Constraining the parameters of GW150914 and GW170104 with numerical relativity surrogates

Prayush Kumar, Jonathan Blackman, Scott E. Field, Mark Scheel, Chad R. Galley, Michael Boyle, Lawrence E. Kidder, Harald P. Pfeiffer, Bela Szilagyi, and Saul A. Teukolsky

Phys. Rev. D **99**, 124005 — Published 7 June 2019

DOI: 10.1103/PhysRevD.99.124005

Constraining the parameters of GW150914 & GW170104 with numerical relativity surrogates

Prayush Kumar,[1,*] Jonathan Blackman,[2] Scott E. Field,[3] Mark Scheel,[2] Chad R. Galley,[4,2] Michael Boyle,[1] Lawrence E. Kidder,[1] Harald P. Pfeiffer,[5] Bela Szilagyi,[2,4] and Saul A. Teukolsky[1,2]

[1] *Cornell Center for Astrophysics and Planetary Science, Cornell University, Ithaca, New York 14853, USA*
[2] *Theoretical Astrophysics, Walter Burke Institute for Theoretical Physics, MC 350-17, California Institute of Technology, Pasadena, CA 91125, USA*
[3] *Department of Mathematics, University of Massachusetts, Dartmouth, MA 02747, USA*
[4] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA*
[5] *Albert Einstein Institute, Am Mühlenberg, Golm, Germany*

(Dated: April 10, 2019)

Gravitational-wave detectors have begun to observe coalescences of heavy black hole binaries ($M \gtrsim 50M_\odot$) at a consistent pace for the past few years. Accurate models of gravitational waveforms are essential for unbiased and precise estimation of the source parameters, such as masses and spins of component black holes. Recently developed surrogate models based on high-accuracy numerical relativity simulations provide ideal models for constraining physical parameters describing these heavy black hole merger events. In this paper we demonstrate the viability of these surrogate models as reliable parameter estimation tools. We show that within a fully Bayesian framework, surrogates can help us extract more information from gravitational wave observations than traditional models. We demonstrate this by analyzing a set of synthetic signals with numerical relativity surrogates and comparing those results against approximate models. We then consider the case of two of the earliest binary black holes detected by the LIGO observatories, GW150914 and GW170104. We reanalyze their data with the generically precessing numerical-relativity-based surrogate model. For these systems we find that our refined analysis estimates the sources of both events to be $20-25\%$ further away than previously estimated. Our analysis also constrains their orientation more tightly around face-on or face-off configurations than in the past. Additionally, for GW150914 we constrain the effective spin parameter more tightly around zero. For all other source parameters, inferences with surrogate templates were quantitatively consistent with those with conventional models. This work is a first step toward eliminating the approximations used in semi-analytic waveform models from GW parameter estimation. It strongly motivates that numerical relativity surrogates be extended to cover more of the binary black hole parameter space.

PACS numbers: 04.30.-w, 04.30.Db

I. INTRODUCTION

General Relativity (GR) predicts that accelerated massive bodies emit energy in the form of gravitational waves (GWs). In 2015, the first direct detection of GWs coming from coalescing binary black holes (BBHs) was made by the LIGO observatories [1]. Since then, many more GW signals from BBHs have been observed by the LIGO and Virgo detectors [2–5], ushering us into the era of GW astronomy. GW searches for BBH signals [6–8], the process of estimating their source properties [9], as well as that of testing GR with them [10] rely heavily on the technique of matched filtering, which tacitly assumes the availability of GW signal models for BBHs.

For heavy black hole binaries (with masses $\gtrsim 50M_\odot$), such as those that have dominated the event rates of LIGO-Virgo detectors so far [3, 11], a large fraction of the observable signal consists of the last few tens of orbits prior to the binary's merger. In this regime, the dynamical effects of GR are substantial, making analytic treatment difficult. Instead, numerical solutions of Einstein's equations [12–17] must be used.

The inspiral of a binary system of black holes along a quasi-circular trajectory, and their subsequent merger and ringdown, is completely describable by 8 parameters - the masses of both holes, and their spin vectors. Conventional parameter estimation (PE) algorithms search through this parameter space to estimate parameters that best describe the signal embedded in LIGO-Virgo data. In the process, they can take a large number ($\mathcal{O}(10^6)$) of steps, each requiring evaluation of a new waveform. Although we have the technical capability to perform full numerical relativity (NR) simulations over a good fraction of the multi-dimensional parameter space, each simulation still takes a large amount of computing and human time. Therefore, it has remained impractical to use NR simulations directly with conventional PE methods for estimating physical parameters of BBHs. There have been two possible alternatives that have been utilized in the past: (a) using phenomenological waveform models containing free parameters that are tuned to a (relatively) small number of NR simulations [18, 19]; and (b) using grid-based parameter estimation methods [20] with NR templates. While phenomeno-

* prayush@astro.cornell.edu

logical waveform models SEOBNR and IMRPhenomPv2 have been used extensively in previous LIGO-Virgo publications [1, 11, 21], they still have many shortcomings. On one hand, IMRPhenomPv2 uses a post-Newtonian theory based waveform amplitude prescription, and captures BH spin effects using only 2 (of total 6) spin degrees of freedom [22]; while on the other hand, only aligned-spin SEOBNR models are computationally inexpensive enough to be used in PE analyses [23]. Moreover, neither of them presently have precession dynamics near merger fitted against NR simulations, and have been shown to break down close to the parameter-space boundary of their calibration domain [24, 25] or even within it [26]. Grid-based PE methods (alternative (b)) have been recently applied to GW observations [27, 28]; however, they have so far only demonstrated the ability to constrain a subset of physical parameters because of the sparse parameter space coverage of available NR simulations.

A novel alternative arises from the development of surrogate models [29, 30] for numerical relativity waveforms [31]. Such data-driven models are constructed over a training set of specially selected NR simulations. The waveforms from these simulations are then "interpolated" in parameter space. The resulting NR surrogate model is able to quickly generate new waveforms at arbitrary points within the training region with the largest surrogate model errors typically comparable to the largest errors in the numerical relativity simulations. After a couple of simpler versions [31, 32], Blackman et al [33] published a surrogate model NRSur7dq2 for generically spinning-precessing binaries. This NRSur7dq2 model was developed using 744 new simulations [31] spanning a range of the 7-dimensional space[1] bounded in mass ratio $q \leq 2$, and BH spin magnitudes $|\vec{\chi}_{1,2}| \leq 0.8$. It provides all $\ell \leq 4$ waveform multipoles, and we use it in two configurations: including all available modes (NRSur7dq2HM) and including only the dominant $\ell = 2$ multipoles (NRSur7dq2L2). We emphasize that this is the only model in literature that both includes $\ell > 2$ GW modes and captures unabbreviated BH spin dynamics through all 6 degrees of freedom. In this paper, we demonstrate the viability of using NRSur7dq2 in the follow-up parameter estimation of GW signals coming from heavy BBHs such as those that LIGO-Virgo have observed multiple times already [11]. In particular, we use it to estimate all physical parameters of the first two heavy BBH events GW150914 and GW170104, significantly extending their past analyses [4, 27, 34].

We first perform controlled tests by injecting 48 synthetic GW signals (details in Table I) into zero noise and inferring their source parameters with both NRSur7dq2 surrogate models. We compare these results against the IMRPhenomPv2 model, which captures spin-orbit precession effects and has been used extensively in recent LIGO-Virgo analyses [22, 35, 36]. We vary the source parameters of injections as follows. Mass ratio is varied from $q = 1.2 - 1.5$, source location between $500 - 1500$Mpc, orbital inclination between close to face-on and edge-on configurations, and component spins are chosen from four distinct configurations with magnitudes $0.4 - 0.65$. These values are deliberately chosen to enhance spin-induced orbital precession. We find that even NRSur7dq2L2 can noticeably improve on IMRPhenomPv2 when it comes to measuring masses and mass ratios of binary sources out to $\sim$ 1Gpc. This can be seen from Fig. 2. All other binary parameters such as component spins and source location are recovered consistently by both models. We further find that the inclusion of higher order $\ell = \{3, 4\}$ GW modes in NRSur7dq2HM allows us to measure luminosity distance and orbital inclination more accurately for sources out to $\sim$ 1Gpc. This improvement is especially noticeable for edge-on configurations, which is expected since higher order modes contribute relatively more when we observe the source at a larger angle. For such sources, NRSur7dq2HM also measures binary mass ratios somewhat more precisely than IMRPhenomPv2 and NRSur7dq2L2. Finally, we find that BH spins are measured broadly consistently by all three models. For the closest sources (within $\sim$ 500Mpc) we do gain some additional information with NRSur7dq2HM templates. This improvement is, however, modest for the investigated cases and we expect it to be more signficant for binary sources with higher mass ratios for which sub-dominant modes carry a larger fraction of the total signal power [37–39]. Since NRSur7dq2 is presently restricted to $1 \leq q \leq 2$, our results provide incentive for extending the domain of NR based surrogates to higher mass ratios.

Having established the performance of NRSur7dq2 surrogates within a fully Bayesian parameter recovery framework, we next analyze the first-ever recorded BBH merger event: GW150914. The primary improvement we note is in the estimation of the binary's luminosity distance d_L from Earth: with extra information coming from sub-dominant modes, we are able to constrain d_L close to $\sim$ 530Mpc, about 100Mpc further away than all others models' estimation. Simultaneously, the surrogate also constrains the source of GW150914 to be either face-on or face-off[2] more strongly than other models, disfavoring edge-on configurations. Consistent with this, the NRSur7dq2 models estimate the source's chirp/total mass to be marginally higher than what approximate models measure. And finally, having complete 2-body spin information encoded in them, the NRSur7dq2 models constrain the effective-spin of GW150914 to be closer to zero than other models (with the same sampling priors). These results continue to hold when we compare

[1] All simulations can be re-scaled to any point in the eight dimension of total mass M.

[2] We use "face-on" to mean that the observer is close to the north pole of the binary (θ_{JN} close to 0 degrees), and "face-off" to mean that the observer is close to the binary's south pole (θ_{JN} close to 180 degrees).

them with the LVC analysis of this event [23].

We also analyze the second heavy BBH event GW170104 recorded by LIGO detectors in early 2017. As for GW150914, we find that the surrogate constrains the luminosity distance to this event to be larger (by 10%) than what approximate models that include only $\ell = 2$ GW modes do. Similarly, it also constrains the source to be closer to face-on/off than edge-on more strongly than other models. The estimation of mass parameters is consistent between `NRSur7dq2` and semi-analytic models, with the former only recovering the portion of mass ratio posterior with support in $1 \leq q \leq 2$. Finally, we find the estimation of spin parameters to be remarkably similar between `NRSur7dq2`, `IMRPhenomPv2` and `SEOBNRv4`, with little extra information coming from the use of NR surrogates. These results are all consistent with the the first analysis of this event by the LVC [4]. A summary in the form of median estimates and symmetric 90% credible intervals for inferred quantities is given in Table II.

From the analyses of GW150914 and GW1701014, we learn that one consistently recovers additional information that helps break the luminosity distance - inclination degeneracy for BBH events with NR surrogate templates, allowing us to constrain GW source and orientation location better. We also learn that, in some cases, one could constrain BBH effective spins better with the NR surrogates since they contain unabbreviated nonlinear GR information. However, spin measurements are sensitive to the choice of sampling priors employed [34, 40], and we defer an investigation of their effect on spin inferences for both events to future work. Our results are encouraging and we propose that `NRSur7dq2HM` and future NR surrogate models be used as part of standard GW event follow-ups. We also encourage the NR community to further the development of surrogate models to higher-mass ratios, so that more BBH sources can be studied with them. In order to enable further analysis by the community, we provide full posterior samples from Bayesian parameter estimation of LIGO/Virgo data for GW150914 and GW170104, with `NRSur7dq2HM` and `IMRPhenomPv2`. These can be obtained from https://github.com/prayush/GW150914_GW170104_NRSur7dq2_Posteriors.

The remainder of this paper is organized as follows. In Section II we describe the surrogate and approximate waveform models used in this paper, as well as the details of our Bayesian parameter estimation machinery. In Section III we present results from studies involving parameter recovery from synthetic signals. In Section IV and V we present results of our re-analysis of GW150914 and GW1701014 using the new NR surrogate model. And finally, in Section VI we summarize our findings and present the future outlook for this research.

II. PRELIMINARIES

A. Numerical Relativity Surrogates

A surrogate waveform model is one that takes a set of pre-computed waveforms generated by an underlying model as input, and interpolates in parameter space between these waveforms to produce waveforms for arbitrary parameter values. The underlying waveform model can be analytic, phenomenological, or purely numerical. Surrogates can often be evaluated in a fraction of the time that takes for the underlying model to generate a waveform, and was in fact originally proposed as a way to reduce the computational cost of otherwise expensive waveform models when used with MCMC-based parameter estimation algorithms applied to GW events [29]. With interpolation comes an additional source of modeling error, called the *surrogate* error. In principle, this error can be arbitrarily reduced by using a sufficiently large set of pre-computed waveforms to cover the parameter space. In practice, when using NR waveforms the cost may become prohibitive.

The `NRSur7dq2` model of Ref. [33] spans the 7-dimensional space of spin-precessing non-eccentric black hole binaries. It is built from the results of 744 NR simulations performed using the Spectral Einstein Code `SpEC` [41] and has already found several applications [42, 43]. That it spans all spin-precession degrees of freedom comes at the cost of limiting its domain to comparable mass ratios $q = m_1/m_2 \leq 2$ and black hole spins with magnitudes ≤ 0.8 of their extremal values. The choice of NR simulations used to train this surrogate was based on a combination of methods including sparse grids [44, 45] (as detailed in Appendix A of [33]), a template-metric-type stochastic sampler, and existing NR simulations. Taken together, these choices maximized the coverage of the binary parameter space with as few simulations as possible while simultaneously keeping the surrogate error sufficiently small. Instead of modeling waveform modes directly across the parameter space, the strategy of `NRSur7dq2` is to interpolate quantities that have as little structure (such as oscillations) as possible. Ref. [33] constructs surrogate models for combinations of waveform modes in the *coorbital frame*, as well as for orbital phase and spin-related quantities that are required to transform these modes back to an inertial frame. They choose to parameterize these fits using *instantaneous* spins and mass ratio, instead of *initial* spins, as they find this choice improves the quality of fits. Therefore, evaluation of `NRSur7dq2` requires first obtaining the full time-evolution of BH spins, orbital phase, and the unit quaternion that defines the coprecessing frame [35, 46, 47]; and subsequently using these evolutions to construct full inertial frame waveform modes from surrogate evaluation of coorbital frame modes. We refer the reader to [32, 33] for further technical details and reasoning supporting various choices of surrogate construction. Finally, we note that `NRSur7dq2` is limited in length to span the last 20 binary

orbits before merger. In practice, with a lower frequency cutoff of 20 Hz, this restricts its use to binaries with total masses $M = m_1 + m_2 \gtrsim 50 M_\odot$. In the remainder of the paper, we will use `NRSur7dq2L2` to mean the surrogate with only $\ell = 2$ waveform modes included, `NRSur7dq2HM` for the surrogate with all its available $\ell = \{2, 3, 4\}$ modes, and `NRSur7dq2` when discussing the surrogate model in general.

B. Analytic Waveform Models

In this paper, we will consider two waveform families: Effective-One-Body (EOB) and phenomenological (IMRPhenom) [19, 48]. Both of these are semi-analytic models of the complete inspiral-merger-ringdown for spinning BHs with non-eccentric orbits. For both models, we consider the dominant $(\ell, m) = (2, \pm 2)$ spin-weighted spherical harmonic waveform multipoles as only these have been calibrated to NR simulations through merger.

Effective-One-Body: The effective-one-body approach solves for the dynamics of the two-body problem in nonlinear GR by mapping it to the dynamics of an *effective* test particle of mass $\mu = m_1 m_2/(m_1 + m_2)$ and spin $S^*(m_1, m_2, \vec{\chi}_1, \vec{\chi}_2)$ in a background spacetime that is described by a parameterized deformation of the Kerr metric. Both S^* and the background deformation (to leading order) are chosen so that the geodesic followed by the test particle reproduces the perturbative post-Newtonian (PN) dynamics of the original two-body system [49]. This conserved dynamics of the test particle is described by the EOB Hamiltonian, which is also derived to leading order using PN results. The radiative dynamics is introduced through a flux of energy to emitted gravitational radiation, obtained by summing over all PN-expanded waveform modes at future null infinity. All of these model pieces are individually taken beyond known PN orders through resummation and addition of phenomenological parameters that are subsequently calibrated to ensure agreement of the inertial-frame waveform multipoles with NR simulations. This allows the EOB prescription to be extended beyond the slow-motion regime where PN results are valid, all the way up till the two BHs merge. After merger, the ringdown waveform is constructed as a linear superposition of the first eight quasi-normal modes (QNMs) of the Kerr BH formed at merger [50]. This ringdown waveform is suitably matched with the inspiral-merger portion by enforcing continuity of waveform modes and their first time-derivatives.

