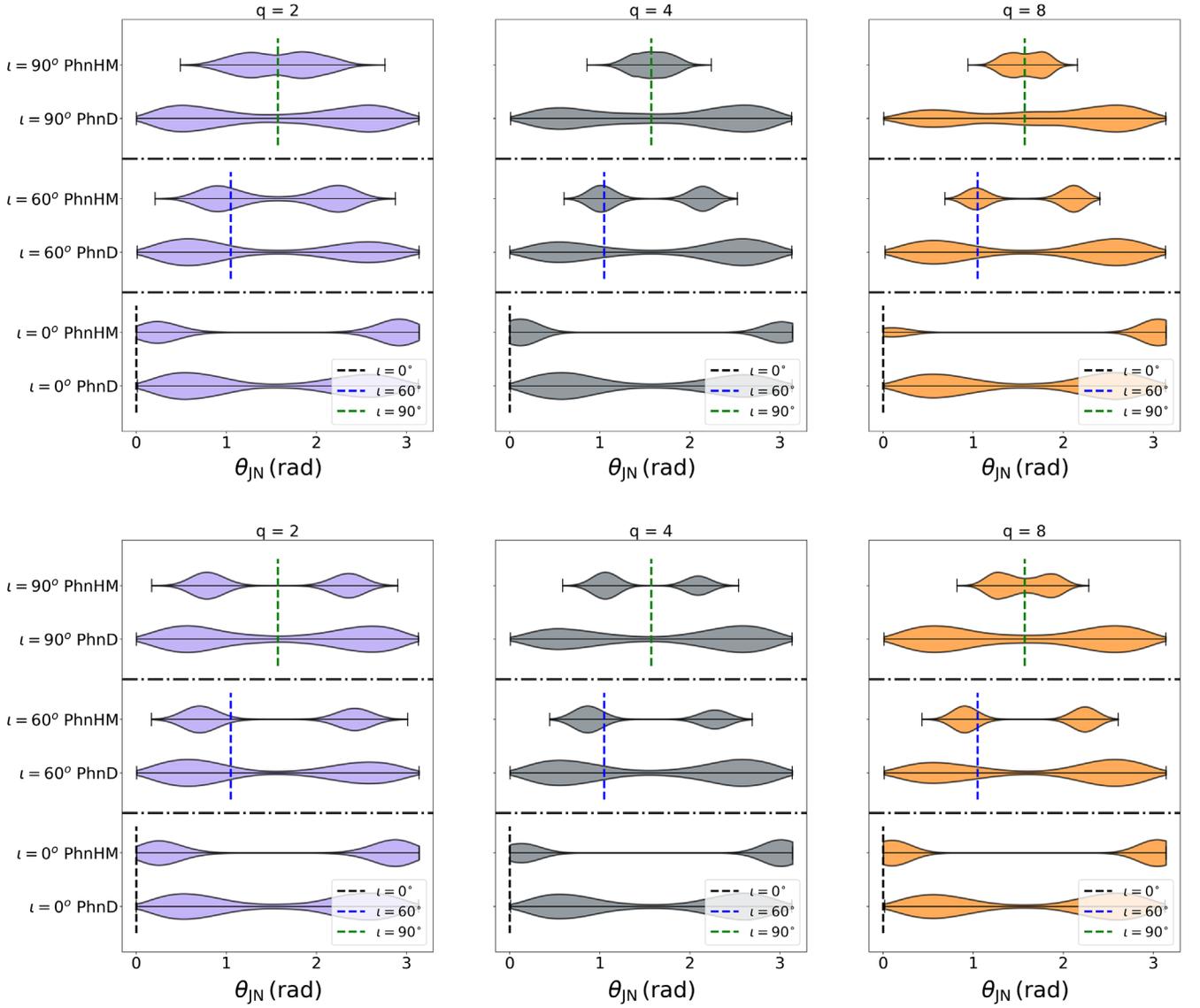


FIG. 4. Recovery of  $\theta_{\text{JN}}$  for IMRPhenomHM injection (top row) and hybrid-NR injections (bottom row) for inclinations  $0^\circ$ ,  $60^\circ$ , and  $90^\circ$  and with IMRPhenomHM and IMRPhenomD as recovery waveform models. Inclination recovery for  $q = 2, 4, 8$  configurations are shown in the left, centre, and right columns, respectively, with the recovery for each inclination separated by horizontal dashed black lines. The true value of the injections are given in dashed black, blue, and red lines for  $0^\circ$ ,  $60^\circ$ , and  $90^\circ$ .

and edge-on systems (for the plus polarization), leading to an uncertainty of roughly a factor of 2 in the distance measurement. The inclination measurement is then dominated by the prior, which is a combination of the inclination dependence of the detector sensitivity (the sensitivity is twice as sensitive to face-on systems) and the inclination probability distribution, which is uniform in  $\cos(\theta_{\text{JN}})$ . The result is a distribution that peaks at  $\sim 30^\circ$  and  $\sim 150^\circ$ , and this is reflected in the IMRPhenomD inclination recovery plots below. In general, as was seen in Refs. [36,37,56,57], the use of higher-order mode templates break the degeneracy present between  $\theta_{\text{JN}}$ ,  $\psi$ , and  $\phi$  that allows for better measurements of inclination value and hence better distance precision. We see that IMRPhenomHM is able to capture

inclination information better than IMRPhenomD and leads to improved constraints on the distance.

