Gravitational wave inference on a numerical-relativity simulation of a black hole merger beyond general relativity

Maria Okounkova[1,2,*] Maximiliano Isi[1,†] Katerina Chatziioannou[3,4,‡] and Will M. Farr[5,1,§]

[1]*Center for Computational Astrophysics, Flatiron Institute, 162 5th Ave, New York, New York 10010, USA*
[2]*Department of Physics, Pasadena City College, Pasadena, California 91106, USA*
[3]*Department of Physics, California Institute of Technology, Pasadena, California 91125, USA*
[4]*LIGO Laboratory, California Institute of Technology, Pasadena, California 91125, USA*
[5]*Department of Physics and Astronomy, Stony Brook University, Stony Brook New York 11794, USA*

We apply common gravitational wave inference procedures on binary black hole merger waveforms beyond general relativity. We consider dynamical Chern-Simons gravity, a modified theory of gravity with origins in string theory and loop quantum gravity. This theory introduces an additional parameter ℓ, corresponding to the length-scale below which beyond-general-relativity effects become important. We simulate data based on numerical relativity waveforms produced under an approximation to this theory, which differ from those of general relativity in the strongly nonlinear merger regime. We consider a system with parameters similar to GW150914 with different values of ℓ and signal-to-noise ratios. We perform two analyses of the simulated data. The first is a template-based analysis that uses waveforms derived under general relativity and allows us to identify degeneracies between the two waveform morphologies. The second is a morphology-independent analysis based on BayesWave that does not assume that the signal is consistent with general relativity. The BayesWave analysis faithfully reconstructs the simulated signals. However, waveform models derived under general relativity are unable to fully mimic the simulated modified-gravity signals and such a deviation would be identifiable with existing inference tools. Depending on the magnitude of the deviation, we find that the templated analysis can under perform the morphology-independent analysis in fully recovering simulated beyond-GR waveforms even for achievable signal-to-noise ratios $\gtrsim 20$–30.

I. INTRODUCTION

Although general relativity (GR) has passed all precision tests to date, it is expected to break down at some scale where gravity is reconciled with quantum mechanics through a beyond-GR theory of gravity. Moreover, considering possible modifications to GR and testing for them can help shed light on the theory itself. Binary black hole (BBH) mergers can probe gravity at its most extreme: in the strong-field, highly nonlinear, highly-dynamical regime, and are therefore an exceptional laboratory for detecting beyond-GR physics [1–3]. Thus far, LIGO-Virgo detections of gravitational wave (GW) signals from BBH systems have been shown to be consistent with GR [4–8].

Tests of GR with GWs can be broadly divided into two categories: unmodeled and modeled. The first kind, such as residuals tests, looks for generic consistency between GR waveform templates and the observed data, without a specific model for GR deviations [6–13]. The second seeks inspiration from beyond-GR dynamics to look for GW properties beyond GR, or introduce phenomenological deviations to the observed signals, such as the parametrized post-Einsteinian framework or ringdown tests [14–20]. For either kind of test, it is beneficial to have concrete examples of plausible waveforms beyond GR; unfortunately, such examples have typically been restricted to regimes that are tractable analytically, namely the binary inspiral or ringdown. Recently, however, there has been progress toward numerical relativity simulations of BBHs in theories beyond GR [21–34]. Even though some of these simulations solve the underlying field equations only approximately, they can still serve as a qualitative example of how beyond-GR dynamics might modify the strongly nonlinear merger BH regime.

Our aim in this study is to explore how existing GW inference tools can recover signals whose merger does not follow GR. We use the beyond-GR waveforms computed through numerical relativity to simulate LIGO [35] and Virgo [36] data and study the behavior of two

*mokounkova@pasadena.edu
†misi@flatironinstitute.org
‡kchatziioannou@caltech.edu
§wfarr@flatironinstitute.org

complementary inference analyses. The first, based on BayesWave [37–39], follows a morphology-independent approach and the second, based on BILBY [40,41], assumes that the signal is consistent with existing waveform templates derived within GR. For this study, we do not consider parametrized tests of GR deviations since those are typically anchored to physical models in the inspiral and ringdown regimes, not the merger.

The BayesWave morphology-independent analysis does not assume that the signal follows the time-frequency evolution prescribed by GR, but it requires that the signal be coherent across the network and that it travel at the speed of light. Though it is possible to relax them in targeted analyses, we make two additional assumptions: (i) the signal contains solely tensor polarization modes [42], and (ii) it is elliptically polarized [39]. With these restrictions, the analysis is expected to recover and faithfully reconstruct beyond-GR signals regardless of their precise time-frequency evolution.