We use the most recent `SEOBNRv4` model [51] (available within the LIGO Algorithms Library (LAL) [52]) in this study. This model describes BBHs with component spins parallel to the orbital angular momentum (i.e., non-precessing binaries), on non-eccentric orbits, and was calibrated to 141 NR simulations. We refer the reader to [51] and references therein for a comprehensive description of the model. In the interest of minimizing computational cost, we use the reduced-order model for `SEOBNRv4` that was also introduced in [51]. We, however, are unable to use the precessing EOB model of [53] in this study due to its high computational cost.

Phenomenological model: `IMRPhenomPv2` is a phenomenological model constructed in the frequency domain that describes GWs emitted by non-eccentric spinning-precessing binaries during their inspiral-merger and ringdown phases [22, 35, 36]. It relies on the approximation that a generic precessing-binary inspiral waveform can be obtained by rotating the waveform for an equivalent spin-aligned system in its quadrupole-aligned frame to the inertial frame using time-dependent rotors (c.f. PN theory) [35, 36]. In the quadrupole-aligned frame, leading order $(\ell, m) = (2, \pm 2)$ modes of the waveform are constructed using the non-precessing IMRPhenomD model [22]. The IMRPhenomD model has a closed form in frequency domain, constructed piecewise in three portions: (i) early inspiral: where both mode amplitude and phasing are given by extensions of PN-theory results; (ii)-(iii) late inspiral and ringdown: where phenomenological ansatzes are taken for waveform amplitude and phasing, and calibrated to enforce high-precision agreement with NR simulations from various numerical relativity groups. Note that `IMRPhenomD` captures BH spin effects on binary inspirals using the effective-spin combination $\chi_{\rm eff} := (m_1 \chi_{1z} + m_2 \chi_{2z})/M$, while `IMRPhenomPv2` uses a precessing-spin parameter χ_p [36] to capture the precession of the quadrupole-aligned frame with respect to inertial observers (instead of using individual BH spin vectors). Also note that `IMRPhenomPv2`(D) belong to the unique class of models that are both closed-form in the frequency domain *and* describe the complete inspiral-merger-ringdown of spin-precessing binaries. These features are ideally suited for GW searches and parameter estimation, which could require the generation of a large number of waveform templates for each event.

C. Parameter Estimation Methodology

Let us denote the collection of measured parameters that describe a GW signal received from a BBH merger event (including the binary's dynamical and kinematic parameters, and other detector-related parameters[3]) as $\vec{\theta}$. The problem statement for PE is to estimate the probability distribution $p(\vec{\theta})$ for the source binary. Using Bayes' theorem, this posterior probability distribution $p(\vec{\theta})$ given data s from GW detectors containing the signal, and a model for GW signals H can be constructed as

$$p(\vec{\theta}|s, H) = \frac{p(s|\vec{\theta}, H)\, p(\vec{\theta}|H)}{p(s|H)}, \tag{1}$$

[3] such as those that describe instrument calibration uncertainty [54, 55]. For these, we take a conservative estimate of 10%/10° uncertainty in amplitude / phase calibration for both LIGO detectors [55]

where: (i) $p(s|H)$ is the *prior* expectation of obtaining the new data s, (ii) $p(\vec{\theta}|H)$ is the expectation on parameters $\vec{\theta}$ for astrophysical sources *prior* to obtaining the new data, and (iii) $p(s|\vec{\theta}, H)$ is the likelihood of obtaining data s (= signal + noise) given $\vec{\theta}$ describes the signal embedded in it. Assuming the detector noise is stationary colored-Gaussian with zero-mean, we can write

$$p(s|\vec{\theta}, H) \propto \exp(-\frac{1}{2}\langle s - h_H(\vec{\theta})|s - h_H(\vec{\theta})\rangle), \quad (2)$$

where $h_H(\vec{\theta})$ is the signal waveform generated with the chosen GW model H, and the noise-weighted inner product $\langle\cdot|\cdot\rangle$ between a and b is defined as

$$\langle a|b\rangle = 4\Re \int_{f_l}^{f_u} \frac{a(f)b^*(f)}{S_n(f)} \mathrm{d}f, \quad (3)$$

with $S_n(f)$ representing the one-sided power spectral density (PSD) of detector noise for LIGO. In this study, we use f_u as the Nyquist frequency corresponding to a sampling rate of 4096Hz. We use the zero-detuning high power design sensitivity curve for Advanced LIGO [56, 57] when not using detector data, and use $f_l = 20$Hz as the lower frequency cutoff. For both events (GW150914 and GW170104) we use LIGO data from its open science center [58], and estimate detector PSD using 1024 seconds of data around the signal concerned as described in [9].

We compute $p(\vec{\theta}|s, H)$ using the Bayesian inference package LALInference [9] that is available as part of the LALSuite software library [52]. LALInference has been extensively used in past analyses published by the LIGO-Virgo Collaborations [1–5], and uses the nested sampling algorithm [59] to estimate source parameters from GW data. We refer the reader to [9] for details of its implementation. As was its original purpose, nested sampling already computes the integrated evidence $Z \equiv p(s|H)$ of the model H. While unimportant to the parameter estimation problem, Z is the key quantity of interest for the purpose of model selection.

For all analyses in this article, we choose sampling priors $p(\vec{\theta}|H)$ identical to those chosen in recent LIGO-Virgo results papers [4, 23], i.e., both BH masses and spin magnitudes are sampled uniformly over their respective ranges, while spin directions are chosen uniformly over a 2−sphere; source distance and sky location are sampled uniformly in 3−D spatial volume out to 2000Mpc, initial inclination angle is sampled uniformly from $[0, \pi]$, and the remaining kinematic parameters are sampled uniformly over their respective ranges. While these priors allow for a direct comparison of our results with published LVC analyses [4, 23], it has been shown [34, 40] that our (common) choice of priors downweights highly spinning binaries for which different choices of prior could improve spin estimation. While neither of these work suggest that GW150914 or GW170104 had large spins, we defer a rigorous study of the effect of priors on the inference of BH spins for these events to future work.

We note from Sec. II A that `NRSur7dq2` is limited to span approximately 40 GW cycles (of the $\ell = |m| = 2$ modes) before merger. Therefore, if the stochastic sampler of LALInference samples a point in binary parameter space for which the complete waveform starting at 20Hz is longer than 40 cycles, the integrated likelihood (c.f. Eq. 3) is automatically reduced due to a reduction of the integration bandwidth to start at the surrogate start frequency instead of 20Hz. We do not, however, *a priori* reject such a jump proposal. Finally, we also note that waveform modes included in `NRSur7dq2` templates that have $m > 2$ (such as the $(3,3), (4,4)$ modes) can start at frequencies above 20Hz. In order to mitigate the Gibbs phenomena brought on by the sudden start of these higher-m modes, we taper all templates at their start. However, some of the information in these modes, contained in frequencies between 20Hz and their start, will be ignored in our analyses (as in previous analyses with numerical simulations [27]). We expect though that these modes contribute the most near merger and that the effect of missing lower frequencies should be minimal [27].

III. PARAMETER ESTIMATION OF SYNTHETIC GW SIGNALS

Since the `NRSur7dq2` surrogate model have not been used to extract BBH parameters from GW signals before, we start with controlled tests using synthetic GW signals to establish the model's viability and benefits for this purpose. We know that `NRSur7dq2` produces NR-level accurate templates [33], and includes $\ell \leq 4$ waveform modes. Therefore there are two reasons why performing parameter estimation with `NRSur7dq2` templates may furnish more accurate results than any waveform model, since approximate models can still struggle with modeling the highly dynamical merger regime [25], and none of the precessing ones used in recent LIGO-Virgo papers include $\ell > 2$ modes [1–5, 11]. With injection tests we will study both reasons together and highlight their distinct effects wherever manifest. We will also investigate when information is lost due to artificially restricted domain of validity of `NRSur7dq2`.

We perform a total of 48 injections in zero noise[4] for both LIGO detectors, and analyze the resulting coincident synthetic data with the `NRSur7dq2L2` model, the `NRSur7dq2HM` model, and the precessing `IMRPhenomPv2` model (with and without artificially restricted priors from `NRSur7dq2`). We do not include `SEOBNRv4` in this section as it only models non-precessing sources, and

[4] "zero noise" implies that data is composed of an injected signal plus zeros. Since detector noise is assumed colored-Gaussian with zero mean, using zero noise with a detector-noise weighted likelihood in Eq. 2 makes our analysis equivalent to the average over an *ensemble* of analyses which use actual noise-realizations [60].

TABLE I. Parameters of injection set. Total 48 injections. Sources' sky location is chosen from a uniform distribution over a 2−sphere, while their polarization angle, i.e. the third Euler angle required to rotate from the source to detector frame, is chosen from a uniform distribution over $[0, \pi]$. Choices for these three are held fixed for all injections. Total mass is held fixed at $60M_\odot$. Their combined network SNR ranges from $\rho = 13 - 83$ for the two-detector Advanced LIGO network.

Injection #	$q \equiv m_1/m_2$	$\vec{\chi}_1$	$\vec{\chi}_2$	$\theta_{\rm JN}$	D (Mpc)	Signal model
0	1.2	$0.65\,(1,1,0)\,/\sqrt{2}$	$-0.65\,(1,1,0)\,/\sqrt{2}$	30°	1500	NRSur7dq2HM
1	1.2	$0.65\,(1,1,0)\,/\sqrt{2}$	$(0,0,0)$	30°	1500	NRSur7dq2HM
2	1.2	$0.4\,(1,1,1)\,/\sqrt{3}$	$0.4\,(1,1,1)\,/\sqrt{3}$	30°	1500	NRSur7dq2HM
3	1.2	$0.4\,(1,1,1)\,/\sqrt{3}$	$-0.4\,(1,1,1)\,/\sqrt{3}$	30°	1500	NRSur7dq2HM
4	1.5	$0.65\,(1,1,0)\,/\sqrt{2}$	$-0.65\,(1,1,0)\,/\sqrt{2}$	30°	1500	NRSur7dq2HM
5	1.5	$0.65\,(1,1,0)\,/\sqrt{2}$	$(0,0,0)$	30°	1500	NRSur7dq2HM
6	1.5	$0.4\,(1,1,1)\,/\sqrt{3}$	$0.4\,(1,1,1)\,/\sqrt{3}$	30°	1500	NRSur7dq2HM
7	1.5	$0.4\,(1,1,1)\,/\sqrt{3}$	$-0.4\,(1,1,1)\,/\sqrt{3}$	30°	1500	NRSur7dq2HM
8	1.2	$0.65\,(1,1,0)\,/\sqrt{2}$	$-0.65\,(1,1,0)\,/\sqrt{2}$	30°	1000	NRSur7dq2HM
9	1.2	$0.65\,(1,1,0)\,/\sqrt{2}$	$(0,0,0)$	30°	1000	NRSur7dq2HM
10	1.2	$0.4\,(1,1,1)\,/\sqrt{3}$	$0.4\,(1,1,1)\,/\sqrt{3}$	30°	1000	NRSur7dq2HM
11	1.2	$0.4\,(1,1,1)\,/\sqrt{3}$	$-0.4\,(1,1,1)\,/\sqrt{3}$	30°	1000	NRSur7dq2HM
12	1.5	$0.65\,(1,1,0)\,/\sqrt{2}$	$-0.65\,(1,1,0)\,/\sqrt{2}$	30°	1000	NRSur7dq2HM
13	1.5	$0.65\,(1,1,0)\,/\sqrt{2}$	$(0,0,0)$	30°	1000	NRSur7dq2HM
14	1.5	$0.4\,(1,1,1)\,/\sqrt{3}$	$0.4\,(1,1,1)\,/\sqrt{3}$	30°	1000	NRSur7dq2HM
15	1.5	$0.4\,(1,1,1)\,/\sqrt{3}$	$-0.4\,(1,1,1)\,/\sqrt{3}$	30°	1000	NRSur7dq2HM
16	1.2	$0.65\,(1,1,0)\,/\sqrt{2}$	$-0.65\,(1,1,0)\,/\sqrt{2}$	30°	500	NRSur7dq2HM
17	1.2	$0.65\,(1,1,0)\,/\sqrt{2}$	$(0,0,0)$	30°	500	NRSur7dq2HM
18	1.2	$0.4\,(1,1,1)\,/\sqrt{3}$	$0.4\,(1,1,1)\,/\sqrt{3}$	30°	500	NRSur7dq2HM
19	1.2	$0.4\,(1,1,1)\,/\sqrt{3}$	$-0.4\,(1,1,1)\,/\sqrt{3}$	30°	500	NRSur7dq2HM
20	1.5	$0.65\,(1,1,0)\,/\sqrt{2}$	$-0.65\,(1,1,0)\,/\sqrt{2}$	30°	500	NRSur7dq2HM
21	1.5	$0.65\,(1,1,0)\,/\sqrt{2}$	$(0,0,0)$	30°	500	NRSur7dq2HM
22	1.5	$0.4\,(1,1,1)\,/\sqrt{3}$	$0.4\,(1,1,1)\,/\sqrt{3}$	30°	500	NRSur7dq2HM
23	1.5	$0.4\,(1,1,1)\,/\sqrt{3}$	$-0.4\,(1,1,1)\,/\sqrt{3}$	30°	500	NRSur7dq2HM
24	1.2	$0.65\,(1,1,0)\,/\sqrt{2}$	$-0.65\,(1,1,0)\,/\sqrt{2}$	75°	1500	NRSur7dq2HM
25	1.2	$0.65\,(1,1,0)\,/\sqrt{2}$	$(0,0,0)$	75°	1500	NRSur7dq2HM
26	1.2	$0.4\,(1,1,1)\,/\sqrt{3}$	$0.4\,(1,1,1)\,/\sqrt{3}$	75°	1500	NRSur7dq2HM
27	1.2	$0.4\,(1,1,1)\,/\sqrt{3}$	$-0.4\,(1,1,1)\,/\sqrt{3}$	75°	1500	NRSur7dq2HM
28	1.5	$0.65\,(1,1,0)\,/\sqrt{2}$	$-0.65\,(1,1,0)\,/\sqrt{2}$	75°	1500	NRSur7dq2HM
29	1.5	$0.65\,(1,1,0)\,/\sqrt{2}$	$(0,0,0)$	75°	1500	NRSur7dq2HM
30	1.5	$0.4\,(1,1,1)\,/\sqrt{3}$	$0.4\,(1,1,1)\,/\sqrt{3}$	75°	1500	NRSur7dq2HM
31	1.5	$0.4\,(1,1,1)\,/\sqrt{3}$	$-0.4\,(1,1,1)\,/\sqrt{3}$	75°	1500	NRSur7dq2HM
32	1.2	$0.65\,(1,1,0)\,/\sqrt{2}$	$-0.65\,(1,1,0)\,/\sqrt{2}$	75°	1000	NRSur7dq2HM
33	1.2	$0.65\,(1,1,0)\,/\sqrt{2}$	$(0,0,0)$	75°	1000	NRSur7dq2HM
34	1.2	$0.4\,(1,1,1)\,/\sqrt{3}$	$0.4\,(1,1,1)\,/\sqrt{3}$	75°	1000	NRSur7dq2HM
35	1.2	$0.4\,(1,1,1)\,/\sqrt{3}$	$-0.4\,(1,1,1)\,/\sqrt{3}$	75°	1000	NRSur7dq2HM
36	1.5	$0.65\,(1,1,0)\,/\sqrt{2}$	$-0.65\,(1,1,0)\,/\sqrt{2}$	75°	1000	NRSur7dq2HM
37	1.5	$0.65\,(1,1,0)\,/\sqrt{2}$	$(0,0,0)$	75°	1000	NRSur7dq2HM
38	1.5	$0.4\,(1,1,1)\,/\sqrt{3}$	$0.4\,(1,1,1)\,/\sqrt{3}$	75°	1000	NRSur7dq2HM
39	1.5	$0.4\,(1,1,1)\,/\sqrt{3}$	$-0.4\,(1,1,1)\,/\sqrt{3}$	75°	1000	NRSur7dq2HM
40	1.2	$0.65\,(1,1,0)\,/\sqrt{2}$	$-0.65\,(1,1,0)\,/\sqrt{2}$	75°	500	NRSur7dq2HM
41	1.2	$0.65\,(1,1,0)\,/\sqrt{2}$	$(0,0,0)$	75°	500	NRSur7dq2HM
42	1.2	$0.4\,(1,1,1)\,/\sqrt{3}$	$0.4\,(1,1,1)\,/\sqrt{3}$	75°	500	NRSur7dq2HM
43	1.2	$0.4\,(1,1,1)\,/\sqrt{3}$	$-0.4\,(1,1,1)\,/\sqrt{3}$	75°	500	NRSur7dq2HM
44	1.5	$0.65\,(1,1,0)\,/\sqrt{2}$	$-0.65\,(1,1,0)\,/\sqrt{2}$	75°	500	NRSur7dq2HM
45	1.5	$0.65\,(1,1,0)\,/\sqrt{2}$	$(0,0,0)$	75°	500	NRSur7dq2HM
46	1.5	$0.4\,(1,1,1)\,/\sqrt{3}$	$0.4\,(1,1,1)\,/\sqrt{3}$	75°	500	NRSur7dq2HM
47	1.5	$0.4\,(1,1,1)\,/\sqrt{3}$	$-0.4\,(1,1,1)\,/\sqrt{3}$	75°	500	NRSur7dq2HM

could not include the precessing EOB model [53] because of its high computational cost of generation [61]. All synthetic signals are generated with the most accurate model available, i.e., `NRSur7dq2HM` including all $\ell \leq 4$ modes. We use the design noise curve for Advanced LIGO detectors [56, 57] while filtering, with a lower frequency cutoff of 20Hz.