In the following sections, we quantify these effects for both IMRPhenomHM and hybrid-NR injections; the former quantify the impact of higher multipoles, while the latter allow us to investigate systematic biases due to approximations in IMRPhenomHM.

### A. $\theta_{\text{JN}}$ recovery

Figure 4 shows the results for inclination recovery for both the IMRPhenomHM and hybrid-NR injections. For both IMRPhenomHM and hybrid-NR injections, at all mass-ratio and inclination configurations,  $\theta_{\text{JN}}$  recovery by IMRPhenomD

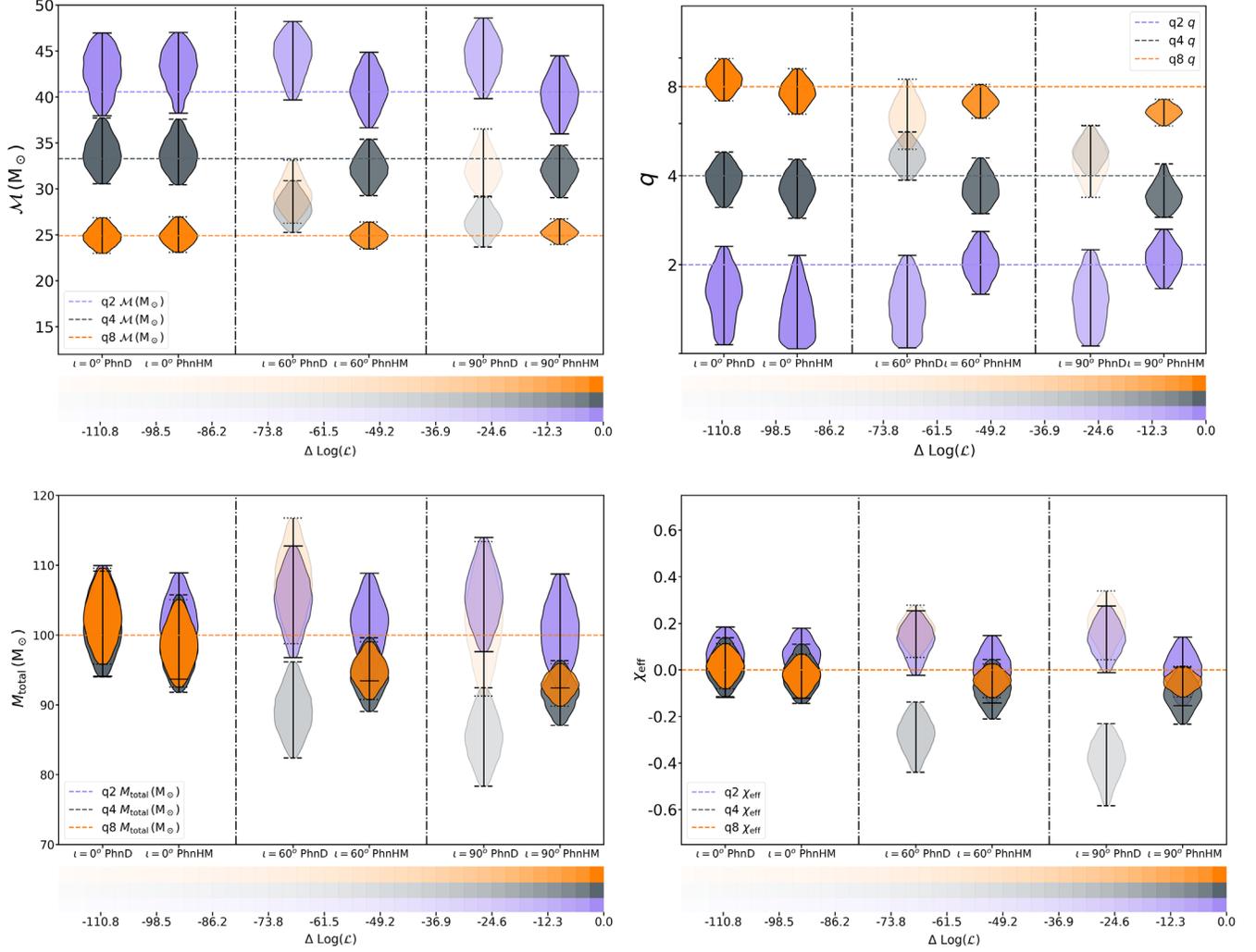


FIG. 5. Posteriors of intrinsic parameters ( $\mathcal{M}_c$ ,  $q$ ,  $M_{\text{total}}$ ,  $\chi_{\text{eff}}$ ) for hybrid-NR waveform injected at  $q = 2$ ,  $q = 4$ ,  $q = 8$  with  $\theta_{\text{JN}} = 0^\circ$ ,  $60^\circ$ , and  $90^\circ$ . Posteriors for  $q = 2$  ( $q = 4$ ) [ $q = 8$ ] are shown in blue (gray) (orange) with the opacity of each determined from the maximum likelihood value of that run. The variation of opacity over the likelihood values is shown at the bottom of each graph.

shows a similar bimodal behavior and mostly follows the prior, as discussed above. Inclination recovery is unaffected by mass ratio or inclination value for the quadrupole-only model and thus it is not possible to differentiate between a noninclined and inclined system.