The BILBY template-based analysis, meanwhile, uses quasicircular BBH waveforms computed within vacuum GR with no extra physics, to model signals and estimate their physical source parameters within the GR BBH parameter space.

This analysis can illuminate degeneracies between GR and beyond-GR theories, as well as potential biases in recovered GR BBH parameters, as illustrated in Fig. 1. If beyond-GR modifications are degenerate with GR, then the template-based analysis will be able to fully model the signal, albeit with a potential *stealth* bias on the BBH parameters [43,44]. If, however, the GR and beyond-GR modifications are nondegenerate, then the template-based analysis will leave a residual signal behind.

In this study, we focus on BBH mergers in dynamical Chern-Simons (dCS) gravity. dCS is a beyond-GR theory that adds a pseudoscalar field coupled to spacetime curvature to the Einstein-Hilbert action, and has origins in string theory and loop quantum gravity [45–48]. This coupling is governed by a parameter ℓ, which has dimensions of length and corresponds to the scale below which beyond-GR effects become important; GR is recovered for $\ell = 0$. We use the waveforms from [22] to simulate data for a system consistent with GW150914 [49], for a variety of dCS coupling constants ℓ (including $\ell = 0$) and signal-to-noise ratios (SNRs).

We find that the morphology-independent BayesWave analysis can reconstruct the injected beyond-GR signals for all values of ℓ, with the fidelity of the reconstruction improving with the SNR of the simulated signal. Meanwhile, the GR-templated BILBY analysis *cannot* fully reconstruct the injected beyond-GR signals and recovers biased source parameters. This shows that the effects introduced in the merger phase of this beyond-GR waveform [22] are nondegenerate with GR.
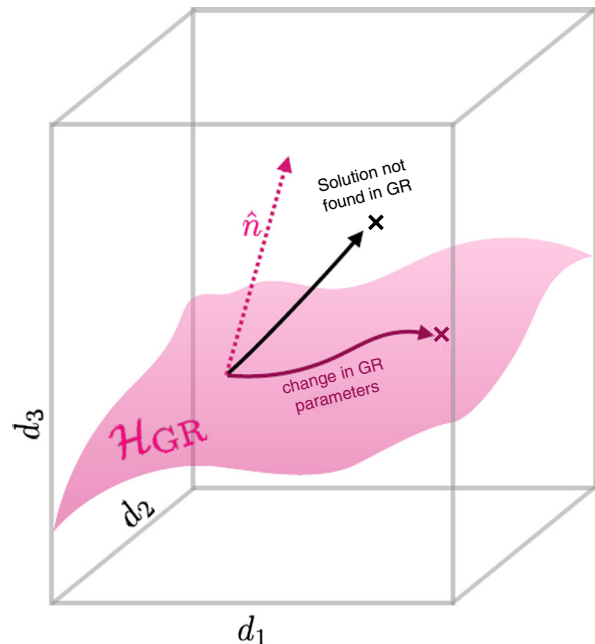

FIG. 1.   Illustration of degeneracy between GR and beyond-GR theories. BBH GW data lives in a space $\mathcal{D}$ (represented by the gray cube), where each dimension (here only d_1, d_2, and d_3 are shown) corresponds to the value of the waveform at discrete frequencies. The set of GR vacuum BBH gravitational waveforms forms a hypersurface $\mathcal{H}_{\mathrm{GR}}$ (represented by the pink surface) in $\mathcal{D}$. In other words, the gravitational waveform for any BBH merger in GR will be found on $\mathcal{H}_{\mathrm{GR}}$. Changing the parameters of the GR BBH system, such as the masses and the spins, corresponds to movement along $\mathcal{H}_{\mathrm{GR}}$ (maroon curve) from one GR solution to another. Let us consider a point in $\mathcal{H}_{\mathrm{GR}}$, and let $\hat{n}$ (dashed pink line) be the normal to $\mathcal{H}_{\mathrm{GR}}$ at that point. Then introduce a beyond-GR modification to the gravitational waveform: if beyond-GR and GR are morphologically nondegenerate, this modification will include a component along $\hat{n}$ (and can include a component along $\mathcal{H}_{\mathrm{GR}}$). If, however, there is degeneracy between a beyond-GR theory and GR, then beyond-GR modifications will only include movement along $\mathcal{H}_{\mathrm{GR}}$, so the non-GR effect can be accommodated by a change in GR parameters, amounting to stealth bias.