For the analyses to be relevant to heavy BH binaries, we fix the total mass to $M = 60M_\odot$ for all signals. All other parameters are chosen in the following manner. Mass ratio is chosen from two values $q = \{1.2, 1.5\}$. BH spins are sampled from 4 distinct configurations: (i) both spins with magnitude 0.65 and both initially parallel to orbital plane with $\hat{\chi}_1 = -\hat{\chi}_2$; (ii) spin on bigger BH with magnitude 0.65 and parallel to orbital plane, with $\vec{\chi}_2 = 0$; (iii) both spins with magnitude 0.4 and mutually parallel with $\hat{\chi}_1 = \hat{\chi}_2 = \frac{1}{\sqrt{3}}(1,1,1)$; and (iv) both spins with magnitude 0.4 and anti-parallel with $\hat{\chi}_1 = -\hat{\chi}_2 = \frac{1}{\sqrt{3}}(1,1,1)$. These spin configurations are chosen to enhance the effects of spin-induced orbital precession. The initial inclination of binary's total angular momentum with the detectors' line of sight is chosen from two values, one close to nearly face-on with $\theta_{\rm JN} = 30°$ and another close to edge-on with $\theta_{\rm JN} = 75°$. Each chosen source is then placed at a luminosity distance $d_L = \{500, 1000, 1500\}$Mpc from the detectors. Together these choices form a grid of $2 \times 4 \times 2 \times 3 = 48$ injections. We list these parameter choices for injections in Table I.

Against these synthetic GW signals, we infer posterior probability distributions for source parameters as described in Sec. II C. For a pedagogical overview, we start with examining a few select injections and study their parameter recovery with `NRSur7dq2`. We choose 3 injections corresponding to the same binary, with $M = 60M_\odot$, $q = 1.2$, $\vec{\chi}_1 = -\vec{\chi}_2 = \frac{0.65}{\sqrt{2}}(1,1,0)$, located at distances of 1500Mpc, 1000Mpc, and 500Mpc. All of these 3 are inclined to the line of sight at an angle $\theta_{\rm JN} = 30°$. These are labelled $\#0, 8, 16$ in Table I. We also consider a fourth injection (labelled #32) that has the same physical parameters as injection #8, *but* is inclined at $\theta_{\rm JN} = 75°$. For all four, we show the recovery of their mass parameters in the top sub-figure of Fig. 1, and their orbital inclination and luminosity distance in the bottom sub-figure. For each case, solid vertical lines in all panels showing 1D histograms, as well solid vertical and horizontal lines in 2D panels, indicate true injected parameter values. Dashed vertical lines indicate measured bounds on them in the form of inferred 90% credible intervals. Let us focus first on the top sub-figure of this figure that focuses on chirp mass $\mathcal{M}_c$ and mass ratio q. Looking first at $\mathcal{M}_c$ recovery for the first 3 cases, we notice a stark reduction in the width of measured 90% credible intervals with increasing SNR (or decreasing distance). The measurement of q also improves as the source moves closer from 1500 to 500Mpc, albeit more slowly than for chirp mass. The fourth injection (#32, shown in cyan) is nearly edge-on with respect to the line of sight to LIGO detectors. Comparing it with the others, we immedi-

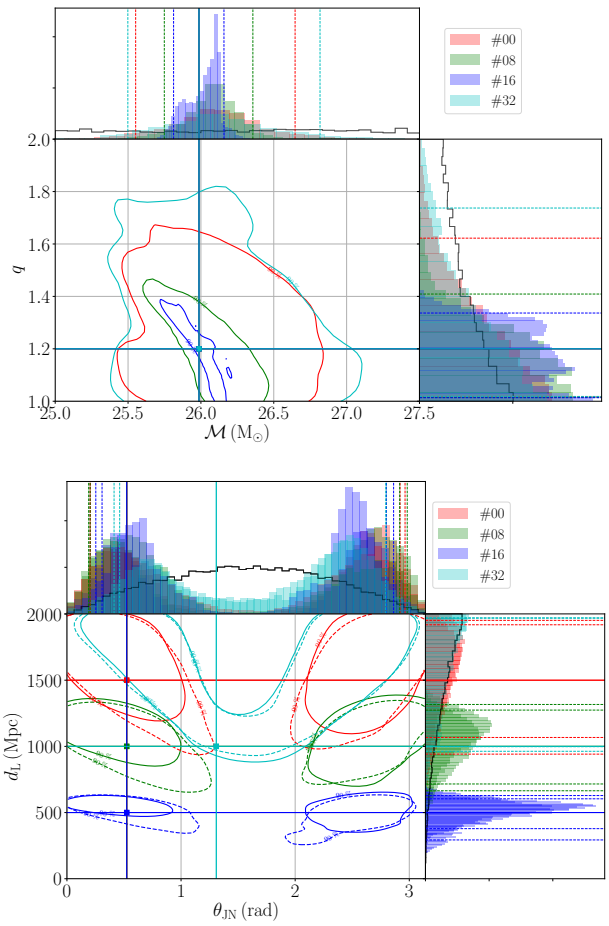

FIG. 1. Parameter recovery for select synthetic injections with `NRSur7dq2HM` templates. Here we show the mass (top) and source distance / inclination angle (bottom) recovery for four injections, labeled $\#0, 8, 16, 32$ in Table. I. Both top and bottom sub-figures have three panels: one that shows two dimensional 90% credible regions for the joint measurement of both parameters (for that sub-figure), while the other two show one dimensional marginalized probability distributions measured for each of the same two source parameters (for that sub-figure). All 4 injections have identical source mass and spin parameters. The first three have identical source inclination angles as well, but differ in the distance at which their source is located: 1500Mpc, 1000Mpc, and 500Mpc respectively. The fourth injection (#32) is similar to injection #8 except that its orbital inclination is much closer to edge-on, i.e. $\theta_{\rm JN} = 75°$ for #32. In the bottom sub-figure, we additionally show results from analyses with `NRSur7dq2L2` templates as dashed 2D contours. These can be directly contrasted with solid contours to read the effect of including $l > 2$ modes in templates. In both 2D and 1D panels, solid colored lines mark the true injected parameter value, and dashed vertical lines show the limits of 90% credible regions for the relevant parameter. In all 1D panels, black curves show the sampling prior for that parameter. See text for further discussion.

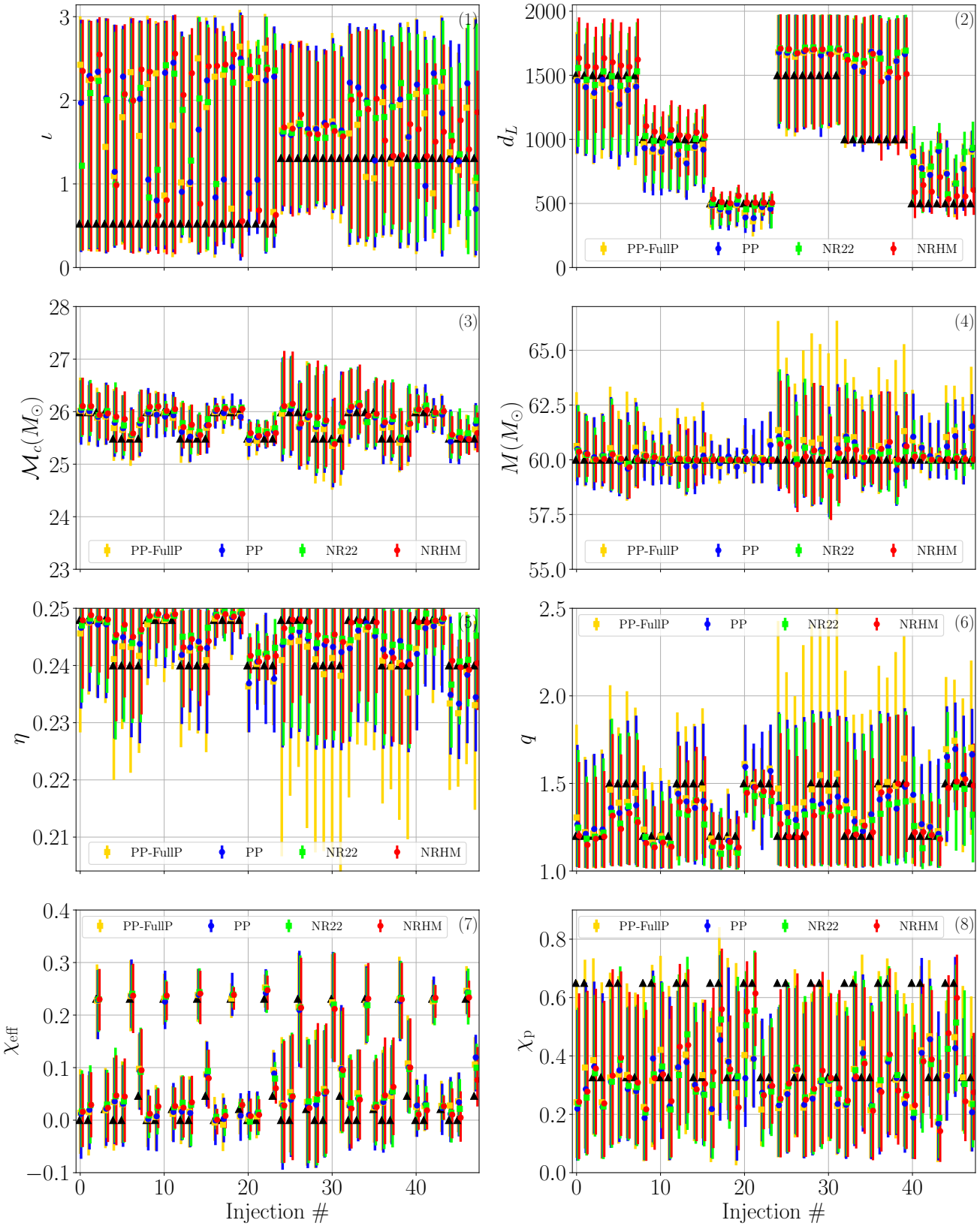

FIG. 2. Estimated masses, spins, and other physical parameters for a set of synthetic GW signals, analyzed using different approximants. In different colors, we show results for three template models: `IMRPhenomPv2` (blue, labeled PP), `NRSur7dq2L2` (labeled NR22) and `NRSur7dq2HM` (labeled NRHM), each restricted to the domain of validity of the `NRSur7dq2` surrogate model. In addition we show results for `IMRPhenomPv2` with the model allowed to explore its entire domain of support [22, 35, 36] (yellow, labeled PP-FullP). Parameters of injected signals are given in Table I and described in text. True values of injection parameters are marked by black triangles. Vertical line segments show measured 90% credible intervals, and colored circles show the corresponding median estimates. Panel numbers are indicated in top right corners. See text for discussion.

ately see how increasing the source's inclination angle toward $\frac{\pi}{2}$ makes the measurement of BH masses significantly worse. This is because $\theta_{\rm JN} \to \frac{\pi}{2}$ decreases the contribution of dominant $\ell = |m| = 2$ modes, and therefore reduces the overall SNR. Further - in the bottom sub-figure of Fig. 1, we show the recovery of sources' inclination and luminosity distance from LIGO detectors. All presentation attributes of this sub-figure are identical to those of the top, with one addition. In panels showing 2D credible regions, while solid contours still correspond to `NRSur7dq2HM`, we have added corresponding *dashed* contours for `NRSur7dq2L2`. We immediately see that the first three injections, which are nearly face-on, have similar 90% credible regions - each two-lobed around face-on and face-off orientations. This is as we expect since both orientations are degenerate and maximize the contribution of dominant $\ell = |m| = 2$ waveform modes. We find that the presence of $l > 2$ modes in recovery templates restricts the distance-inclination posterior further, as seen by comparing solid and dashed 2D contours. Regions of the posterior that underestimate luminosity distance are ruled out more aggressively near the lobe around $\theta_{\rm JN} \to \theta^{\rm true}_{\rm JN}$ than around $\theta_{\rm JN} \to \pi - \theta^{\rm true}_{\rm JN}$ for this fiducial binary. We find that this asymmetry between face-on/face-off posterior lobes for distance and inclination also depends on the intrinsic parameters of the source, which modulate the relative signal power content between $l = 3$ and $l = 2$ modes. Further: $\theta_{\rm JN} = 75^\circ$ being close to $\pi/2$, we expect a systematic overestimation of distance for case #32 as the true value is located toward the lower "U" end of the $d_L - \theta_{\rm JN}$ degeneracy contours. We find, accordingly, that distance for the fourth injection is indeed grossly overestimated, in contrast with the other three cases. This is consistent with past results for highly inclined binaries [60].

So far we have illustrated select cases of parameter estimation with the `NRSur7dq2` surrogate, highlighting differences between dominant-mode and higher-mode templates. Next we will investigate the improvements in BBH parameter recovery brought upon by both (a) the presence of higher-order modes, and (b) NR-level accuracy of merger modeling, in `NRSur7dq2HM` templates. We will do so analyzing all injections together. We will quantify our results using marginalized 1D 90% credible intervals as measures of statistical error, with the estimated median values furnishing any corresponding systematic errors. Our results are shown in Fig. 2 for all injections. In each panel, the horizontal axis shows the injection index, which ranges from $0-47$, and was introduced in Table I. Further, black triangles show injected (true) parameter values. Injections are arranged first according to their inclination angle, then according to their source distance, then mass ratio, and finally by the bigger BH's spin magnitude. This implies that the first 24 injections shown have source $\theta_{\rm JN} = 30^\circ$ and the next 24 have $\theta_{\rm JN} = 75^\circ$. Within each of these two blocks of 24 injections, the first 8 have sources at $d_L = 1500$Mpc, next 8 at $d_L = 1000$Mpc and the last 8 at $d_L = 500$Mpc. Within each of these blocks of 8 injections, the first 4 have $q = 1.2$ and the next 4 have $q = 1.5$. And finally, within each block of 4 injections, the first 2 have $|\vec{\chi}|_1 = 0.65$ and the next 2 have $|\vec{\chi}|_1 = 0.4$. This arrangement is manifest in the locations of black triangles in all panels. The median value of the measured marginalized posterior distribution for each parameter is shown in solid circles and the associated 90% credible intervals are shown as vertical line segments. Colors distinguish between template models. Labels NRHM and NR22 correspond to `NRSur7dq2HM` and `NRSur7dq2L2` templates. Two sets of results are shown with `IMRPhenomPv2`, one where its sampling priors are artificially restricted to the domain of `NRSur7dq2` (labeled PP), and the other where they span the entire domain of validity of `IMRPhenomPv2` (labeled PP-FullP).