For IMRPhenomHM recovery, the bimodality for the inclination posterior persists, but the posteriors are better constrained. At face-on configurations for IMRPhenomHM and hybrid-NR injections, IMRPhenomHM sees the system as strongly face-on or face-off. For  $60^\circ$  IMRPhenomHM injections, the recovered inclination is peaked near the true value and the constraint on the inclination improves with increasing mass ratio. The edge-on IMRPhenomHM injection posteriors show a similar behavior. For hybrid-NR injections, inclination recovery for  $60^\circ$  is peaked just off the true value and for edge-on  $q = (2, 4)$  systems, the recovery is strongly biased. This is not surprising: the systematic errors in the IMRPhenomHM model enter almost entirely into the higher

multipoles, and so the largest systematic errors will be observed for edge-on systems, where the higher multipoles contribute most to the signal. We find, however, that it is *only* for the edge-on cases that this bias appears. Since edge-on systems are still roughly half as strong as equivalent face-on systems, they are eight times less likely to be observed.

## B. Distance recovery

Recall that injections were made such that the signal's SNR was 25 in all cases. Since the signal strength decreases as mass ratio increases, and also as the inclination varies from face-on to edge-on, the edge-on  $q = 8$  system is injected at a much smaller distance (199 Mpc) than the face-on  $q = 2$  system (895 Mpc). Although there is some variation in the injection distance between the IMRPhenomHM and hybrid-NR injections, due to the differences in their higher-multipole content, these are small and are always

less than 10%. All of the injection distances are given in the caption to Fig. 6.

Figure 6 shows the results for distance recovery. We plot the relative percent distance error, which we define as

$$\Delta d_L = 100 \left( \frac{p(d_L^{\text{posterior}}) - d_L^{\text{injected}}}{d_L^{\text{injected}}} \right). \quad (13)$$

For IMRPhenomHM injections, the true distance value lies within the 90% confidence intervals for most of IMRPhenomD and all of IMRPhenomHM recovered posteriors. At larger inclinations, the quadrupole model tends to overestimate the distance to the binary. For  $q = 2$  hybrid-NR injections, at inclinations  $60^\circ$  and  $90^\circ$ , 90% CIs for  $d_L$  recovered by IMRPhenomHM do not include the true value. For all other