At sufficiently high SNR ($\gtrsim$20–30, depending on the value of ℓ), existing inference tools can identify discrepancies between the observed signal and the GR expectation. Fully interpreting such a discrepancy, however, would require additional consideration and careful studies of waveform systematics even within GR [50].

This paper is organized as follows: In Sec. II, we detail the beyond-GR numerical relativity merger waveforms we use. In Sec. III, we describe the methods for the morphology-independent and the template-based analysis, while in Sec. IV we describe the results of the two methods and how they compare. We conclude in Sec. V.

II. BEYOND-GR GRAVITATIONAL WAVEFORMS AND SIMULATED DATA

We consider dCS [45], a beyond-GR theory that treats the Einstein-Hilbert action of GR as the first term in a higher-order expansion in spacetime curvature. The higher-order curvature terms, though classical, are inspired by quantum gravity corrections [46–48]. This expansion is governed by a parameter ℓ, which corresponds to the length-scale below which beyond-GR gravity effects become important. The dCS action takes the form (setting the speed of light $c = 1$ throughout)

$$S \equiv \int d^4x \sqrt{-g} \left(\frac{R}{16\pi G} - \frac{\ell^2}{16\sqrt{2\pi G}} \vartheta\, {}^*RR - \frac{1}{2}(\partial\vartheta)^2 \right), \quad (1)$$

where the first term is the Einstein-Hilbert action of GR and the second term introduces a quadratic curvature effect through the Pontryagin density, ${}^*RR \equiv {}^*R^{abcd}R_{abcd}$, with coupling ℓ^2.[1] In order to make the theory dynamical, we couple the quadratic curvature term to an axionic scalar field ϑ, where the final term in the action corresponds to the kinetic term for the scalar field.

In order to ensure well-posedness of the dCS evolution equations [51], the corresponding numerical relativity simulations are performed in an *order-reduction scheme* [52], in which the spacetime metric and the dCS scalar field are perturbed around GR. Because the coupling in Eq. (1) is ℓ^2, each order n in the expansion will take the power ℓ^{2n}. Zeroth order (ℓ^0) corresponds to a GR background spacetime. At first order (ℓ^2), the curvature of the GR background sources the leading-order scalar field. At second order (ℓ^4), the curvature of the GR background and the leading-order scalar field source the leading-order correction to the spacetime metric. This scheme gives us access to the background GR strain waveform h_{GR} and the leading order dCS correction to the strain, $\ell^4 \delta h_{\text{dCS}}$. The total leading-order corrected dCS waveform is the sum of the two using the dCS coupling constant as

$$h = h_{\text{GR}} + (\ell/GM)^4 \delta h_{\text{dCS}}, \quad (2)$$

where M the total mass of the binary. The larger the value of ℓ, the larger the beyond-GR effects. The order-reduction scheme allows us to evaluate h_{GR} and h_{dCS} once given some system parameters, and then generate a beyond-GR waveform for any value of the coupling constant. However, due to the perturbative nature of the scheme, there is an instantaneous regime of validity (see [21,22,52] for technical details), that limits the allowed values of ℓ/GM. The

resulting waveform is not an exact solution to the dCS field equations, but it can still be used as an example of beyond-GR dynamics in a data analysis setting.

A. Simulated data parameters

For this study, we use the waveform from [22] with parameters consistent with GW150914 [49,53–55] in the $\ell = 0$ (GR) case. We have chosen such a system because GW150914 is well-studied, including with many tests of GR [5,9,18]. We choose a total mass of $M \simeq 68 M_\odot$, consistent with GW150914 [53]. This choice further ensures that most of the signal observed by LIGO-Virgo is near the merger phase where the dCS modification is the strongest. As discussed in [22], the order-reduction scheme leads to *secular* growth during the inspiral between the 'full' and 'perturbed' dCS solutions. To avoid this secular growth, the numerical simulations of [22] give a beyond-GR waveform with the dCS effects smoothly ramped-on close to merger, thus producing a combination of a GR inspiral with a dCS merger. Future work will include dCS modifications to the inspiral phase as well [56].