In the top row of Fig. 2, panels (1) and (2) show the recovery of source inclination and luminosity distance with respect to LIGO detectors. We know that the effect of both of these parameters on GW signals incident on Earth is degenerate, i.e., signals with inclination $\theta_{\rm JN}$ are degenerate with sources with $\theta_{\rm JN} \to \pi - \theta_{\rm JN}$ at similar distances, as well as with sources with $\theta_{\rm JN} \to \pi/2$ at smaller distances or $\theta_{\rm JN} \to 0$ at larger distances. The general shape of this degeneracy is visually appreciable from the 2D cyan contours in the lower left corner of the lower 2D panel in Fig. 1. Therefore, we find that median $\theta_{\rm JN}$ estimates in panel (1) of Fig. 2 are either close to the true value of $\theta_{\rm JN}$ or π minus the true value. For most of nearly face-on cases (with $\theta_{\rm JN} = 30^\circ$), neither `IMRPhenomPv2` nor `NRSur7dq2L2` constrains binary's initial orbital inclination very well, with 90% credible intervals nearly spanning the entire prior range $[0, \pi]$. For more inclined configurations (i.e. with $\theta_{\rm JN} = 75^\circ$), the effect of higher-order waveform modes is enhanced, and we accordingly find that `NRSur7dq2HM` constrains $\theta_{\rm JN}$ better than the other two models with only $\ell = |m| = 2$ modes. This is especially noticeable for sources closer than 1Gpc (c.f. injections $\#32 - 47$). In panel (2) we show luminosity distance measurements. We notice immediately that distance estimates can improve significantly for closer sources with `NRSur7dq2HM` templates. For sources with $\theta_{\rm JN} = 30^\circ$, we find that `NRSur7dq2HM` measures source distance more accurately and precisely than the other two models, with both the median estimate being closer to the true value and 90% credible intervals being smaller. For the same sources, we also find that while `NRSur7dq2L2` does not improve the precision of luminosity distance measurement by much, it does improve its accuracy, as the median estimates with `NRSur7dq2L2` are closer to true values than with `IMRPhenomPv2`. Contrasting the benefit of including higher-order modes in `NRSur7dq2HM` with improved merger modeling of `NRSur7dq2`, we find the missing subdominant modes in `IMRPhenomPv2` to be the leading cause of loss of information in $d_L - \theta_{\rm JN}$ measurements. For closer to edge-on configurations (with $\theta_{\rm JN} = 75^\circ$, injections $\#24 - 47$), we find that d_L is *systematically biased*

toward larger values, which is expected given the nature of $d_L - \theta_{\rm JN}$ degeneracy, as we discussed above in the context of Fig. 1. Even then, `NRSur7dq2HM` estimates are both more accurate and more precise for all injections, especially note the loudest ones: $\#40-47$. Finally, we point out that restricted priors for `NRSur7dq2` do not reduce the quality of either orbital inclination or luminosity distance measurements with the model for chosen injections.

Next we look at the recovery of the binary's mass parameters. We show these for four different mass combinations: chirp mass $\mathcal{M}_c = M\eta^{3/5}$, total mass M, dimensionless mass ratio $\eta = m_1 m_2/M^2$, and mass ratio q, in panels $(3)-(6)$ respectively. For both chirp and total masses, we immediately notice that both are constrained categorically worse for nearly edge-on cases than for nearly face-on ones. For closer sources (out to 500Mpc), we find that both `NRSur7dq2HM` and `NRSur7dq2L2` can measure both $\mathcal{M}_c$ and M more accurately and precisely than `IMRPhenomPv2` (especially total mass). The pattern holds for both $\theta_{\rm JN} = 30^\circ$ and 75° configurations. A similar pattern is seen for the estimation of mass ratio q (or η[5]) in panels $(5)-(6)$. Overall, mass ratios are recovered better for nearly face-on systems. For close binaries (out to 500Mpc) we again find that mass ratios can be better estimated by both `NRSur7dq2HM` and `NRSur7dq2L2` templates than with `IMRPhenomPv2` ones. This is especially evident for injections $\#16-23$ and $\#40-47$. These improvements clearly illustrate one benefit of the NR-based surrogate that does not simply depend on the inclusion of extra information through sub-dominant GW modes, but on its intrinsic ability to reproduce GR signals more faithfully. We therefore surmise that, as expected, *the benefit of following up GW signals with NR surrogate templates is genuinely twofold.* We do note, however, that for highly inclined configurations, the artificially restricted prior of `NRSur7dq2` shows up as the leading factor affecting the quality of M, η, q measurements (but not of $\mathcal{M}_c$). This is seen by comparing the two set of results with `IMRPhenomPv2` (PP and PP-FullP) for injections $\#24-39$ in panels $(3)-(6)$. This clearly motivates the development of NR-based surrogate models for more unequal mass ratios.

Finally, we consider BH spins. We show the recovery of two different combinations of both BH's spins in panels $(7)-(8)$ of Fig. 2: the effective spin $\chi_{\rm eff}$, and the in-plane precessing spin χ_p. From panel (7) we note that effective spin is consistently well estimated by all template models. For the closest sources at 500Mpc, we find that $\chi_{\rm eff}$ is estimated more *precisely* by both `NRSur7dq2HM` and `NRSur7dq2L2`, i.e. with narrower 90% credibility intervals, than other approximants; see, for instance, injections $\#8-23$ and $\#40-47$. This (marginal) improvement comes from the more faithful modeling of spin effects in the dominant GW mode of `NRSur7dq2`, as it does not depend on the inclusion of higher-order modes. From panel (9), we note that χ_p is overall poorly constrained for heavy BBHs such as the cases considered here. The measured 90% credible regions almost span 90% of the entire prior range, implying that very little information about χ_p is available. This is to be expected because the timescale of orbital precession is considerably larger than the orbital timescale. Short signals from heavy BBHs barely span a couple of precession cycles, making measurements of precession-related parameters challenging. We still note, however, that for the closest injected sources at moderate inclination ($\theta_{\rm JN} = 30^\circ$, i.e. injections $\#16-23$), both `NRSur7dq2HM` and `NRSur7dq2L2` recover χ_p somewhat more accurately than `IMRPhenomPv2`. For these spin measurements, we again find no substantial influence of artificially restricted priors of `NRSur7dq2` models.

A more succinct way of summarizing information from all injections is to compute an averaged measure of systematic biases and statistical uncertainties associated with the recovery of various physical parameters θ by different template models. For each parameter we therefore first compute the relative systematic bias $\delta\theta^i_{\rm syst}$ and relative statistical uncertainty $\delta\theta^i_{\rm syst}$ for each injection (indexed $\#i$) as

$$\begin{aligned} \delta\theta^i_{\rm syst} &:= \left|\theta^i_{\rm median} - \theta^i_{\rm true}\right| / \theta^i_{\rm true}, \\ \delta\theta^i_{\rm stat} &:= \left|\Delta\theta^i_{90\%}\right| / \theta^i_{\rm true}, \end{aligned} \tag{4}$$

where $\Delta\theta^i_{90\%}$ is the size of the measured 90% credible region. For parameters whose possible values include 0, such as BH spins and their combinations, we do not divide by $\theta^i_{\rm true}$ in both parts of Eq. 4. We then take the algebraic mean of both $\delta\theta^i_{\rm syst}$ and $\delta\theta^i_{\rm stat}$ over all injections to obtain our combined measures of parameter estimation accuracy and precision: $\overline{\delta\theta_{\rm syst}}$ and $\overline{\delta\theta_{\rm stat}}$. We find that both of these measures are significantly affected by the inclination angles of injections being averaged over. We therefore average over two sets of injections separately, one where the injected $\theta_{\rm JN} = 30^\circ$ and another where $\theta_{\rm JN} = 75^\circ$. We show summary error measures for both sets of injections in the two sub-figures of Fig. 3. The left sub-figure corresponds to injections with $\theta_{\rm JN} = 30^\circ$ and the right one corresponds to those with $\theta_{\rm JN} = 75^\circ$. We show results for `NRSur7dq2L2`, `NRSur7dq2HM`, and two configurations of `IMRPhenomPv2` templates (using unrestricted and `NRSur7dq2`'s restricted sampling priors, labeled IMRPPv2 and IMRPPv2-FullP respectively). As in Fig. 2, colors indicate different template models.

Let us focus on the left sub-figure of Fig. 3. Considering the recovery of binary mass combinations first, we immediately note that for $\mathcal{M}_c, M, \eta, q$: $\overline{\delta\theta_{\rm syst}} \ll \overline{\delta\theta_{\rm stat}}$, and therefore statistical errors dominate their measurement. By comparing results from different template models, we can see that *both* `NRSur7dq2HM` and `NRSur7dq2L2` recover all four mass combinations substantially more *precisely* than `IMRPhenomPv2` templates. Most noticeable is the

[5] the one-to-one map: $\eta = q/(1+q)^2$ ensures that patterns that hold for posteriors of q will hold for η, and vice-versa.

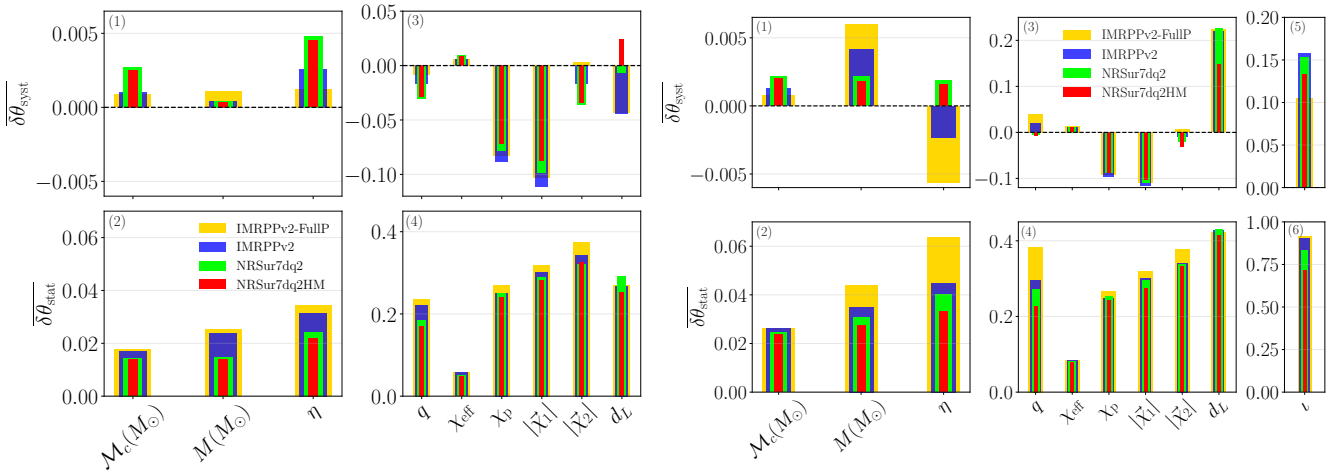

FIG. 3. *Left*: Shown are the mean systematic biases $\overline{\delta\theta_{\rm syst}}$ and statistical uncertainties $\overline{\delta\theta_{\rm stat}}$ for various binary parameters averaged over all injections with $\theta_{\rm JN} = 30^\circ$. Four distinct template configurations are considered: `NRSur7dq2L2` and `NRSur7dq2HM` models, and `IMRPhenomPv2` with and without being artificially restricted to the domain of validity of `NRSur7dq2`. The full prior for `IMRPhenomPv2` extends over $1 \leq q \leq 8$ and spin magnitudes $|\vec{\chi}_{1,2}| \leq 0.89$. Since the signal model is `NRSur7dq2HM`, when the recovery model is also `NRSur7dq2HM` the mean systematic biases and statistical uncertainties reflect the shape of the posterior itself rather than modeling error. *Right*: Same as left panel, except the averaging of $\overline{\delta\theta_{\rm syst}}$ and $\overline{\delta\theta_{\rm stat}}$ is performed over all injections with $\theta_{\rm JN} = 75^\circ$.

improvement in measuring total mass. This improvement is unlikely to be due to artificially restrictive sampling priors of `NRSur7dq2`, as the effect of the same priors is minimal on `IMRPhenomPv2` analyses (as can be seen by comparing blue and yellow bars in panel (2)). We therefore conclude that *the improved modeling of dominant GW mode by* `NRSur7dq2` (in the nonlinear merger regime) *is responsible for this improvement in our capability to measure BBH masses*. Next, we turn our attention to BBH spin combinations $\chi_{\rm eff}, \chi_{\rm p}, |\vec{\chi}_1|$, and $|\vec{\chi}_2|$. Results for these are shown in panels (3)−(4). We again note that the ratio $\overline{\delta\theta_{\rm syst}}/\overline{\delta\theta_{\rm stat}}$ is below 10% for all four, implying that statistical errors dominate. From panel (3), we read that while for $\chi_{\rm p}$ and $|\vec{\chi}_1|$ the surrogates record smaller systematic measurement biases, for $\chi_{\rm eff}$ and the smaller BH's spin its the opposite. However, this improvement is moot unless we improve on the dominant statistical errors. We therefore turn to panel (4). We find that while both `NRSur7dq2` models slightly improve the precision of measurement for individual BH spins and $\chi_{\rm p}$, this improvement is contaminated by the restricted sampling priors of `NRSur7dq2`. This can be seen by comparing yellow and blue bars for the three spin combinations in panel (4). Having said this, we remind the reader that for these spin combinations, little information is actually recovered from data as measurements tend to follow sampling priors for such heavy BBHs. Lastly, we find a small improvement in the measurement precision for effective spins with `NRSur7dq2L2(HM)`, which is too marginal for us to draw generic conclusions. Looking back at panel (7) of Fig. 2 we remind the reader that `NRSur7dq2HM` improves the measurement of $\chi_{\rm eff}$ only for the closest sources (out to 500Mpc), and this improvement gets washed out when we average over all other injections at 1000 − 1500Mpc. Finally, we assess the measurement quality for BBH luminosity distances. Looking at panel (4), we find that both surrogate and `IMRPhenomPv2` templates measure d_L with comparable *precision*. From the panel right above, however, we find that `NRSur7dq2HM` systematically overestimates d_L by only about 2.5% while `IMRPhenomPv2` under-estimates it by twice that amount. We also find that while the former systematically overestimates d_L, the latter underestimates it. This can be understood heuristically by considering the shape of $d_L - \theta_{\rm JN}$ degeneracy contours. Look back at the illustrations in Fig. 1, and focus on the lower sub-figure. The 2D contours show that `NRSur7dq2HM` improves upon the measurement with `NRSur7dq2L2` by ruling out the lower portions of the $d_L - \theta_{\rm JN}$ degeneracy regions. These portions correspond to highly-inclined close-by configurations. Since `NRSur7dq2HM` eliminates them from the posterior, the resulting posterior is bound to move toward larger values of d_L. This can be seen by comparing any of the red/green/blue solid contours with dashed ones in the same 2D panel. This explains why $\overline{\delta\theta_{\rm syst}}$ for d_L is strictly positive with `NRSur7dq2HM`, but is negative with the dominant-mode-only models. However, since the ratio $\overline{\delta\theta_{\rm syst}}/\overline{\delta\theta_{\rm stat}}$ for d_L is 10% with `NRSur7dq2HM`, 4% for `NRSur7dq2L2`, and 16% for `IMRPhenomPv2`, statistical errors still dominate the measurement of d_L *on average*. Reconciling this observation with panel (2) of Fig. 2 (injections #16 − 23), we conclude that `NRSur7dq2HM` can improve the *accuracy* of d_L measurement provided the GW source is close enough ($\lesssim$ 500Mpc). Note that we do not discuss the measurement quality for $\theta_{\rm JN}$ itself as we find to be nearly identical between all template mod-

els.

Moving forward, we focus on nearly edge-on injections in the right sub-figure of Fig. 3. Compared to the nearly face-on cases, we immediately note that the measurement of mass parameters including M, η, and q by `NRSur7dq2` are uncertain enough to be dominated by the restriction on the models' domain of validity. In other words, the full posterior distributions for these parameters have substantial support outside the domain of `NRSur7dq2` and therefore these NR surrogate models will likely produce a biased estimate for them. Having said this, we also note that the chirp mass is measured fairly consistently with all models. It is measured with high precision, and its recovered posteriors are narrow enough to lie completely within the domain of `NRSur7dq2` models, unaffected by sampling prior restrictions. Looking at spin parameters $\chi_{\rm eff}, \chi_{\rm p}, |\vec{\chi}_1|$, and $|\vec{\chi}_|$ next, we find qualitatively very similar features as we did for nearly face-on injections in the left sub-figure of Fig. 3. The measurement of $\chi_{\rm eff}$ is consistent between all four template choices, while that of the other three spin combinations is mildly influenced by the the sampling prior restrictions. Lastly, we focus on the measurement of source distance and inclination. From panels $(3)-(4)$ we find that while all four template models measure luminosity distances to comparable precision, the inclusion of sub-dominant waveform modes in `NRSur7dq2HM` does improve its *accuracy* substantially. Since $\overline{\delta\theta_{\rm syst}}/\overline{\delta\theta_{\rm stat}}$ is close to 50% for d_L here, this improvement in $\overline{\delta\theta_{\rm syst}}$ by `NRSur7dq2HM` templates is substantial[6]. Turning to panels $(5)-(6)$ we find a similar story. The inclusion of higher-order modes in `NRSur7dq2HM` again lead to a substantial reduction in both the systematic and statistical errors associated with measuring orbital inclination. Looking back at panel (1) of Fig. 2 we confirm that this is especially true for closer sources (out to 500Mpc).