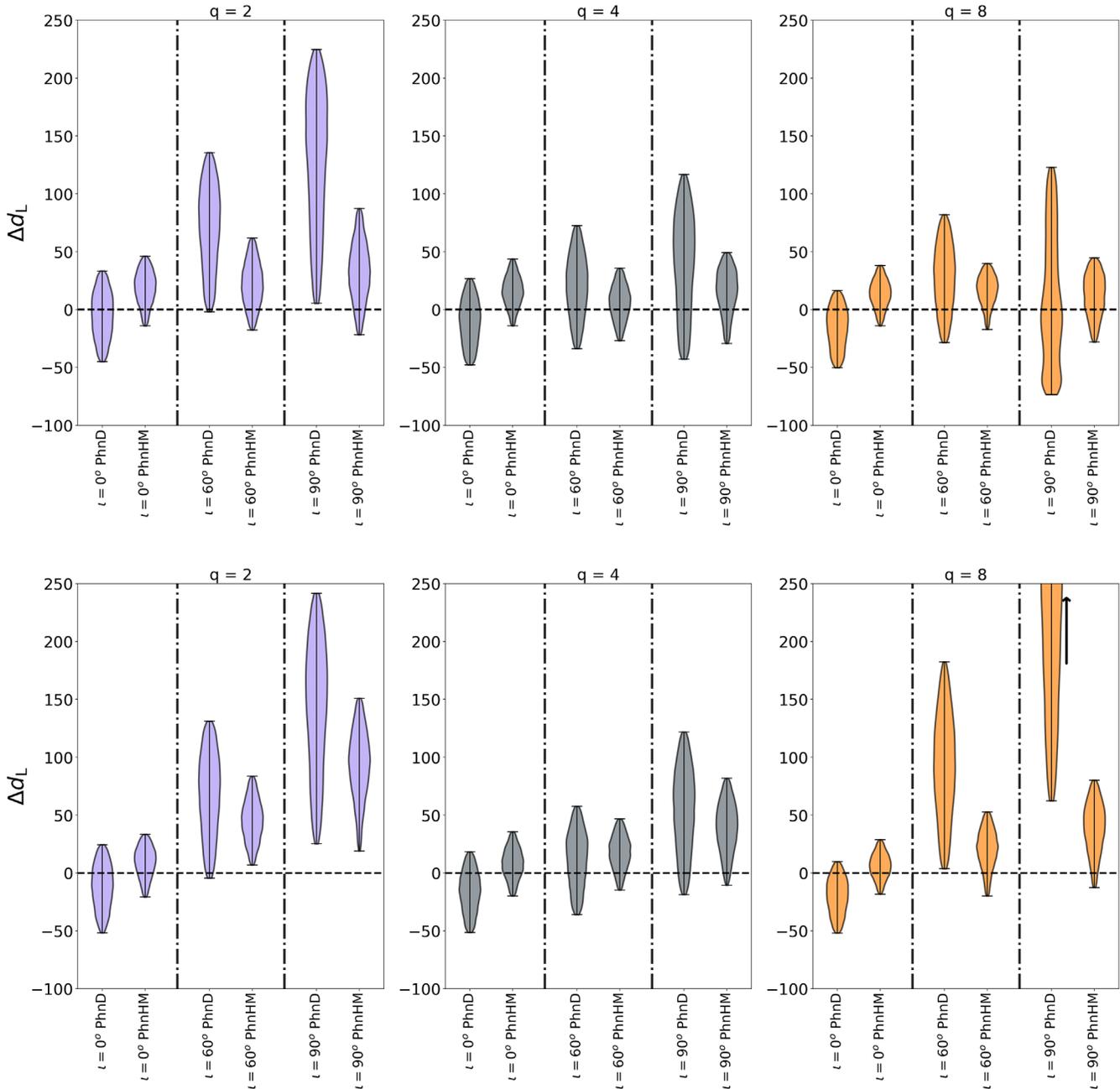


FIG. 6. Recovered percent distance error  $\Delta d_L$  [see Eq (13)] recovery for IMRPhenomHM injection (top row) and hybrid-NR injections (bottom row) for inclinations  $0^\circ$ ,  $60^\circ$ , and  $90^\circ$  with IMRPhenomHM and IMRPhenomD as recovery waveform models. Distance recovery for  $q = 2, 4, 8$  configurations are shown in the left, center, and right columns, respectively. The  $\Delta d_L = 0$  line is shown in horizontal dashed black line, with the vertical dashed lines separating recovery for different inclinations. The injected distance value for IMRPhenomHM injection for  $q = (2, 4, 8)$   $0^\circ$  is (895, 624, 388) [880, 639, 398] Mpc,  $60^\circ$  is (537, 404, 258) [523, 376, 249] Mpc, and for  $90^\circ$  is (387, 307, 199) [367, 253, 183] Mpc. For  $q = 8, 90^\circ$ , hybrid-NR recovery by IMRPhenomD,  $\Delta d_L$  extends up to 400%.

situations though, 90% CIs for  $d_L$  recovered by IMRPhenomHM contain the true injected value.

For quadrupole-only templates, as the recovered inclination is the same for all injected inclinations, the recovered distance for nonzero inclinations tend toward overall larger values leading to larger  $\Delta d_L$ . For IMRPhenomHM injection—IMRPhenomHM recovery, where injected  $\theta_{\text{IN}}$  lies within the 90% CIs of recovered  $\theta_{\text{IN}}$ , the real distance is recovered at all times. For the  $q = 2$  and  $q = 4$  hybrid-NR injection—IMRPhenomHM template, recovered  $\theta_{\text{IN}}$  at  $60^\circ$  ( $90^\circ$ ), is slightly (completely) off the true value which causes the recovered distance to be overestimated from the true value. This is likely due to the different mode content in the signal and template and the waveform inaccuracies in IMRPhenomHM. But, these results do indicate that use of multimode template waveform will lead to better distance measurements.

The improved constraints on inclination for IMRPhenomHM recovery translate to improved constraints on the measured distance of the system as compared to IMRPhenomD recovery. We see this behavior for all the configurations. At face-on configurations, IMRPhenomHM constrains the distance about  $\sim 20\%$ – $25\%$  better as compared to IMRPhenomD. For higher inclinations, the constraint improves by about  $\sim 30\%$ – $60\%$ .

## VI. CONCLUSIONS

This is the first study that quantifies the accuracy of inferred source parameters using a multimode aligned-spin model waveform. To do that, we consider two families of nonspinning multimode signal waveforms (IMRPhenomHM and hybrid-NR) over a range of mass ratios ( $q = 2, 4, 8$ ) and inclinations (face-on,  $60^\circ$ , and edge-on), with fixed total mass, and compare the parameters recovered by multimode and quadrupole-only templates. In all cases, we consider an SNR of 25. We fix the total mass of injected signals at  $100 M_\odot$ ; for three reasons (i) the relative measurable signal power in the higher modes increases with mass, and so the choice of a high mass allows us to provide an estimate of the largest impact of higher modes in likely LIGO-Virgo observations, for which total binary masses above  $100 M_\odot$  will be rare. (ii) IMRPhenomHM is an approximate waveform model of the subdominant modes that is not tuned to NR waveforms, and the most uncertain part of the IMRPhenomHM modeling is in the merger and ringdown phases. Hence, the choice of a high total mass also allows us to also make a conservative estimate of the systematic errors due to waveform inaccuracies. (iii) IMRPhenomHM is a more computationally expensive model than its quadrupole-only counterpart, and signals with total mass  $100 M_\odot$  allow us to quickly perform a large PE study. Although, as previous studies have shown [33–35], systematic errors due to neglecting higher-order modes in template waveforms increase at higher masses. Optimized versions of the model will make it easier to

perform a much more extensive study over a wider range of parameters, including much lower masses.

Here is a summary of our main results.