The physical parameters of the simulated system are consistent with those of the numerical relativity waveform in Fig. 1 of [49] as well as follow-up studies [55,57,58]. The initial dimensionless spins vectors are $\vec{\chi}_1 = 0.330\hat{z}$ and $\vec{\chi}_2 = -0.440\hat{z}$, aligned and antialigned with the orbital angular momentum respectively, leading to no spin-precession effects. The ratio of the component masses is $q = 0.819$, and the remnant BH has final Christodoulou mass 0.9525 in units of the initial mass of the system, and dimensionless spin 0.692 purely in the $\hat{z}$ direction. In vacuum numerical relativity simulations, the total mass of the system is scaled out, and thus when performing injection studies, we can introduce an arbitrary total mass. To be consistent with GW150914, we choose $M = 68 M_\odot$ [53]. When projecting the waveform into a detector, we choose a geocenter GPS time of 1126259460 s, a right ascension of 1.95 radians, a declination of -1.27 radians, and a binary inclination of π radians. We show waveforms for various values of ℓ in Fig. 2. The largest value we choose for ℓ corresponds to the maximum-allowed value for our order-reduction-schemes to be within the instantaneous regime of validity (cf. [22]). We choose spacing of ℓ values roughly even in ℓ^4, the order at which beyond-GR affects appear in the waveform.

We use these waveforms to simulate data observed with LIGO Hanford (LHO), LIGO Livingston (LLO), and Virgo using the infrastructure of [59] and assuming a zero-noise realization. We vary the strength of the signal by changing the luminosity distance of the system, computing the SNR using the Advanced LIGO *design* sensitivity [60] curve for LHO, LLO, and Virgo. Current BBH observations have typical SNRs ~10–25 [61], with GW150914 having a network SNR of 24 in two LIGO detectors [49] with the O1 sensitivity. Next-generation GW detectors will detect BBH

---

[1] ${}^*R^{abcd}R_{abcd}$ refers to the dual of a the Riemann tensor contracted with itself, which can be expressed using the fully antisymmetric Levi-Civita tensor ϵ_{abcd} as ${}^*R^{abcd}R_{abcd} = \frac{1}{2}\epsilon^{abef}R_{ef}{}^{cd}R_{abcd}$.
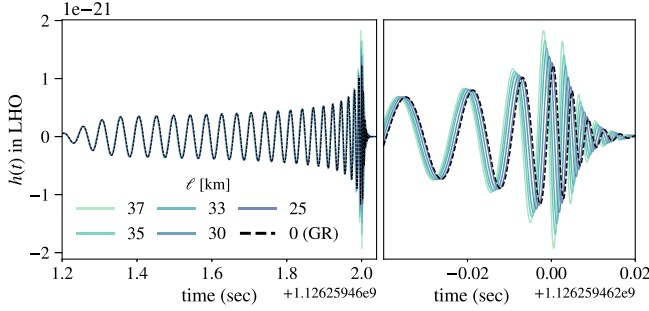
FIG. 2. Gravitational strain in LHO for various values of the dCS coupling parameter ℓ and an injected distance of 500 Mpc. The waveforms are aligned at the start and are smoothly ramped on from zero. The black curve with $\ell = 0$ corresponds to the GR waveform, while the remaining curves show beyond-GR modified waveforms for various values ℓ. With increasing ℓ, the phase of the beyond-GR waveform evolves more rapidly relative to that of GR, leading to the beyond-GR waveforms peaking earlier [22]. The beyond-GR waveforms also have a larger amplitude at merger [22] and thus have increased network SNRs at the same luminosity distance.

mergers with SNRs reaching and exceeding ~ 100 [62,63]. We thus consider network SNRs in the range 25–125 in order to span a wide range of detector capabilities.

III. METHODS

Given the simulated data described in Sec. II, we consider both a flexible morphology-independent and a GR-template based analysis and describe them in this section. Both analyses target the same data and a frequency band of $(f_{\text{low}}, f_{\text{high}}) = (25, 1024)$ Hz. The low-frequency limit is determined by the finite length of the numerical relativity simulation, while the high-frequency limit is chosen such that the analysis includes the binary merger and ringdown.