From these results, we conclude that parameter recovery with `NRSur7dq2HM` templates can be an improvement over conventional precessing template model that have been used so far to analyze LIGO-Virgo BBH observations [1–5, 11]. These surrogate templates can help estimate source masses better for moderately inclined BBH configurations, and for comparable-mass close-by sources they help resolve the luminosity distance-orbital inclination degeneracy and improve the measurement of both. Our results emphasize the impact of two factors that set NR surrogates apart from other models: (a) NR-level accurate modeling of the dominant GW modes, and (b) inclusion of higher-order harmonics. Measurement improvements that we find here due to (b) are consistent with past work on higher harmonics [37, 38, 62, 63], while those due to (a) are a novel result. Based on these findings, we encourage the GW community to utilize NR surrogates for detailed follow-ups of heavy BBH coalescences. We also motivate the NR community to continue further development of surrogate models, as extending their domain to higher mass ratios can broaden the scope of their applicability.

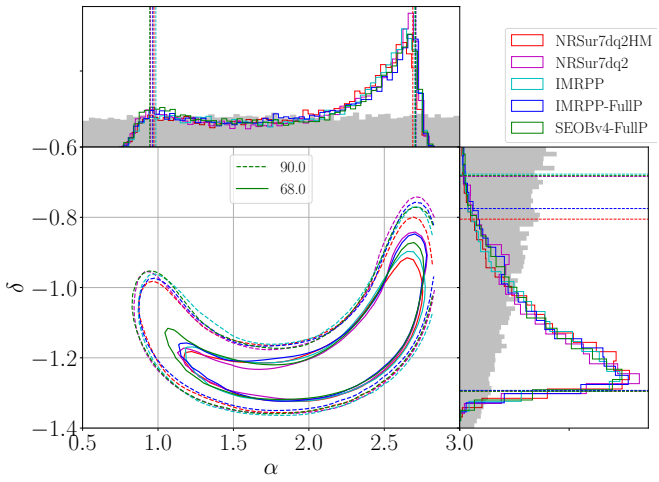

FIG. 4. Estimated sky location for GW150914, using different approximants: `NRSur7dq2HM`, `NRSur7dq2L2`, `IMRPhenomPv2` (labeled IMRPP), and `SEOBNRv4` (labeled SEOBv4). The X-FullP results correspond to an analysis with model X that allows for unrestricted mass ratios $1 \leq q \leq 8$, and spin magnitudes up to $a_{1,2} \lesssim 0.89$ for `IMRPhenomPv2` and $a_{1,2} \lesssim 0.98$ for `SEOBNRv4`. For all others, we *a priori* restrict sampling to $1 \leq q \leq 2$ and $0 \leq a_{1,2} \leq 0.8$, i.e. to the range where NR surrogate models are valid. In all panels showing 1-D posterior distributions, the shaded region shows our prior belief. Vertical dashed lines in 1-D posteriors mark 90% credible regions. The 2-D posteriors show both the 90% (dashed line) and 68% (solid line) credible regions.

We remind the reader that the choice of injected parameters here is made to enhance the effect of precession, and so could be considered a sample of "moderately" precessing sources. It is not, however, drawn from an astrophysically motivated distribution, and is therefore *not* representative of an astrophysical BBH population (in any case, a sample size of 48 over an 8D parameter space is unlikely to be *statistically* representative of any chosen distribution). Therefore, knowing how much benefit we will reap with NR surrogates for a (future) LIGO-Virgo BBH population would require the additional knowledge of how the source parameters of LIGO-Virgo BBH sources are distributed in nature; a study of this is beyond the scope of this article. Finally, note that the choice of using zero noise instead of a particular noise realization ensures that our results hold *on average*, where the averaging is meant in the sense of an ensemble average over an infinite set of noise realizations embedding the same signal. When real instrument noise is present, these results will get shifted depending on the exact nature of the noise realization.

[6] Also note that for edge-on injections (right sub-figure of Fig. 3), luminosity distance d_L is systematically *over*-estimated by *all* models, as has been found before [60] (see Fig. 4)

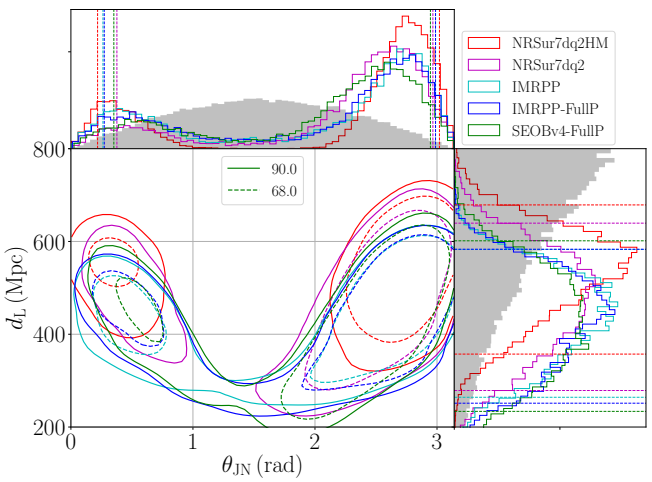

FIG. 5. Estimated source orbital orientation / luminosity distance for GW150914, using different approximants: `NRSur7dq2HM`, `NRSur7dq2L2`, `IMRPhenomPv2` (labeled IMRPP), and `SEOBNRv4` (labeled SEOBv4). All figure attributes are similar to Fig. 4.

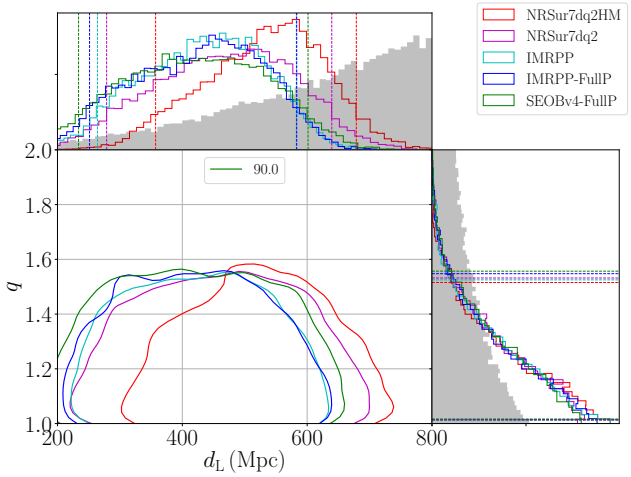

FIG. 6. Luminosity distance and mass ratio measurement for GW150914. All figure attributes are similar to Fig. 4. We find that the samples at large luminosity distances actually correspond to *smaller* mass ratios, and therefore the shifting of distance measurement to larger values when using `NRSur7dq2HM` is not a symptom of the model's restricted sampling priors.

IV. GW150914

Having assessed the performance of `NRSur7dq2` surrogate models in characterizing gravitational-wave signal sources in a fully Bayesian framework, we now analyze the first ever observed GW event GW150914 with these models. As before, we use the nested sampling algorithm in LALInference to perform parameter recovery on the event, and use non-precessing `SEOBNRv4` templates in addition to `NRSur7dq2` and `IMRPhenomPv2`. We perform two analyses with `IMRPhenomPv2`: one where we artificially restrict sampling priors to the domain of `NRSur7dq2` and another where we do not. We, however, find that both analyses furnish almost identical results, and therefore conclude that the effect of sampling priors on GW150914's analyses is minimal. In the analysis with `SEOBNRv4`, we do not artificially restrict the sampling prior. As in the previous section, we do not use the precessing EOB model of [53] due to its high computational cost. Results from all of the above analyses are posterior probability distributions for physical parameters describing the GW source, which are shown in Figs. 4 - 9. In all of these figures, black curves in 1D posterior distributions will show prior distributions for respective parameters.

In Fig. 4, we show the recovery of the source's sky location angles (right ascension α and declination δ) by both `NRSur7dq2` models and compare it with those for semi-analytic models. We immediately note that the recovery of sky location of GW150914's source with `NRSur7dq2L2` is remarkably similar to that from semi-analytic models, but adds little extra information. In Fig. 5, we show the recovery of the source's luminosity distance d_L from LIGO detectors and its total angular momentum's initial inclination $\theta_{\rm JN}$ with respect to the line of sight. These two parameters are strongly degenerate, as can be seen from the 2-D posterior slices showing 1-σ and 90% credible regions for both. Note the effect of higher-order waveform multipoles included in `NRSur7dq2HM` on the measurement of both d_L and $\theta_{\rm JN}$. From their 90% credible intervals in 1D marginalized posteriors, we can see that `NRSur7dq2HM` places GW150914 at $\sim$ 530Mpc, while other models, including `NRSur7dq2L2`, place it at $\sim$ 430Mpc. The primary LVC analyses of the event also inferred the source to be at $\sim$ 410Mpc [23]. Therefore `NRSur7dq2HM` locates the source of GW150914 about 25% further away than what other template models have so far. The difference between d_L posteriors estimated from `NRSur7dq2HM` and `NRSur7dq2L2` strongly implies that the difference in `NRSur7dq2HM`'s luminosity distance estimation is indeed due to the inclusion of higher-order waveform modes. Similarly, the inclination angle is more precisely constrained by `NRSur7dq2HM` to be either face-on or face-off, with edge-on configurations being more strongly disfavored by it than all other models. *These are some of the key findings of this paper. They were inaccessible to the original LVC analyses [4, 23, 27], which were limited by modeling approximations and the availability of a sufficient number of NR simulations.* In Fig. 6 we show the correlated posterior distribution for luminosity distance and mass ratio. We see immediately that the increase in the estimated value of d_L by `NRSur7dq2HM` is not an artifact of the model's restricted domain of validity since the region of the posterior at large distances actually corresponds to nearly equal-mass binaries.

Next, we show the recovery of mass parameters for GW150914 in Fig. 7. While for individual masses and mass ratio, `NRSur7dq2HM` and `NRSur7dq2L2` give us very

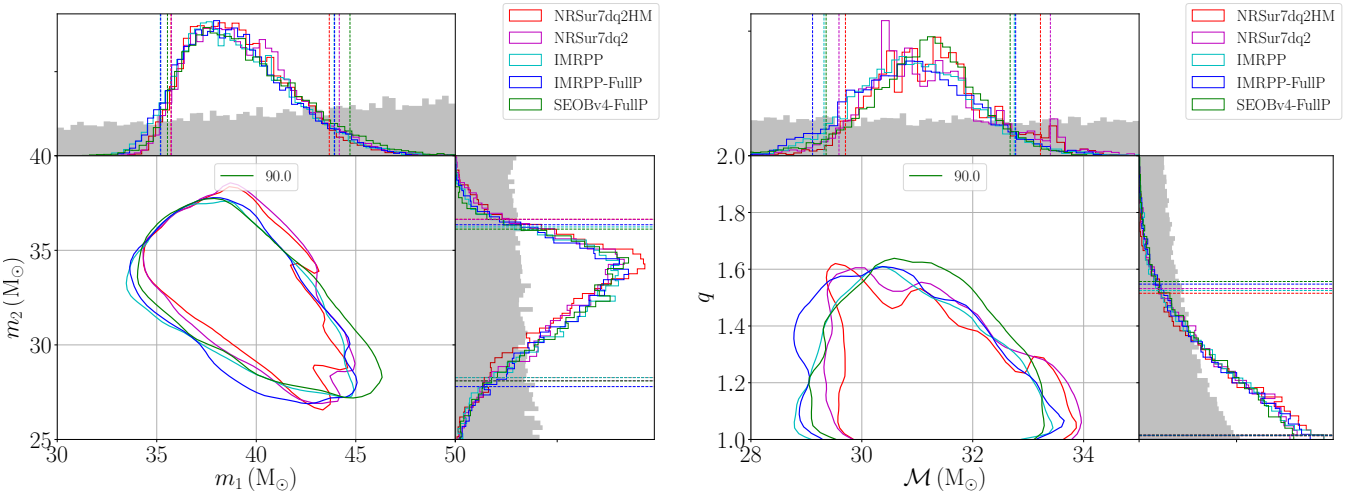

FIG. 7. Estimated masses for GW150914, using different approximants: `NRSur7dq2HM`, `NRSur7dq2L2`, `IMRPhenomPv2` (labeled IMRPP), and `SEOBNRv4` (labeled SEOBv4). The X-FullP results correspond to an analysis with model X that allows for unrestricted mass ratios $1 \leq q \leq 8$, and spin magnitudes up to $a_{1,2} \lesssim 0.89$ for `IMRPhenomPv2` and $a_{1,2} \lesssim 0.98$ for `SEOBNRv4`. For all others, we *a priori* restrict sampling to $1 \leq q \leq 2$ and $0 \leq a_{1,2} \leq 0.8$, i.e. to the range where NR surrogate models are valid. In all panels showing 1-D posterior distributions, the shaded region shows our prior belief. Vertical dashed lines in 1-D posteriors mark 90% credible regions. The 2-D posteriors show the 90% credible regions as a solid line.

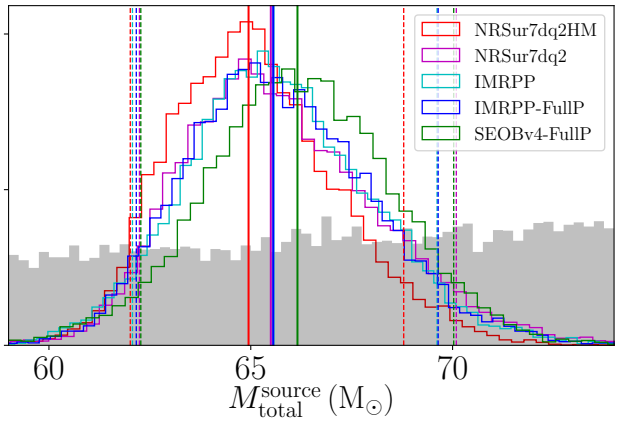

FIG. 8. Estimated total mass for GW150914, as measured in the source frame. Four different approximants are shown: `NRSur7dq2HM`, `NRSur7dq2L2`, `IMRPhenomPv2` (labeled IMRPP), and `SEOBNRv4` (labeled SEOBv4). Figure attributes are identical to Fig. 7. The shaded region shows our prior belief. Vertical dashed lines mark 90% credible regions, and vertical solid lines show the distribution median.

similar posterior distributions to what we obtain from approximate waveform models, the binary's chirp mass is estimated to be somewhat higher by both surrogate models. While this difference is marginal, it is consistent with `NRSur7dq2HM`'s estimation of luminosity distance to larger values, as the GW signal strength depends on the ratio of the two. Further, given that the distance estimated by `NRSur7dq2HM` and other waveform models differs by $\sim 15\%$, we expect the measured source-frame mass to also differ by $\delta M^{\mathrm{source}} \sim -M\delta z$ where δz is the corresponding difference in the inferred redshift of GW150914 (assuming standard cosmology [64]) between models, and M is the estimated total mass. This can be seen from Fig. 8 where we show the posterior distribution for the total mass of the binary in its source frame. We find that `NRSur7dq2HM`'s median estimate to be approximately $0.5 - 1M_\odot$ *lower* than others, which is consistent with our estimate of $-0.15 \times 0.1 \times 65 \approx -1M_\odot$ (see also Table II).