Our key results on measurements of intrinsic parameters are in Figs. 2 and 5. For face-on systems, there is no appreciable bias in the quadrupole-only recovery at any mass ratio. There are biases in the total mass of up to 10% at  $60^\circ$  inclination and 20% or 30% for edge-on configurations. For the edge-on  $q = 8$  configuration (IMRPhenomHM injection), the signal matches the quadrupole-only model so poorly that the recovery is bimodal. Mass ratio is strongly biased toward equal-mass recovery for small mass ratios, but shows less bias at higher mass ratios. (This point is relevant to the observation GW170729 [8], where the templates with higher-order modes were able to resolve  $q$  between 1.25 and 3.3 at 90% confidence interval, i.e., providing strong evidence that the mass ratio was bounded away from equal mass, while quadrupole-only models gave a 90% CI from  $q = 1$  up to  $q = 2.5$ .) The effective spin  $\chi_{\text{eff}}$  shows a bias of up to 0.2 for  $60^\circ$  inclination and cannot be measured at all in some edge-on cases. Our results suggest that bimodal parameters or, in less extreme cases, double-peaked posteriors, might occur (for some cases) when measuring high-mass-ratio systems with a quadrupole-only model which could be resolved with multimode models. The overall lower likelihood of recovered parameters by the quadrupole-only model (compared to multimode model) at high-mass-ratio high-inclination combination suggests that the model cannot be trusted for accuracy in that region.

For IMRPhenomHM injections, recovery with a higher-mode model not only removes these biases (as we would expect, since the injection and recovery use the same model), but also increases the precision of the measurement. Figure 3 shows the percentage improvement in the size of the 90% credible intervals over using a quadrupole-only model. The improvement is up to 50% for  $60^\circ$  inclination and up to 80% for edge-on configurations; the improvement is roughly linear with mass ratio.<sup>1</sup> Improvement in parameter recovery was also considered in Ref. [37], but using a higher-mode model that did not include spin. We find that the addition of the spin dimension can significantly increase the widths of the confidence intervals. For example, recovering a  $q = 4$ ,  $M = 100 M_\odot$ ,  $60^\circ$  inclination SNR = 18 signals with a nonspinning model leads to an uncertainty in the chirp mass of  $\Delta \mathcal{M}_{\text{obs}} / \mathcal{M}_{\text{obs}} = 0.056$  [37]. If we naively rescale to an SNR of 25, the uncertainty would decrease to  $\sim 0.04$ . By contrast, the uncertainty when using a spinning higher-mode model is 0.168, i.e., four times larger.

NR-hybrid injections show broadly consistent results, indicating that the systematic errors in the IMRPhenomHM model are in general small. The IMRPhenomHM recovery

<sup>1</sup>Note that improvement of 80% for  $q = 8$  edge-on is due to the bimodal recovery of quadrupole-only model.

gives comparable results between the IMRPhenomHM and NR-hybrid injections, except for biases of up to 5% in the total mass for  $q = 4$  and  $q = 8$  non-face-on cases; this is less than the bias incurred by using a quadrupole-only model. Results show larger differences in the IMRPhenomD recovery between the two injection sets, but these are cases where the log-likelihood is poorer, and so we ascribe these less significance. Our conclusion is that the IMRPhenomHM model leads to improved parameter measurements over a quadrupole-only model in *all* cases; and except for high-mass high-mass-ratio high-inclination signals with an SNR of 25 or higher, systematic errors will not affect results.

This is quantified further in Fig. 7. Here we follow Ref. [36], and plot the ratio of the systematic error to the statistical error,  $\delta\beta_{\lambda^i}$ , for the parameter  $\lambda^i$ . By “systematic error” we mean the difference between the true injected parameter and the mean of the marginalized one-dimensional (1D) recovered posterior (in Ref. [36], the authors use the difference between the maximum *a posteriori* value and injected value), and by “statistical error” we mean the standard deviation of the posterior. Since the standard deviation corresponds to the 68% C.I.,  $\delta\beta_{\lambda^i}$  is a more conservative estimate than if we had used the 90% CI.

As such, if the ratio is below  $\sim 1.645$ , then the systematic error is within the 90% CI and the measurement is not considered to be biased. Figure 7 shows this error ratio for the chirp mass, total mass, mass ratio, and effective spin. Also shown are the results from Ref. [36], although care should be taken in comparing the results, since that study considered lower masses, higher SNRs, and different mode content for the injections.

Inclination recovery is always improved when recovering with a higher-mode model. With quadrupole-only recovery, the distance-inclination degeneracy means that largely the same posterior is recovered, regardless of inclination (see Fig. 4), while the higher-mode model is able to constrain the inclination. The trend in the IMRPhenomD distance recovery is consistent with expectations: for a signal with a high nonzero inclination, a quadrupole-only template model gets more support from nonedge on inclinations, suggesting that the (comparatively) weak signal is from further away and therefore  $d_L$  would be overestimated. A multimode template model can better constrain the degeneracy between the inclination, phase, and polarization values, leading to improved constraints on the inclination, which then translates to a better