We quantify how well the two analyses reconstruct the simulated signal through the overlap between different waveforms, see e.g., [10,64]. For two waveforms A and B the overlap is defined as

$$\mathcal{O}_{A,B} \equiv \frac{\langle h_A, h_B \rangle_N}{\sqrt{\langle h_A, h_A \rangle_N \langle h_B, h_B \rangle_N}}, \tag{3}$$

where

$$\langle h_A, h_B \rangle_N = \sum_i^N \langle h_A^i, h_B^i \rangle, \tag{4}$$

$$\langle h_A^i, h_B^i \rangle = 4\text{Re} \int_{f_{\text{low}}}^{f_{\text{high}}} \frac{\tilde{h}_A^i(f) \tilde{h}_B^{i*}(f)}{S_n^i(f)} df. \tag{5}$$

Here h_A^i and h_B^i are the two target waveforms as seen in detector i, $\tilde{h}$ denotes the frequency-domain waveform, $S_n^i(f)$ is the noise power spectral density (PSD) of

detector i, and N is the total number of detectors. The optimal network SNR is $\sqrt{\langle h, h \rangle}$. For waveforms that agree perfectly, the overlap is 1. We also define the *mismatch* between two waveforms as

$$\Delta_{A,B} \equiv 1 - \mathcal{O}_{A,B}. \tag{6}$$

A. Morphology-independent analysis

We use BayesWave [37–39] to perform a morphology-independent analysis that does not impose that the signal be consistent with GR dynamics. BayesWave models the data through a sum of sine-Gaussian wavelets [65], requiring only that the signal be coherent across the detector network, travel at the speed of light between detectors, contain only tensor polarizations in an elliptical configuration, and do not disperse while traveling between the detectors. These assumptions are satisfied by our simulated data. Indeed, in dCS, GWs travel at the speed of light between detectors and contain only tensor polarizations [45]. Moreover, the numerical relativity simulation we use in thus study (cf. Sec. II A) does not exhibit spin-precession that could break the elliptical polarization assumption. Besides these, BayesWave makes no further assumption about the time-frequency content of the signal being consistent with GR.

Details about the wavelet model and the sampler implementation, settings, and priors are provided in [37–39,66]. Here we specifically use the "signal model" with a fixed PSD, and sample the multi-dimensional posterior for the number of the sine-Gaussian wavelets, their parameters, and the signal extrinsic parameters using a collection of (reversible jump) Markov Chain Monte Carlo [67] samplers. Under the default configuration and priors, BayesWave is more sensitive to BBH signals and times close to merger, where the GW amplitude peaks [10,68,69].

B. Template-based analysis

For the template-based analysis with GR waveforms, we use the parameter estimation software library BILBY [40], which has been used in parameter estimation studies of the LIGO-Virgo transient catalogs [41,61,70]. BILBY uses nested sampling [71,72] to sample the posterior distribution for the BBH parameters. The BBH parameter space within GR includes the component masses m_1 and m_2, the BH spin vectors $\vec{\chi}_1$, $\vec{\chi}_2$, the sky location and luminosity distance of the source, the inclination angle of the total angular momentum of the binary, the polarization angle, as well as a time and phase. We use standard priors and settings [41]. For the template we use the waveform model IMRPhenomXPHM [73] as implemented in LALSuite [74] due to its computational efficiency. This model includes effects of spin-precession and higher-order multiple moments and was used in the LIGO-Virgo GWTC-3 analysis [61]. We have verified that the mismatch between our numerical relativity GR waveform and the

IMRPhenomXPHM waveform with the corresponding set of GR parameters is lower than the mismatch between our numerical dCS waveform and GR waveform (at the 10^{-4} level), showing that we can resolve dCS effects.

Rather than the component spins, we express results through common spin *combinations*. The effective inspiral spin χ_{eff} [75,76] is the mass-weighted ratio sum of the spin components along the Newtonian orbital angular momentum $\hat{L}$

$$\chi_{\text{eff}} = \frac{\vec{\chi_1} \cdot \hat{L} + q\vec{\chi_2} \cdot \hat{L}}{1 + q} \in [-1, 1], \qquad (7)$$

and it is approximately conserved under spin-precession [77]. Complementary to χ_{eff} is the effective precession spin χ_p, which measures the mass-weighted in-plane spin component and characterizes spin precession [78,79]

$$\chi_p = \max \left\{ \chi_{1,\perp}, \frac{q(4q + 3)}{4 + 3q} \chi_{2,\perp} \right\} \in [0, 1], \qquad (8)$$