Finally, we focus on the recovery of binary spins for GW150914 in Fig. 9. The left sub-figure shows marginalized 1D and 2D posteriors for individual BH spin magnitudes (labeled $a_{1,2} \equiv |\vec{\chi}_{1,2}|$), and the right one focuses on their effective-spin $\chi_{\mathrm{eff}} := (m_1\chi_{1z} + m_2\chi_{2z})/M$ and precessing-spin χ_p [36] combinations. From the left sub-figure we note that spin recovery with `NRSur7dq2HM` and `NRSur7dq2L2` closely follows what we measure with our approximate precessing model `IMRPhenomPv2`. Restricted priors of `NRSur7dq2` have no significant effect. With `SEOBNRv4` we find spin magnitudes to be constrained along the region with $|\vec{\chi}_1| \simeq |\vec{\chi}_2|$. This is as expected given that the effective spin combination is constrained close to 0 (right sub-figure), which necessitates $\vec{\chi}_1 \simeq -\vec{\chi}_2$ for a comparable-mass binary such as this, and $\vec{\chi}_1$, $\vec{\chi}_2$ are always (anti-)parallel for `SEOBNRv4`. From the right sub-figure, we immediately note that the surrogate model does not recover any additional information about the binary's precessing spin component, as its posterior appears to be sampling the prior with little information being added by data. However, *it does constrain the source's effective-spin to be somewhat closer to zero.* This is most clearly seen by how the 1D 90% credible in-

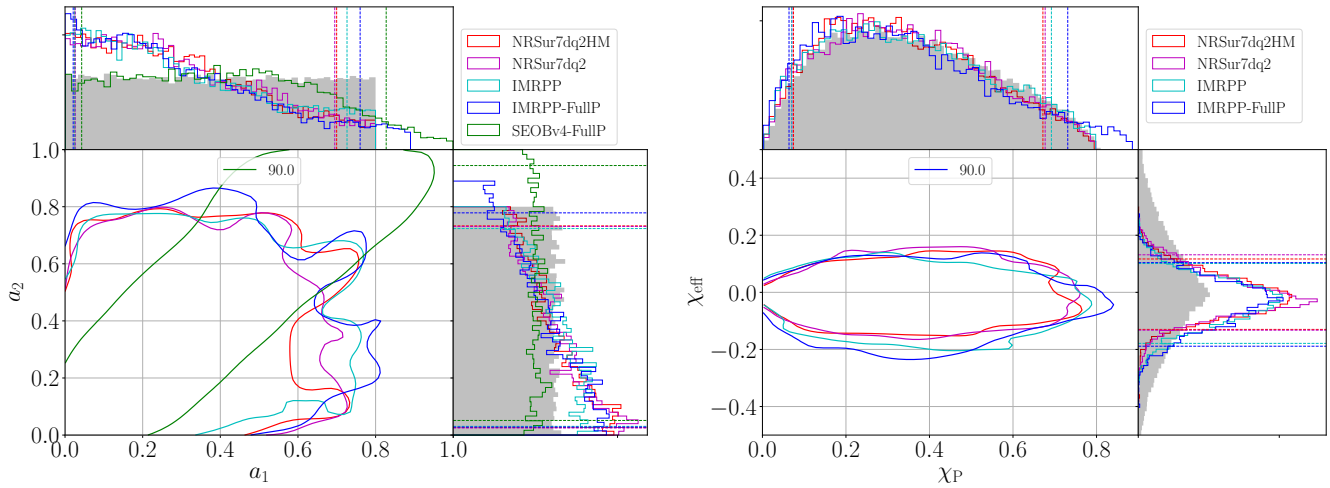

FIG. 9. Estimated spins for GW150914, using different approximants and different prior probability distributions. Shown are spin magnitudes for both BHs and the tilt angles between BH spins and the orbital angular momentum at $f_{\rm ref}$. Figure attributes are identical to Figs. 4 and 7.

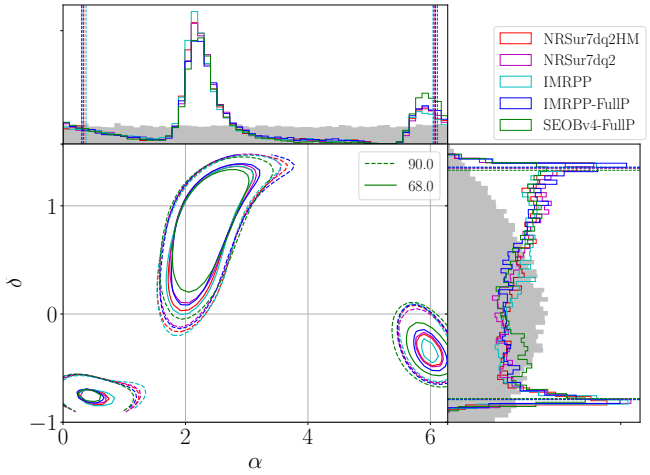

FIG. 10. Estimated sky location for GW170104, using different approximants: `NRSur7dq2HM`, `NRSur7dq2L2`, `IMRPhenomPv2` (labeled IMRPP), and `SEOBNRv4` (labeled SEOBv4). The X-FullP results correspond to an analysis with model X that allows for unrestricted mass ratios $1 \leq q \leq 8$, and spin magnitudes up to $a_{1,2} \lesssim 0.89$ for `IMRPhenomPv2` and $a_{1,2} \lesssim 0.98$ for `SEOBNRv4`. For all others, we *a priori* restrict sampling to $1 \leq q \leq 2$ and $0 \leq a_{1,2} \leq 0.8$, i.e. to the range where NR surrogate models are valid. In all panels showing 1-D posterior distributions, the shaded region shows our prior belief. Vertical dashed lines in 1-D posteriors mark 90% credible regions. The 2-D posteriors show both the 90% (dashed line) and 68% (solid line) credible regions.

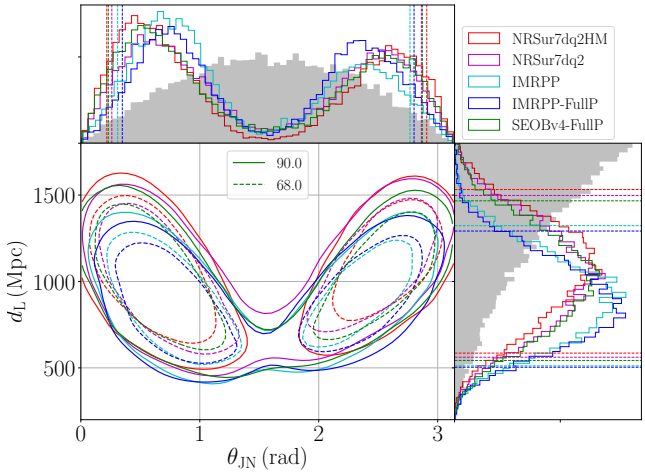

FIG. 11. Estimated source orientation / luminosity distance for GW170104, using different approximants: `NRSur7dq2HM`, `NRSur7dq2L2`, `IMRPhenomPv2` (labeled IMRPP), and `SEOBNRv4` (labeled SEOBv4). All figure attributes are identical to Fig. 10.

tervals differ between `NRSur7dq2` and `IMRPhenomPv2`. As both `NRSur7dq2HM` and `NRSur7dq2L2` provide for better estimation of $\chi_{\rm eff}$, we conclude that this may be because of additional spin information in the surrogate models that is not included in the `IMRPhenomPv2` model.

Overall, we conclude that *the use of* `NRSur7dq2` *surrogate improves the estimation of source distance and inclination for GW150914 substantially. This is primarily because of the inclusion of higher-order GW modes.* ***`NRSur7dq2`*** *also helps constrain GW150914's effective spin somewhat better (closer to zero). This appears to be because* ***`NRSur7dq2`*** *models capture effects of BH spins on quadrupolar GW emission better than approximate waveform models.* These results are further quantified in Table II, which can be directly compared with Table I of [23].

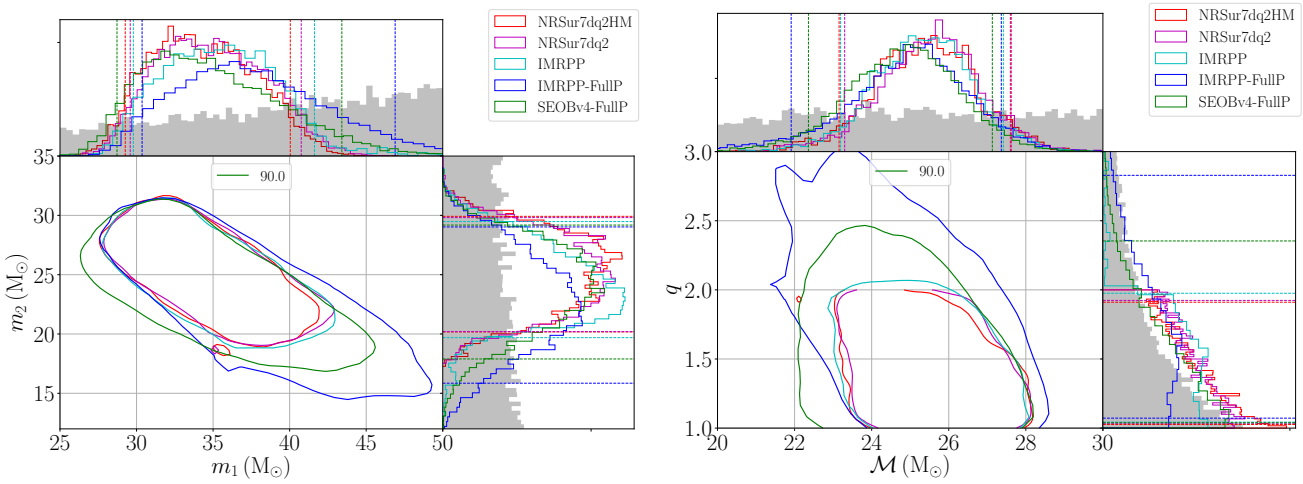

FIG. 12. Estimated masses for GW170104, using different approximants: `NRSur7dq2HM`, `NRSur7dq2L2`, `IMRPhenomPv2` (labeled IMRPP) and `SEOBNRv4` (labeled SEOBv4). The X-FullP results correspond to an analysis with model X that allows for unrestricted mass ratios $1 \leq q \leq 8$, and spin magnitudes up to $a_{1,2} \lesssim 0.89$ for `IMRPhenomPv2` and $a_{1,2} \lesssim 0.98$ for `SEOBNRv4`. For all others, we *a priori* restrict sampling to $1 \leq q \leq 2$ and $0 \leq a_{1,2} \leq 0.8$, i.e. to the range where NR surrogate models are valid. In all panels showing 1D posterior distributions, the shaded region shows our prior belief. Vertical dashed lines in 1D posteriors mark 90% credible regions. The 2D posteriors show 90% credible regions as solid contours.

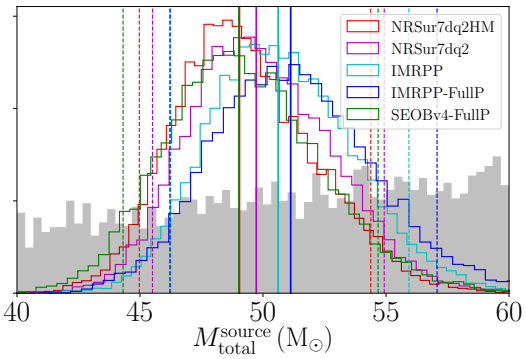

FIG. 13. Total mass for GW170104, as measured in the source frame. Four different approximants are shown: `NRSur7dq2HM`, `NRSur7dq2L2`, `IMRPhenomPv2` (labeled IMRPP), and `SEOBNRv4` (labeled SEOBv4). All figure attributes are identical to Fig. 12.

V. GW170104

The second heavy binary black hole merger was detected by the two LIGO detectors on January 4, 2017[7]. We perform identical analyses on this event as we did for GW150914. Results are shown in Figs. 10 - 14.

In Fig. 10 we show the recovery of GW170104's sky location with different models and choices of priors. We immediately note that all models yield remarkably similar estimates for its sky location, with the inclusion of higher-order modes in `NRSur7dq2HM` not yielding much additional information. In Fig. 11 we show the recovery of source's luminosity distance d_L and its initial inclination angle θ_{JN} with respect to the line of sight. Qualitatively similar to GW150914, we find that `NRSur7dq2HM` narrows the range of plausible θ_{JN} values to be closer to face-on and face-off configurations (as opposed to edge-on ones). It estimates the source to be located at a median distance of 1080Mpc which is 20% further away than what we get when using approximate precessing and non-precessing models models here (882Mpc) as well as in published LVC analysis of the event (880Mpc) [4]. As was explicitly shown for GW150914 in Fig. 6, for GW1701014 too we find that the increase in the estimated value of d_L with `NRSur7dq2HM` is not due to the model's restricted domain of validity.

Next, we show the estimation of binary mass parameters for GW1701014 using different models in Fig. 12. We note that the 2D posterior distribution for individual BH masses has support at mass ratios larger than $q = 2$ and therefore `NRSur7dq2` models only recover a fraction of the whole posterior. More specifically, it appears that `NRSur7dq2` models miss out on the low-M-high-q portion of the posterior. This would explain why even `NRSur7dq2L2` templates recover a slightly higher value for d_L than approximate models, as seen in the right panel of Fig. 11, since an increase in total mass estimate increases the estimated distance for a given signal with fixed SNR. However, the marginalized probability distributions estimated for chirp mass by all precessing models are consistent, while the non-precessing `SEOBNRv4` model constrains it less stringently. Overall we find re-

[7] The second actual detection was GW151226 [2]. In the context of this paper, this event was a "light" BBH merger, in contrast to "heavy" BBHs that we focus on here.

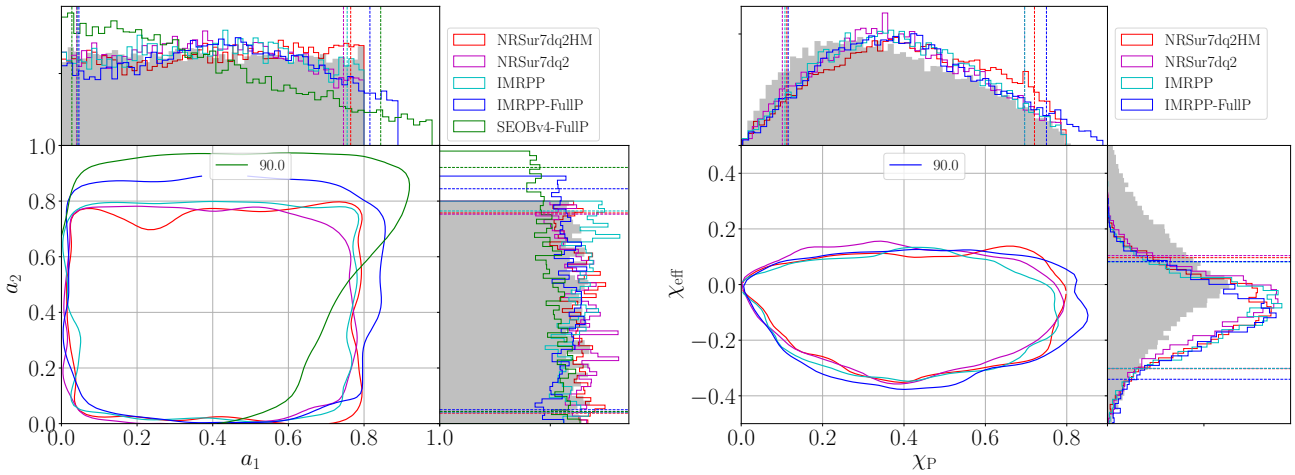

FIG. 14. Estimated spins for GW170104, using different approximants and different prior probability distributions. Shown are spin magnitudes for both BHs and the tilt angles between BH spins and the orbital angular momentum at $f_{\rm ref}$. Figure attributes are identical to Figs. 10 and 12.

sults from semi-analytic models to be consistent with NR surrogate estimates. Similar to GW150914, we expect the source-frame mass of GW170104 to be measured differently by `NRSur7dq2` than other waveform models. We confirm this through Fig. 13 where we show the posterior distribution for the total mass of the binary in its source frame. We find that `NRSur7dq2HM`'s median estimate to be approximately $1M_\odot$ lower than others, see Table II for other mass parameters measured in source frame.

In Fig. 14 we demonstrate how well we estimate component BH spins for GW170104 with the NR surrogate model, and compare it with what we get from semi-analytic ones. In the left sub-figure we show the estimation of spin magnitudes, while in the right sub-figure we show the same for the effective spin $\chi_{\rm eff}$ and precessing spin χ_p combinations. While all models estimate $\chi_{\rm eff} \simeq -0.1$ for this event, they recover little information for either $a_{1,2}$ or χ_p, with their respective 1D posteriors following closely their sampling priors. For all spin combinations considered, we note that the recovery from all models is remarkably similar, despite the additionally restricted priors of `NRSur7dq2` models. For this signal, therefore, the `NRSur7dq2` and approximate models provide essentially identical spin information.