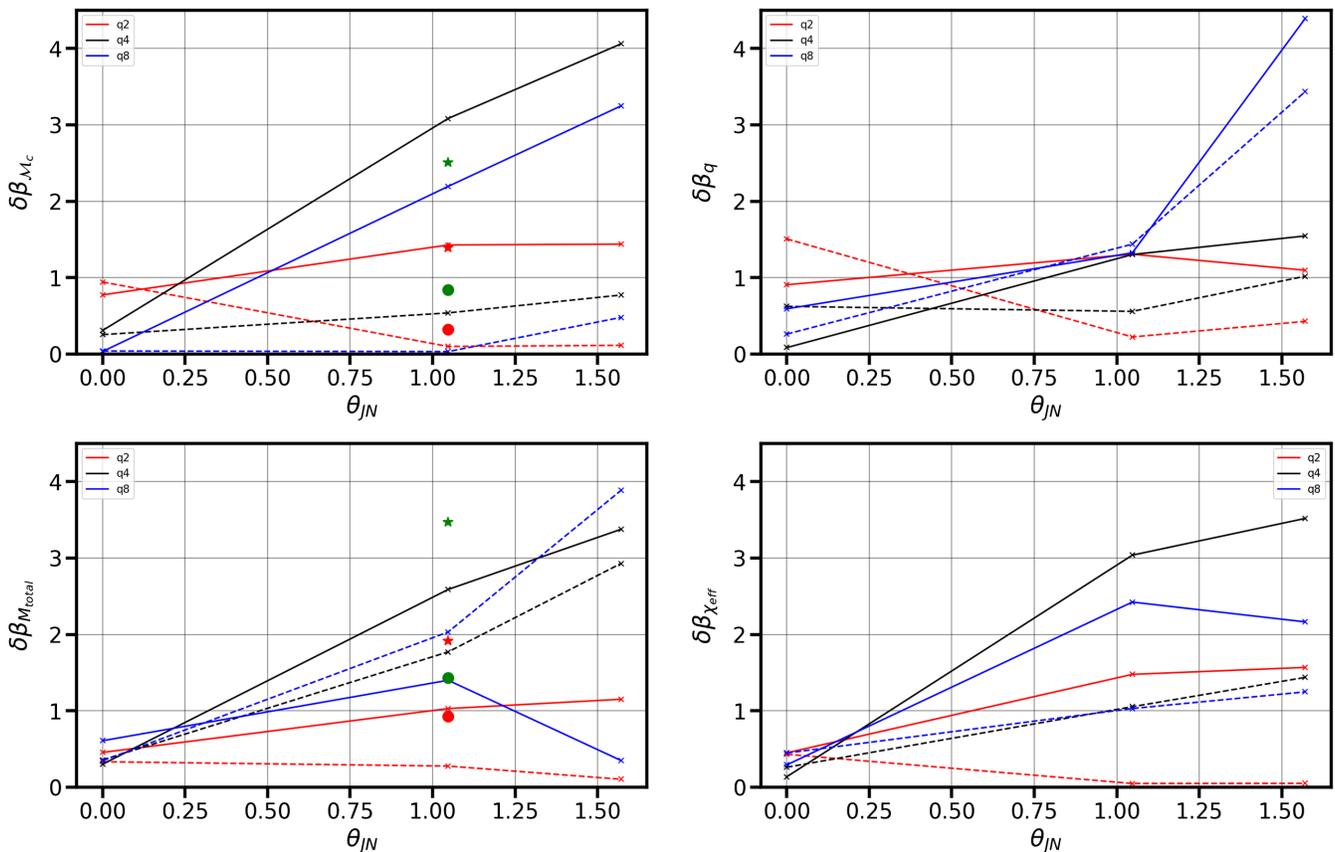


FIG. 7. We plot the quantity  $\delta\beta_{\lambda^i}$  for the parameters ( $M_c$ ,  $M_{\text{total}}$ ,  $q$ ,  $\chi_{\text{eff}}$ ) for hybrid-NR injection results with the solid (dashed) lines indicating the bias value for IMRPhenomD (IMRPhenomHM) recovery.  $\delta\beta_{\lambda^i}$  for  $q = 2, 4, 8$  are shown in red, black, and blue, respectively. The systematic bias for the  $q = 2, 6$ ,  $M/M_{\odot} = 51, 56$  and SNR = 48 configurations from Ref. [36] are shown in red (green) with the quadrupole (multimode) recovered bias shown with a star (circle).

constrained measurement of distance. Distance recovery is greatly improved with the higher-mode model. At face-on configurations, IMRPhenomHM constrains the distance about 20%–25% better as compared to IMRPhenomD. For higher inclinations, the constraint improves by about 30%–60%. With improved multimode models, we can expect improved inclination constrains and hence, distance measurements, with additional improvement from those three-detector observations that have good polarization measurements [53,61].

Systematic errors in IMRPhenomHM are worst for edge-on higher-mass-ratio systems, where the approximations used to produce the higher modes are least applicable, and this shows up in the inclination recovery: we do not identify the system as so clearly edge-on in the NR-hybrid injections compared to the IMRPhenomHM injections, as seen in the top panels of Fig. 4. As it is more likely that systematics due to the model inaccuracies dominate at larger inclinations, we can expect accurate parameter recovery at lower inclinations.