where $\chi_{i,\perp}$ is the component of spin perpendicular to the direction of the Newtonian orbital angular momentum. A vanishing χ_p corresponds to a system with no spin-precession, and hence spins aligned with $\hat{L}$.
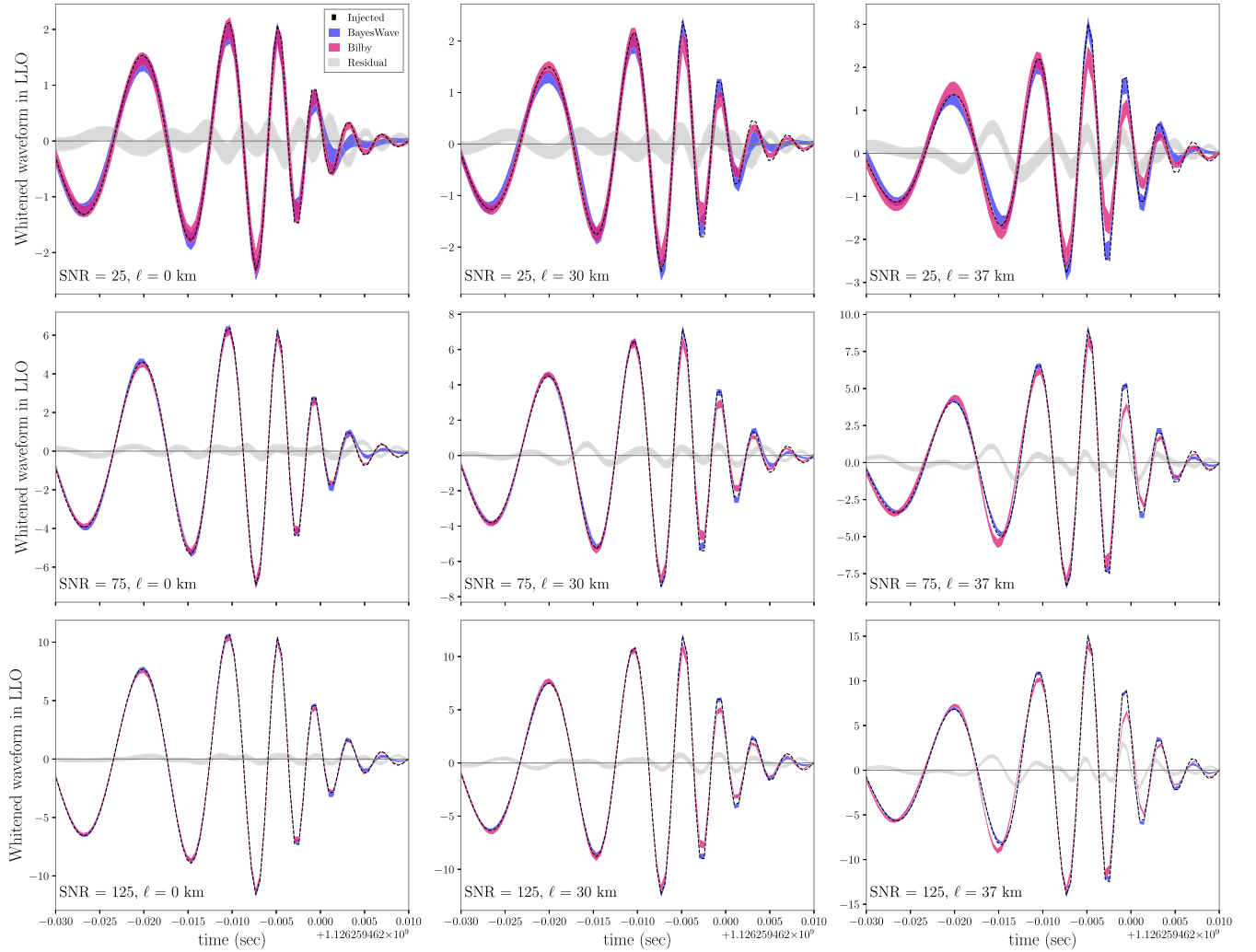

FIG. 3. Time-domain whitened waveforms in LLO for different values of ℓ (left to right) and injected network SNR (top to bottom). The dashed black curve corresponds to the simulated signal, while blue and red shaded regions denote the 90% credible intervals for the BayesWave and BILBY reconstructions respectively. The BayesWave reconstruction fully overlaps with the injected waveform for all ℓ, with the agreement improving with SNR. The BILBY reconstruction, on the other hand, is inconsistent with the injected signal and the BayesWave reconstruction for sufficiently high ℓ and SNR. We further demonstrate this by including a gray band showing the residual between the BayesWave and BILBY results in each panel. We compute the residual interval pointwise in time by subtracting the maximum likelihood BILBY waveform from the bounds on the 90% credible BayesWave interval. In the $\ell = 0$ case, the residual is consistent with zero at each time, but as we increase the value of ℓ, the residual disagrees with zero at certain times, especially near the merger (note that the scale of the ordinate axis varies for each plot). We quantitatively assess this disagreement between the BayesWave and BILBY results in Fig. 5.