Overall, we conclude that with the NR surrogate model `NRSur7dq2` we estimate the source of GW170104 to be approximately 20% further away than was previously estimated using semi-analytic waveform models [4]. The same surrogate furnishes little extra information for source mass and spin parameters of GW170104 though. Our results are summarized in Table II.

VI. DISCUSSION

The population of binary black hole mergers that the LIGO-Virgo detector network has observer thus far comprises of many loud signals coming from "heavy" black hole binaries, with each hole measuring around $20-30$ times the mass of the Sun [11]. Coalescing binaries of such heavy black holes radiate gravitational waves at lower frequencies than their lighter counterparts, and therefore enter the sensitive frequency band of current GW detectors only a few orbits before they merge. During this pre-merger period of the two-body evolution that is visible to LIGO-Virgo, inspiraling binaries' orbits evolve rapidly from being well approximated as a sequence of slowly shrinking spheres (or circles) to being highly dynamical as both holes enter each other's strong-field regions at highly relativistic velocities. Describing their motion in the pre-merger regime, and consequently the form of emitted gravitational radiation, is beyond the reach of traditional perturbative methods that typically rely on the dynamical timescale of gravity being large and/or binary motion being non-relativistic.

Fully numerical solutions of nonlinear Einstein equations is the most powerful (and only) approach that can tackle the physics in the pre-merger regime. This, however, comes at a non-trivial computational cost that precludes performing numerical simulations for an arbitrary number of binary mergers. With present day technology and budgets, it is possible to perform approximately $\mathcal{O}(10^3)$ simulations in a calendar year. However, when trying to determine the physical parameters of the source of a BBH merger event from its observed GW data, one typically needs to matched-filter the data

TABLE II. Summary of parameters that characterize GW150914 and GW170104. For model parameters we report the median value as well as the range of the symmetric 90% credible interval [65]; where useful, we also quote 90% credible bounds. The source redshift and source-frame masses assume standard cosmology [64]. The spin-aligned `SEOBNRv4` (labeled SEOB) and precessing `IMRPhenomPv2` (labeled IMRPP) waveform models are described in the text, as is the NR surrogate labeled here NR22 (`NRSur7dq2L2`) and NRHM (`NRSur7dq2HM`). Results for the effective precession spin parameter $\chi_{\rm p}$ used by `IMRPhenomPv2` are not shown as we effectively recover the prior; see left panels of Figs. 9 and 14. The `SEOBNRv4/IMRPhenomPv2` values stated here are directly comparable to Table I of [23] for GW150914 and to Table I of [4] for GW170104, and are broadly consistent with published LIGO analyses.

	GW150914 SEOB / IMRPP / NR22 / NRHM	GW170104 SEOB / IMRPP / NR22 / NRHM
Detector-frame total mass $M/M_\odot$	$72.1^{+3.5}_{-4.0}$ / $71.6^{+4.1}_{-3.8}$ / $71.9^{+5.1}_{-3.1}$ / $72.2^{+4.8}_{-3.1}$	$58.8^{+5.8}_{-5.4}$ / $60.2^{+5.7}_{-5.3}$ / $59.9^{+4.6}_{-5.3}$ / $59.5^{+4.8}_{-4.6}$
Detector-frame chirp mass $\mathcal{M}/M_\odot$	$31.2^{+1.5}_{-1.8}$ / $30.9^{+1.9}_{-1.8}$ / $31.0^{+2.4}_{-1.5}$ / $31.2^{+2.0}_{-1.5}$	$24.9^{+2.2}_{-2.5}$ / $25.1^{+2.3}_{-3.0}$ / $25.5^{+2.1}_{-2.4}$ / $25.4^{+2.2}_{-2.1}$
Detector-frame primary mass $m_1/M_\odot$	$39.0^{+5.7}_{-3.5}$ / $38.7^{+5.2}_{-3.5}$ / $39.1^{+5.1}_{-3.3}$ / $38.9^{+4.8}_{-3.2}$	$34.3^{+9.1}_{-5.7}$ / $37.0^{+9.7}_{-6.6}$ / $34.9^{+5.8}_{-5.3}$ / $34.1^{+5.9}_{-4.8}$
Detector-frame secondary mass $m_2/M_\odot$	$32.9^{+3.2}_{-4.8}$ / $32.9^{+3.4}_{-5.1}$ / $33.1^{+3.5}_{-5.0}$ / $33.5^{+3.2}_{-5.2}$	$24.1^{+5.1}_{-6.1}$ / $22.7^{+6.3}_{-6.6}$ / $24.7^{+4.9}_{-4.7}$ / $25.4^{+4.5}_{-5.1}$
Detector-frame final mass $M_{\rm f}/M_\odot$	$68.7^{+3.1}_{-3.6}$ / $68.2^{+3.7}_{-3.4}$ / $68.5^{+4.5}_{-2.8}$ / $68.8^{+4.2}_{-2.7}$	$56.3^{+5.4}_{-4.9}$ / $57.8^{+5.7}_{-4.9}$ / $57.3^{+4.3}_{-4.8}$ / $56.9^{+4.4}_{-4.2}$
Source-frame total mass $M^{\rm source}/M_\odot$	$66.2^{+3.9}_{-3.9}$ / $65.6^{+4.1}_{-3.4}$ / $65.5^{+4.6}_{-3.3}$ / $64.9^{+3.8}_{-2.9}$	$49.0^{+5.7}_{-4.7}$ / $51.1^{+5.9}_{-4.9}$ / $49.7^{+5.2}_{-4.2}$ / $49.0^{+5.2}_{-3.9}$
Source-frame chirp mass $\mathcal{M}^{\rm source}/M_\odot$	$28.6^{+1.6}_{-1.7}$ / $28.3^{+1.8}_{-1.5}$ / $28.3^{+2.0}_{-1.5}$ / $28.1^{+1.7}_{-1.4}$	$20.7^{+2.2}_{-2.0}$ / $21.2^{+2.3}_{-2.3}$ / $21.2^{+2.2}_{-1.9}$ / $20.9^{+2.4}_{-1.7}$
Source-frame primary mass $m_1^{\rm source}/M_\odot$	$35.8^{+5.4}_{-3.3}$ / $35.5^{+5.0}_{-3.2}$ / $35.6^{+4.6}_{-3.2}$ / $35.0^{+4.5}_{-2.9}$	$28.6^{+8.3}_{-4.8}$ / $31.5^{+9.0}_{-6.0}$ / $29.0^{+5.6}_{-4.6}$ / $28.3^{+4.8}_{-4.3}$
Source-frame secondary mass $m_2^{\rm source}/M_\odot$	$30.2^{+3.2}_{-4.4}$ / $30.2^{+3.1}_{-4.6}$ / $30.1^{+3.3}_{-4.5}$ / $30.1^{+3.0}_{-4.6}$	$20.0^{+4.4}_{-4.8}$ / $19.2^{+5.4}_{-5.2}$ / $20.6^{+4.4}_{-3.8}$ / $20.9^{+4.3}_{-4.2}$
Source-fame final mass $M_{\rm f}^{\rm source}/M_\odot$	$63.0^{+3.5}_{-3.5}$ / $62.5^{+3.7}_{-3.0}$ / $62.4^{+4.1}_{-3.0}$ / $61.9^{+3.5}_{-2.7}$	$46.9^{+5.5}_{-4.4}$ / $49.0^{+5.9}_{-4.6}$ / $47.6^{+4.9}_{-3.9}$ / $46.9^{+4.8}_{-3.7}$
Mass ratio q	$1.2^{+0.4}_{-0.2}$ / $1.2^{+0.4}_{-0.2}$ / $1.2^{+0.4}_{-0.2}$ / $1.2^{+0.4}_{-0.1}$	$1.4^{+0.9}_{-0.4}$ / $1.6^{+1.1}_{-0.6}$ / $1.4^{+0.5}_{-0.4}$ / $1.3^{+0.6}_{-0.3}$
Effective inspiral spin parameter $\chi_{\rm eff}$	$-0.01^{+0.11}_{-0.15}$ / $-0.03^{+0.14}_{-0.16}$ / $-0.02^{+0.15}_{-0.12}$ / $-0.01^{+0.13}_{-0.12}$	$-0.12^{+0.21}_{-0.28}$ / $-0.10^{+0.19}_{-0.23}$ / $-0.07^{+0.17}_{-0.24}$ / $-0.09^{+0.19}_{-0.20}$
Dimensionless primary spin mag. a_1	$0.40^{+0.43}_{-0.36}$ / $0.26^{+0.50}_{-0.24}$ / $0.27^{+0.43}_{-0.24}$ / $0.27^{+0.43}_{-0.25}$	$0.32^{+0.53}_{-0.29}$ / $0.42^{+0.40}_{-0.37}$ / $0.38^{+0.36}_{-0.34}$ / $0.41^{+0.35}_{-0.37}$
Dimensionless secondary spin mag. a_2	$0.50^{+0.44}_{-0.45}$ / $0.30^{+0.48}_{-0.28}$ / $0.29^{+0.44}_{-0.27}$ / $0.30^{+0.44}_{-0.27}$	$0.45^{+0.47}_{-0.41}$ / $0.46^{+0.39}_{-0.41}$ / $0.38^{+0.38}_{-0.34}$ / $0.39^{+0.37}_{-0.35}$
Final spin $a_{\rm f}$	$0.69^{+0.04}_{-0.07}$ / $0.67^{+0.05}_{-0.06}$ / $0.68^{+0.05}_{-0.05}$ / $0.68^{+0.04}_{-0.05}$	$0.62^{+0.10}_{-0.15}$ / $0.61^{+0.08}_{-0.18}$ / $0.64^{+0.06}_{-0.10}$ / $0.64^{+0.07}_{-0.09}$
Luminosity distance $D_{\rm L}$/Mpc	$423.4^{+178.0}_{-189.7}$ / $434.1^{+149.2}_{-182.7}$ / $472.0^{+167.5}_{-193.2}$ / $538.5^{+140.2}_{-181.3}$	$1000.9^{+467.0}_{-457.9}$ / $882.3^{+407.6}_{-376.9}$ / $1016.4^{+469.8}_{-456.3}$ / $1079.3^{+441.3}_{-487.1}$
Source redshift z	$0.09^{+0.03}_{-0.04}$ / $0.09^{+0.03}_{-0.04}$ / $0.10^{+0.03}_{-0.04}$ / $0.11^{+0.03}_{-0.04}$	$0.20^{+0.08}_{-0.09}$ / $0.18^{+0.07}_{-0.07}$ / $0.20^{+0.08}_{-0.08}$ / $0.21^{+0.07}_{-0.09}$
Upper bound on primary spin mag. a_1	0.74 / 0.65 / 0.62 / 0.62	0.85 / 0.82 / 0.75 / 0.76
Upper bound on secondary spin mag. a_2	0.94 / 0.78 / 0.73 / 0.73	0.92 / 0.85 / 0.75 / 0.76
Upper bound on mass ratio q	1.56 / 1.55 / 1.53 / 1.52	2.34 / 2.79 / 1.93 / 1.90

against $\mathcal{O}(10^{6-8})$ distinct GW templates[8]. Therefore there is a large gap between the demand of matched-filtering templates and their availability through direct numerical simulations. This gap is traditionally bridged by introducing *phenomenological* extensions to perturbative waveform models, and *calibrating* these extensions to agree with numerical simulations where they can. Examples of such models would include those within the Effective-one-body family [48] and the phenomenological family [19]. Although these models now span a fair region of the full 7-dimensional parameter space that describes arbitrary binary black hole coalescences, there is always scope for inaccuracies in one corner or another [25, 26]. An ab-initio more accurate and reliable approach would be to develop a 7-dimensional numerical interpolant (a *surrogate* model) for the gravitational-wave strain based on select numerical relativity simulations. Such a surrogate model would have been too expensive to construct in the past. With improvements to numerical relativity technology in recent years, Blackman et al [32, 33] developed the first usable surrogate models based solely on numerical relativity simulations. They span a restricted region of the full 7-D binary black hole parameter space, but within that region the model describes arbitrary precessing binary orbits to NR-level accuracy.

The primary purpose of this paper is to use the numerical relativity-based surrogate model `NRSur7dq2` [33] in a fully Bayesian framework and demonstrate both its viability and efficacy in estimating source parameters from heavy binary black hole merger signals. This work paves

[8] Recently proposed grid-based methods [28, 34] can recover a *subset* of binary parameters with much fewer ($\mathcal{O}(10^3)$) template evaluations. However, approximations used in these methods include interpolating Bayesian likelihood on unstructured high-dimensional grids. The (physical and technical) impact of these approximations still needs to be quantified more thoroughly.

the way for future surrogate models that will gradually span the entire parameter space for most GW events. We also use the same surrogate [33] to re-analyze data from the first two heavy BBH merger events: GW150914 and GW170104. While we find improvements in the precision of measuring mass and spin parameters for these events' source binaries, our primary finding is that both of these events were located $15-20\%$ further away than what approximate waveform-model based analyses have found, including published LIGO-Virgo results [4, 23].

We first perform controlled tests by injecting synthetic GW signals into zero noise, reconstructing their parameters using `NRSur7dq2` templates as filters, and comparing the parameter recovery with both the true parameters as well as what other waveform models furnish. We perform a total of 48 such injections that are described in Table I. The injected source parameters are varied as follows: mass ratio takes on values $\in \{1.2, 1.5\}$, while the total mass is fixed to $60M_{\odot}$; source spins are chosen from four precessing configurations; source distance is varied over $\{500, 1000, 1500\}$Mpc, while source inclination is allowed two values - one nearly face-on and the other nearly edge-on; and sky location angles are chosen uniformly over a 2-sphere. We use the full `NRSur7dq2HM` to model synthetic signals, and use both `NRSur7dq2L2` and `NRSur7dq2HM`, in addition to `IMRPhenomPv2` (a phenomenological model for spin-precessing binaries with an effective description of spin degrees of freedom), as filter templates. We use `IMRPhenomPv2` in two configurations: first, where templates are artificially restricted to be sampled with the same prior restrictions for BBH parameters as numerical surrogates, and second, where they can be sampled freely. We find that both total mass and mass ratio are better recovered by both `NRSur7dq2L2` and `NRSur7dq2HM` templates than by `IMRPhenomPv2`. This is noticeable in Fig. 3. For BBH spins, we find that all models produce broadly consistent results, with the effective spin being measured somewhat more accurately by `NRSur7dq2`. For all other intrinsic BBH parameters, including the chirp mass, the `NRSur7dq2` templates furnish results that are broadly consistent with those from `IMRPhenomPv2`. Amongst extrinsic parameters, we find that the degeneracy in measuring source distance and orbital inclination is largely reduced by the addition of $\ell = \{3, 4\}$ multipoles in `NRSur7dq2HM`, and with them we can recover both of these parameters better than all other template models, including `NRSur7dq2L2`. We find this improvement to be especially pronounced when the source is highly inclined to the line of sight and therefore emits more strongly in $\ell = \{3, 4\}$ GW modes. Past work applying higher-mode information from NR to GW parameter estimation could only use it to measure a subset of source parameters, and relied on the interpolation of Bayesian likelihood on unstructured grids [27, 28, 34]. We point out that our tests described above are the first *comprehensive* usage of higher-mode information from NR without additional approximations. Overall, we observe that `NRSur7dq2` templates improve estimation of BBH parameters both with and without the inclusion of $\ell > 2$ GW modes. This is understandable since it models both the leading and sub-leading order GW modes more accurately than approximate GW models [33], while most approximate models do not yet include $\ell > 2$ GW modes. Most of this improvement is moderate, however, and we expect it to be more pronounced when the signals themselves have either a larger relative contribution from $l \neq 2$ modes, such as for binaries with higher mass ratios, or their sources have larger spin magnitudes[9]. Our results therefore provide strong motivation to extend the `NRSur7dq2` model to span a larger range of binary mass ratios and black hole spins.