As expected, priors can bias results for statistically less likely configurations, i.e., edge-on, even when using a higher-mode model. See, for example, the edge-on recovery of the mass ratio  $q$  for the  $q = 8$  case in Fig. 2.

We note that this study is limited to binaries with total mass  $M = 100 M_{\odot}$  and does not include spinning signals, or precession. The effect of black-hole spin on higher-mode contributions is much weaker than the effect of mass ratio, and given that LIGO-Virgo observations to date suggest that astrophysical black holes in binaries predominantly have very low spins [6,62], we expect the results of this study to be relevant to the majority of observations made with second-generation detectors. Since the impact of higher modes decreases for lower masses, the results we report here are likely to represent the largest impact higher modes will have on GW observations. However, studies

that include spinning configurations, lower masses, and precession, are still needed, to quantify the impact of higher modes for yet stronger signals, or larger spins, where their effects have not yet been quantified. For this study, we had fixed the azimuth phase to zero and as was shown in [33], this parameter can strongly affect parameter estimates. It would be interesting to perform a similar study, but with different azimuth phase values. Studies that consider yet higher SNRs will also benefit from more accurate models tuned to NR simulations, although we expect that the IMRPhenomHM model will be sufficiently accurate for all but the most extreme observations.

## ACKNOWLEDGMENTS

We thank Alessandro Nagar for providing the EOB code used in generating the hybrids and Edward Fauchon-Jones for an early version of the hybridization code. We also thank Lionel London, Sebastian Khan, and Frank Ohme for useful discussions. This work was supported by Science and Technology Facilities Council (STFC) Grant No. ST/L000962/1, European Research Council Consolidator Grant No. 647839, and the National Science Foundation under Grant No. NSF PHY-1748958. We are grateful for computational resources provided by Cardiff University and funded by an STFC grant supporting UK Involvement in the Operation of Advanced LIGO. This research has made use of data, software, and/or web tools obtained from the Gravitational Wave Open Science Centre (<https://www.gw-openscience.org>), a service of LIGO Laboratory, the LIGO Scientific Collaboration, and the Virgo Collaboration. LIGO is funded by the U.S. National Science Foundation. Virgo is funded by the French Centre National de Recherche Scientifique, the Italian Istituto Nazionale della Fisica Nucleare, and the Dutch Nikhef, with contributions by Polish and Hungarian institutes.

- 
- [1] L. London, S. Khan, E. Fauchon-Jones, C. García, M. Hannam, S. Husa, X. Jiménez-Forteza, C. Kalaghatgi, F. Ohme, and F. Pannarale, *Phys. Rev. Lett.* **120**, 161102 (2018).
  - [2] B. P. Abbott *et al.*, *Phys. Rev. Lett.* **116**, 061102 (2016).
  - [3] J. Aasi *et al.* (LIGO Scientific Collaboration), *Classical Quantum Gravity* **32**, 115012 (2015).
  - [4] F. Acernese *et al.* (Virgo Collaboration), *Classical Quantum Gravity* **32**, 024001 (2015).
  - [5] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **119**, 141101 (2017).
  - [6] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. X* **9**, 031040 (2019).
  - [7] B. Abbott, R. Abbott, T. Abbott, F. Acernese *et al.*, *Phys. Rev. Lett.* **119**, 161101 (2017).
  - [8] K. Chatziioannou *et al.*, *Phys. Rev. D* **100**, 104015 (2019).
  - [9] P. Schmidt, M. Hannam, S. Husa, and P. Ajith, *Phys. Rev. D* **84**, 024046 (2011).
  - [10] R. O’Shaughnessy, B. Vaishnav, J. Healy, Z. Meeks, and D. Shoemaker, *Phys. Rev. D* **84**, 124002 (2011).
  - [11] M. Boyle, R. Owen, and H. P. Pfeiffer, *Phys. Rev. D* **84**, 124011 (2011).
  - [12] E. Poisson and C. M. Will, *Phys. Rev. D* **52**, 848 (1995).
  - [13] C. Cutler and E. E. Flanagan, *Phys. Rev. D* **49**, 2658 (1994).
  - [14] P. Ajith *et al.*, *Phys. Rev. Lett.* **106**, 241101 (2011).
  - [15] E. Baird, S. Fairhurst, M. Hannam, and P. Murphy, *Phys. Rev. D* **87**, 024035 (2013).
  - [16] A. Buonanno, Y.-b. Chen, and M. Vallisneri, *Phys. Rev. D* **67**, 104025 (2003); **74**, 029904(E) (2006).