IV. ANALYSIS RESULTS

Figure 3 shows the whitened time-domain simulated signal and 90%-credible intervals for the BayesWave and BILBY reconstructions in LHO. In the case of GR ($\ell = 0$), both methods return reconstructions that agree with the simulated data, and increasingly so for higher SNR signals. While this is the expected behavior for BayesWave's morphology-independent approach, it is a nontrivial statement for the BILBY analysis. Waveform models such as IMRPhenomXPHM are subject to systematics and are thus expected to not perfectly recover GR signals at sufficiently high SNRs. However, our results confirm that IMRPhenomXPHM remains faithful to this GR numerical relativity simulation even at the highest SNRs we consider. Thus, any unfaithfulness we find when $\ell \neq 0$ can be attributed to beyond-GR effects and not within-GR waveform systematics.

When ℓ is nonzero, the BayesWave reconstruction continues to accurately reproduce the simulated data, thus living up to the expectations for a morphology-independent analysis. However, the BILBY credible interval *fails* to fully contain the simulated beyond-GR waveform, with the discrepancy increasing as ℓ increases. This suggests that the beyond-GR effect we consider here is not degenerate with GR, corresponding to the black-cross case from Fig. 1. For a real observed signal, we will thus be able to compare the BayesWave and BILBY reconstructions and identify unexpected dynamics in the merger phase.

We further quantify this through the "residual," computed as the difference between the BayesWave and BILBY reconstructions. Specifically, we subtract the maximum likelihood BILBY waveform from the upper and lower 90% credible intervals of the BayesWave reconstruction as a function of time to obtain a region for the residual. This illustrates the extent to which the BILBY best-fit waveform is contained in the BayesWave interval at that specific time. In the GR case this residual is consistent with zero where the signal is strong. As ℓ increases the residual becomes inconsistent with zero during the merger phase, with smaller uncertainty as the SNR increases. This would also manifest as residual power left after subtracting the best fit BILBY reconstruction from the data, as would be measured be the residuals test formulated in [7,8]. Though our analysis demonstrates that such residuals can be identified with current data analysis tools, further quantitative estimates about the SNR or the amount of deviation required depend on the exact simulated signal considered.

This behavior is further quantified in Fig. 4, where we show the mismatch between the median BayesWave reconstruction and the injected waveform as well as the maximum-likelihood BILBY reconstruction and the injected
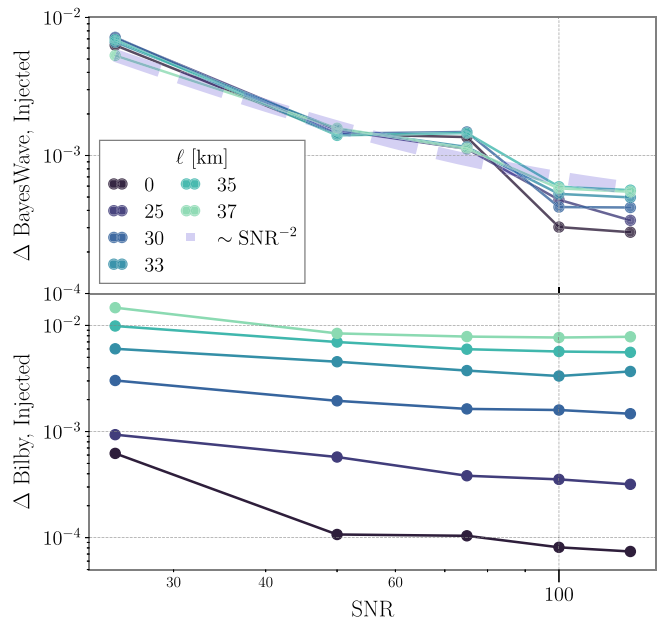

FIG. 4. Mismatch between injected and reconstructed waveforms as a function of the network SNR for different values of ℓ. The top panel corresponds to the mismatch between the injected waveforms and the median BayesWave reconstructed waveforms, while the lower panel corresponds to the mismatch between the injected waveforms and the maximum-likelihood BILBY reconstructed waveform. The dashed purple curve corresponds to the $1/\mathrm{SNR}^2$ behavior. For all ℓ, the BayesWave mismatch decreases with high SNR as expected, however the BILBY mismatch deviates from this trend and plateaus with increasing ℓ.