From these controlled tests, we establish the viability of using NR information directly in a traditional Bayesian parameter estimation framework for GW events. We next proceed to analyze the first BBH merger event ever to be recorded: GW150914. We analyze it with our NR surrogates - `NRSur7dq2HM` and `NRSur7dq2L2` - in addition to `IMRPhenomPv2` and `SEOBNRv4`. The latter two (or their variants) were used in the original published analyses for this binary [23]. We find that with `NRSur7dq2HM` we place the source of this event to be at a luminosity distance of $\simeq 530$Mpc, which is about 25% further away than what other models estimate (including previous LVC analysis of the event [23]). If we remove $\ell = \{3, 4\}$ modes and restrict the analysis to `NRSur7dq2L2`, we find that the measured luminosity distance agrees with the originally estimated value, indicating clearly that this new information is extracted by the sub-dominant waveform multipoles in `NRSur7dq2HM`. Simultaneously, `NRSur7dq2HM` also helps narrow down the allowed inclination configurations for the source to be either face-on or face-off with more confidence than earlier. Both of these improvements can be seen from Fig. 5. As would be consistent with a larger luminosity distance, the NR surrogates estimate the chirp mass of GW150914 to be marginally higher than what approximate model estimate. Consistent with a larger redshift, `NRSur7dq2HM` estimates GW150914's mass in its source frame to be approximately 1% lower than other models. Finally, with full GR information implicitly contained within them, both `NRSur7dq2L2` and `NRSur7dq2HM` constrain the effective-spin of GW150914 more tightly around $\chi_{\rm eff} \simeq 0$ than previous estimates. Components of spin that are orthogonal to the orbit and that contribute to its precession are not constrained much better than phenomenological models, and this is as expected because for short signals there is simply not enough time for the binary to complete a few precession cycles. However, measurement of BH spins can be sensitive to the choice of sampling priors employed [34, 40]. We defer an investigation of their effect on spin inferences with `NRSur7dq2HM` to future work.

[9] In the latter case, however, the sampling priors imposed on spins can alter their estimation appreciably [40] and must be carefully chosen.

Finally, we move on to the second heavy BBH merger event: GW170104. This event differs from GW150914 in the sense that its measured posterior probability distributions (by approximate models) for binary masses have support outside the domain of validity of `NRSur7dq2HM`, specifically for binaries with mass ratios $q > 2$. In practice, however, we find that this restriction does not bias the recovery of other parameters by `NRSur7dq2HM` in any noticeable manner. Similar to GW150914, `NRSur7dq2HM` constrains the luminosity distance to this event to be approximately 20% larger than what approximate models that include only the dominant $\ell = |m| = 2$ modes give [4]. The orientation of this source is constrained more tightly around face-on or face-off configurations, with edge-on configurations being strongly disfavored. The estimation of mass parameters is consistent between `NRSur7dq2` and other models, although the former can only recover a partial posterior distribution for mass ratio. Lastly, we note that the estimation of spin parameters by `NRSur7dq2HM` remains remarkably similar to what we get from `IMRPhenomPv2` and `SEOBNRv4`, and is therefore consistent with them. The effective spin for the event is the only well-measured spin parameter. It is constrained to be small but negative by all models. For all other spin combinations, all models essentially recover the sampling prior as the posterior, with data adding little information.

From our results it is clear that there are certainly advantages of using numerical relativity surrogates for following up heavy binary black hole merger events. One of them is the inclusion of accurate $\ell > 2$ modes in `NRSur7dq2HM`, which facilitates the resolution of the luminosity distance - inclination angle degeneracy. This degeneracy often leads to systematic bias in providing point estimates of the distance to GW sources, which subsequently percolates to the calculations of astrophysical binary merger rates [66], estimation of Hubble's constant [67], etc. These applications could therefore potentially benefit from `NRSur7dq2HM`-based follow-ups of GW events. Another benefit comes as improvement in the measurement of BBH masses by `NRSur7dq2` (both `NRSur7dq2HM` and `NRSur7dq2L2`) as it captures even the quadrupolar GW mode more accurately than approximate models. The primary disadvantage of using `NRSur7dq2` templates for GW event follow-up is its limited domain of applicability. One of the lines of active research we are currently pursuing is to extend the domain of `NRSur7dq2` to model binary emitters with mass ratios $q > 2$. Finally, we note that the computational cost of using `NRSur7dq2HM` in parameter estimation is approximately $3 - 4\times$ the cost of using the frequency-domain `IMRPhenomPv2` / reduced order model for `SEOBNRv4`. A bulk of this extra cost is due to the extra Fourier transform required to transform each time-domain surrogate template to frequency-domain. This extra cost can, however, be mitigated in two ways. The first is to develop a reduced order model for the surrogate, along the lines of [30]. Another is using a recently developed *rapid* parameter estimation scheme [34] with `NRSur7dq2HM` templates.

Our results are encouraging and we propose for `NRSur7dq2` and its follow-up models to used in standard GW event follow-up analyses in order to maximize the science output from GW detector data. We provide full posterior samples (as supplemental materials) from Bayesian parameter estimation of LIGO/Virgo data for GW150914 and GW170104, with `NRSur7dq2HM` and `IMRPhenomPv2`, to enable further analysis by the community. These can be obtained from https://github.com/prayush/GW150914_GW170104_NRSur7dq2_Posteriors.

ACKNOWLEDGMENTS

We gratefully acknowledge support for this research at Cornell from the Sherman Fairchild Foundation and NSF grant PHY-1606654; at Caltech from the Sherman Fairchild Foundation and NSF grant PHY-1404569; at CITA from NSERC of Canada, the Ontario Early Researcher Awards Program, the Canada Research Chairs Program, and the Canadian Institute for Advanced Research; and at Princeton from NSF grant PHY-1305682 and the Simons Foundation. SEF was partially supported by NSF grant PHY-1806665. PK would like to thank FAPESP grant 2016/01343-7 for hospitality during his visits to ICTP-SAIFR where part of this work was completed. This research has made use of data, software and/or web tools obtained from the LIGO Open Science Center (https://losc.ligo.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 (CNRS), the Italian Istituto Nazionale della Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by Polish and Hungarian institutes.

[1] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].
[2] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 241103 (2016), arXiv:1606.04855 [gr-qc].
[3] B. P. Abbott *et al.* (Virgo, LIGO Scientific), Phys. Rev. **X6**, 041015 (2016), arXiv:1606.04856 [gr-qc].
[4] B. P. Abbott *et al.* (VIRGO, LIGO Scientific), Phys. Rev. Lett. **118**, 221101 (2017), arXiv:1706.01812 [gr-qc].
[5] B. P. Abbott *et al.* (Virgo, LIGO Scientific), Phys. Rev. Lett. **119**, 141101 (2017), arXiv:1709.09660 [gr-qc].

[6] S. A. Usman *et al.*, Class. Quant. Grav. **33**, 215004 (2016), arXiv:1508.02357 [gr-qc].

[7] K. Cannon *et al.*, Astrophys. J. **748**, 136 (2012), arXiv:1107.2665 [astro-ph.IM].

[8] S. Privitera, S. R. P. Mohapatra, P. Ajith, K. Cannon, N. Fotopoulos, M. A. Frei, C. Hanna, A. J. Weinstein, and J. T. Whelan, Phys. Rev. **D89**, 024003 (2014), arXiv:1310.5633 [gr-qc].

[9] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C.-J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O'Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. Van Der Sluys, A. Vecchio, W. Vousden, and L. Wade, Phys. Rev. D **91**, 042003 (2015).

[10] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 221101 (2016), arXiv:1602.03841 [gr-qc].

[11] B. P. Abbott *et al.* (Virgo, LIGO Scientific), (2018), arXiv:1811.12907 [astro-ph.HE].

[12] F. Pretorius, Phys. Rev. Lett. **95**, 121101 (2005), arXiv:gr-qc/0507014 [gr-qc].

[13] F. Pretorius and M. W. Choptuik, J. Comput. Phys. **218**, 246 (2006), gr-qc/0508110.

[14] F. Pretorius, Class. Quantum Grav. **22**, 425 (2005), gr-qc/0407110.

[15] M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower, Phys. Rev. Lett. **96**, 111101 (2006), arXiv:gr-qc/0511048 [gr-qc].

[16] J. G. Baker, J. Centrella, D.-I. Choi, M. Koppitz, and J. van Meter, Phys. Rev. Lett. **96**, 111102 (2006), arXiv:gr-qc/0511103 [gr-qc].

[17] J. G. Baker, M. Campanelli, F. Pretorius, and Y. Zlochower, Class. Quantum Grav. **24**, S25 (2007), gr-qc/0701016.

[18] A. Buonanno, Y. Pan, H. P. Pfeiffer, M. A. Scheel, L. T. Buchman, and L. E. Kidder, Phys. Rev. D **79**, 124028 (2009), arXiv:0902.0790 [gr-qc].

[19] P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, J. T. Whelan, B. Brügmann, P. Diener, J. Gonzalez, M. H. S. Husa, M. Koppitz, D. Pollney, L. Rezzolla, L. Santamaría, A. M. Sintes, U. Sperhake, and J. Thornburg, Class. Quantum Grav. **24**, S689 (2007), arXiv:0704.3764 [gr-qc].

[20] C. Pankow, P. Brady, E. Ochsner, and R. O'Shaughnessy, Phys. Rev. **D92**, 023002 (2015), arXiv:1502.04370 [gr-qc].

[21] B. P. Abbott *et al.* (Virgo, LIGO Scientific), (2018), arXiv:1811.12940 [astro-ph.HE].

[22] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Purrer, X. Jimenez Forteza, and A. Bohe, Phys. Rev. **D93**, 044007 (2016), arXiv:1508.07253 [gr-qc].

[23] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), Phys. Rev. Lett. **116**, 241102 (2016), arXiv:1602.03840 [gr-qc].

[24] M. Boyle, D. A. Brown, L. E. Kidder, A. H. Mroue, H. P. Pfeiffer, M. A. Scheel, G. B. Cook, and S. A. Teukolsky, Phys. Rev. **D76**, 124038 (2007), arXiv:0710.0158 [gr-qc].

[25] P. Kumar, T. Chu, H. Fong, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, M. A. Scheel, and B. Szilágyi, Phys. Rev. **D93**, 104050 (2016), arXiv:1601.05396 [gr-qc].

[26] 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), arXiv:1507.00103 [gr-qc].

[27] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. **D94**, 064035 (2016), arXiv:1606.01262 [gr-qc].

[28] J. Lange *et al.*, Phys. Rev. **D96**, 104041 (2017), arXiv:1705.09833 [gr-qc].

[29] S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio, Phys. Rev. X **4**, 031006 (2014), arXiv:1308.3565 [gr-qc].

[30] M. Pürrer, Class. Quantum Grav. **31**, 195010 (2014), arXiv:1402.4146 [gr-qc].

[31] 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), arXiv:1502.07758 [gr-qc].

[32] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, Phys. Rev. **D95**, 104023 (2017), arXiv:1701.00550 [gr-qc].

[33] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, C. D. Ott, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and B. Szilagyi, Phys. Rev. **D96**, 024058 (2017), arXiv:1705.07089 [gr-qc].

[34] J. Lange, R. O'Shaughnessy, and M. Rizzo, (2018), arXiv:1805.10457 [gr-qc].

[35] P. Schmidt, M. Hannam, S. Husa, and P. Ajith, Phys. Rev. D **84**, 024046 (2011), arxiv:1012.2879.

[36] P. Schmidt, F. Ohme, and M. Hannam, Phys. Rev. D **91**, 024043 (2015), arXiv:1408.1810 [gr-qc].

[37] L. Pekowsky, J. Healy, D. Shoemaker, and P. Laguna, "Impact of higher-order modes on the detection of binary black hole coalescences," (2012), arXiv:1210.1891 [gr-qc].

[38] J. Healy, P. Laguna, L. Pekowsky, and D. Shoemaker, (2013), 10.1103/PhysRevD.88.024034, arXiv:1302.6953 [gr-qc].

[39] J. Calderon Bustillo, S. Husa, A. M. Sintes, and M. Purrer, Phys. Rev. **D93**, 084019 (2016), arXiv:1511.02060 [gr-qc].

[40] K. Chatziioannou, G. Lovelace, M. Boyle, M. Giesler, D. A. Hemberger, R. Katebi, L. E. Kidder, H. P. Pfeiffer, M. A. Scheel, and B. Szilagyi, Phys. Rev. **D98**, 044028 (2018), arXiv:1804.03704 [gr-qc].

[41] "The Spectral Einstein Code," `http://www.black-holes.org/SpEC.html`.

[42] D. Gerosa, F. Hebert, and L. C. Stein, Phys. Rev. **D97**, 104049 (2018), arXiv:1802.04276 [gr-qc].

[43] C. Talbot, E. Thrane, P. D. Lasky, and F. Lin, (2018), arXiv:1807.00990 [astro-ph.HE].

[44] S. A. Smolyak, in *Dokl. Akad. Nauk SSSR*, Vol. 4 (1963) p. 123.

[45] H.-J. Bungartz and M. Griebel, Acta numerica **13**, 147 (2004).

[46] M. Boyle, R. Owen, and H. P. Pfeiffer, Phys. Rev. D **84**, 124011 (2011), arXiv:1110.2965 [gr-qc].

[47] R. O'Shaughnessy, B. Vaishnav, J. Healy, Z. Meeks, and D. Shoemaker, Phys. Rev. D **84**, 124002 (2011), arXiv:1109.5224.

[48] A. Buonanno and T. Damour, Phys. Rev. D **59**, 084006 (1999), arXiv:gr-qc/9811091 [gr-qc].

[49] L. Blanchet, Living Rev. Rel. **17**, 2 (2014), arXiv:1310.1528 [gr-qc].

[50] K. D. Kokkotas and B. G. Schmidt, Living Rev. Rel. **2** (1999), 2.

[51] A. Bohé, L. Shao, A. Taracchini, A. Buonanno, S. Babak, I. W. Harry, I. Hinder, S. Ossokine, M. Pürrer, V. Raymond, T. Chu, H. Fong, P. Kumar, H. P. Pfeiffer,

M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **95**, 044028 (2017), arXiv:1611.03703 [gr-qc].
[52] L. S. Collaboration, "LSC Algorithm Library software packages LAL, LALWRAPPER, and LALAPPS," .
[53] Y. Pan, A. Buonanno, A. Taracchini, L. E. Kidder, A. H. Mroue, H. P. Pfeiffer, M. A. Scheel, and B. Szilagyi, Phys. Rev. **D89**, 084006 (2014), arXiv:1307.6232 [gr-qc].
[54] S. Vitale, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, I. Mandel, B. Aylott, and J. Veitch, Phys. Rev. **D85**, 064034 (2012), arXiv:1111.3044 [gr-qc].
[55] T. B. Littenberg, M. Coughlin, B. Farr, and W. M. Farr, Phys. Rev. D **88**, 084044 (2013), arXiv:1307.8195 [astro-ph.IM].
[56] G. M. Harry (LIGO Scientific Collaboration), Class. Quantum Grav. **27**, 084006 (2010).
[57] J. Aasi *et al.* (LIGO Scientific Collaboration), Class. Quantum Grav. **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].
[58] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, *Proceedings, 10th International LISA Symposium: Gainesville, Florida, USA, May 18-23, 2014*, J. Phys. Conf. Ser. **610**, 012021 (2015), arXiv:1410.4839 [gr-qc].
[59] J. Skilling, AIP Conference Proceedings **735**, 395 (2004), https://aip.scitation.org/doi/pdf/10.1063/1.1835238.
[60] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Class. Quant. Grav. **34**, 104002 (2017), arXiv:1611.07531 [gr-qc].
[61] T. D. Abbott *et al.* (Virgo, LIGO Scientific), Phys. Rev. **X6**, 041014 (2016), arXiv:1606.01210 [gr-qc].
[62] P. B. Graff, A. Buonanno, and B. S. Sathyaprakash, Phys. Rev. D **92**, 022002 (2015), arXiv:1504.04766 [gr-qc].
[63] L. London, S. Khan, E. Fauchon-Jones, C. Garcia, M. Hannam, S. Husa, X. Jimenez-Forteza, C. Kalaghatgi, F. Ohme, and F. Pannarale, Phys. Rev. Lett. **120**, 161102 (2018), arXiv:1708.00404 [gr-qc].
[64] P. A. R. Ade *et al.* (Planck), Astron. Astrophys. **594**, A13 (2016), arXiv:1502.01589 [astro-ph.CO].
[65] J. Aasi *et al.* (the LIGO Scientific Collaboration, the Virgo Collaboration), Phys. Rev. D **88**, 062001 (2013), arXiv:1304.1775 [gr-qc].
[66] B. P. Abbott *et al.* (Virgo, LIGO Scientific), Astrophys. J. **833**, L1 (2016), arXiv:1602.03842 [astro-ph.HE].
[67] A. Nishizawa, Phys. Rev. **D96**, 101303 (2017), arXiv:1612.06060 [astro-ph.CO].