- [17] P. Schmidt, M. Hannam, and S. Husa, *Phys. Rev. D* **86**, 104063 (2012).
- [18] D. A. Brown, A. Lundgren, and R. O’Shaughnessy, *Phys. Rev. D* **86**, 064020 (2012).
- [19] S. Fairhurst, R. Green, C. Hoy, M. Hannam, and A. Muir, [arXiv:1908.05707](https://arxiv.org/abs/1908.05707).
- [20] 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).
- [21] A. Nagar, G. Pratten, G. Riemenschneider, and R. Gamba, *Phys. Rev. D* **101**, 024041 (2020).
- [22] A. K. Mehta, C. K. Mishra, V. Varma, and P. Ajith, *Phys. Rev. D* **96**, 124010 (2017).
- [23] A. K. Mehta, P. Tiwari, N. K. Johnson-McDaniel, C. K. Mishra, V. Varma, and P. Ajith, *Phys. Rev. D* **100**, 024032 (2019).
- [24] R. Cotesta, A. Buonanno, A. Bohé, A. Taracchini, I. Hinder, and S. Ossokine, *Phys. Rev. D* **98**, 084028 (2018).
- [25] J. Blackman, S. E. Field, C. R. Galley, B. Szilágyi, M. A. Scheel, M. Tiglio, and D. A. Hemberger, *Phys. Rev. Lett.* **115**, 121102 (2015).
- [26] 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).
- [27] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, C. D. Ott, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and B. Szilágyi, *Phys. Rev. D* **96**, 024058 (2017).
- [28] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, *Phys. Rev. Research* **1**, 033015 (2019).
- [29] C. Capano, Y. Pan, and A. Buonanno, *Phys. Rev. D* **89**, 102003 (2014).
- [30] J. C. Bustillo, P. Laguna, and D. Shoemaker, *Phys. Rev. D* **95**, 104038 (2017).
- [31] I. Harry, J. C. Bustillo, and A. Nitz, *Phys. Rev. D* **97**, 023004 (2018).
- [32] J. C. Bustillo, F. Salemi, T. Dal Canton, and K. P. Jani, *Phys. Rev. D* **97**, 024016 (2018).
- [33] J. C. Bustillo, S. Husa, A. M. Sintes, and M. Pürrer, *Phys. Rev. D* **93**, 084019 (2016).
- [34] V. Varma and P. Ajith, *Phys. Rev. D* **96**, 124024 (2017).
- [35] V. Varma, P. Ajith, S. Husa, J. C. Bustillo, M. Hannam, and M. Pürrer, *Phys. Rev. D* **90**, 124004 (2014).
- [36] T. B. Littenberg, J. G. Baker, A. Buonanno, and B. J. Kelly, *Phys. Rev. D* **87**, 104003 (2013).
- [37] P. B. Graff, A. Buonanno, and B. S. Sathyaprakash, *Phys. Rev. D* **92**, 022002 (2015).
- [38] J. Veitch *et al.*, *Phys. Rev. D* **91**, 042003 (2015).
- [39] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Classical Quantum Gravity* **34**, 104002 (2017).
- [40] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044006 (2016).
- [41] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044007 (2016).
- [42] A. Bohé *et al.*, *Phys. Rev. D* **95**, 044028 (2017).
- [43] J. C. Bustillo, A. Bohé, S. Husa, A. M. Sintes, M. Hannam, and M. Pürrer, [arXiv:1501.00918](https://arxiv.org/abs/1501.00918).
- [44] D. J. A. McKechnan, C. Robinson, and B. S. Sathyaprakash, *Classical Quantum Gravity* **27**, 084020 (2010).
- [45] M. Boyle *et al.*, *Classical Quantum Gravity* **36**, 195006 (2019).
- [46] T. Damour and A. Nagar, *Phys. Rev. D* **90**, 044018 (2014).
- [47] A. Nagar, G. Riemenschneider, and G. Pratten, *Phys. Rev. D* **96**, 084045 (2017).
- [48] A. Nagar, T. Damour, C. Reisswig, and D. Pollney, *Phys. Rev. D* **93**, 044046 (2016).
- [49] T. B. Littenberg and N. J. Cornish, *Phys. Rev. D* **91**, 084034 (2015).
- [50] N. J. Cornish and T. B. Littenberg, *Classical Quantum Gravity* **32**, 135012 (2015).
- [51] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, *J. Phys. Conf. Ser.* **610**, 012021 (2015).
- [52] P. Schmidt, I. W. Harry, and H. P. Pfeiffer, [arXiv:1703.01076](https://arxiv.org/abs/1703.01076).
- [53] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, and Virgo), *Living Rev. Relativity* **21**, 3 (2018).
- [54] A. Nitz *et al.*, [gwastro/pycbc: Pycbc release v1.14.1](https://arxiv.org/abs/1908.05707) (2019).
- [55] C. Van Den Broeck and A. S. Sengupta, *Classical Quantum Gravity* **24**, 1089 (2007).
- [56] R. O’Shaughnessy, B. Farr, E. Ochsner, H.-S. Cho, C. Kim, and C.-H. Lee, *Phys. Rev. D* **89**, 064048 (2014).
- [57] R. O’Shaughnessy, B. Farr, E. Ochsner, H.-S. Cho, V. Raymond, C. Kim, and C.-H. Lee, *Phys. Rev. D* **89**, 102005 (2014).
- [58] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, *Phys. Rev. Lett.* **113**, 151101 (2014).
- [59] Y. Pan, A. Buonanno, A. Taracchini, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi, *Phys. Rev. D* **89**, 084006 (2014).
- [60] S. Fairhurst, R. Green, M. Hannam, and C. Hoy, [arXiv:1908.00555](https://arxiv.org/abs/1908.00555).
- [61] S. Fairhurst, *Classical Quantum Gravity* **28**, 105021 (2011).
- [62] V. Tiwari, S. Fairhurst, and M. Hannam, *Astrophys. J.* **868**, 140 (2018).