waveform.[2] The former does not depend on the value of ℓ and decreases as $1/\mathrm{SNR}^2$ for high SNR as expected [10,80–82], showing that BayesWave is able to faithfully recover both GR and beyond-GR injections. However, the latter is a strong function of ℓ even at SNR 25, with larger values of ℓ and thus large deviations from the expected GR signal leading to a larger mismatch. Additionally, the BILBY mismatch decreases less steeply with the SNR as ℓ increases and instead seems to plateau at large SNR. This again shows that the beyond-GR waveform cannot be faithfully reproduced with GR waveforms. Another way

---

[2]The maximum-likelihood BILBY reconstructed waveform corresponds to a true GR waveform in the BBH parameter space, while the median BILBY waveform corresponds to a pointwise (in time) median taken over a set of recovered waveforms, with no guarantee that the resulting waveform corresponds to a physical GR BBH system. In the morphology-independent BayesWave analysis, meanwhile, none of the recovered waveforms necessarily correspond to a GR BBH system, and the maximum-likelihood waveform is typically an outlier in the posterior for the number of wavelets. This is because BayesWave is a transdimensional analysis and the maximum-likelihood waveform is typically the one with the largest number of wavelets, and thus the one that overfits the noise the most. Hence, we use the maximum-likelihood BILBY waveform and the median BayesWave waveform when computing mismatches.
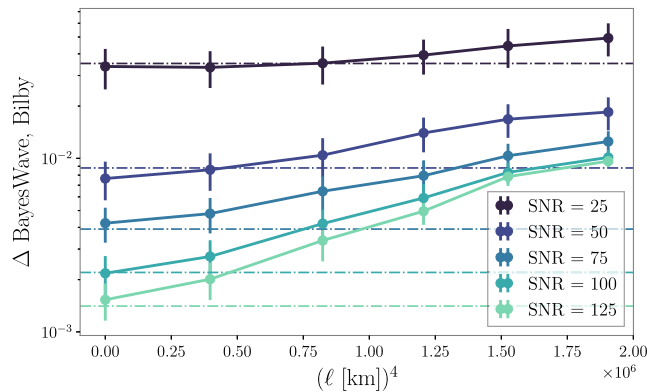
FIG. 5.   Mismatch between BILBY and BayesWave reconstructed waveforms as a function of ℓ for a variety of injection networks SNRs. Recalling that the dCS modification comes in as a power of ℓ^4, Eq. (2), we choose to show this quantity on the horizontal axis. Each curve corresponds to a different injected network SNR. We compute mismatches between 200 BayesWave and BILBY reconstructed waveforms sampled from their respective posterior distributions, and show the median mismatch, with the standard deviation as an errorbar. For each SNR, as the value of ℓ increases, the mismatch between the recovered waveforms increases. This effect persists even as we increase SNR. Each dashed line corresponds to SNR^{-2}.

to demonstrate this is presented in Fig. 5 which shows that the mismatch between the median BayesWave and maximum-likelihood BILBY reconstructions increases with ℓ for constant SNR. This figure makes only use of quantities that are available from analyses of real signals (namely, not the true injected waveform) and again demonstrates how a GR deviation can be flagged from real data.

Figure 6 shows the posterior distributions for mass and spin parameters obtained with BILBY for two values of ℓ. When $\ell = 0$, all posteriors are consistent with the injected parameters as expected. However, when $\ell \neq 0$ we recover much tighter and, in many cases, biased posteriors. The parameters with the largest posterior difference are the mass ratio q and the two spin parameters. The total mass posterior is the least biased one, possibly since it is well-measured from the merger frequency. We take a closer look at the mass ratio and spins for more values of ℓ in Fig. 7. Biases in the recovered parameters are indicative of "stealth bias", in which different values of GR parameters can better (though not perfectly) reproduce a beyond-GR waveform (in Fig. 1 this corresponds to movement towards the beyond-GR solution parallel to $\mathcal{H}_{GR}$).

We can interpret the parameter biases with increasing ℓ as follows. The beyond-GR waveforms have roughly the same merger frequency as their GR counterparts (cf. Fig. 2), but the merger is moved at earlier times and has a larger amplitude at merger. Thus, the total mass remains approximately the same to preserve the merger frequency but χ_{eff} moves to lower values, which reduces the length of the waveform due to the reduced orbital hang up effect arising
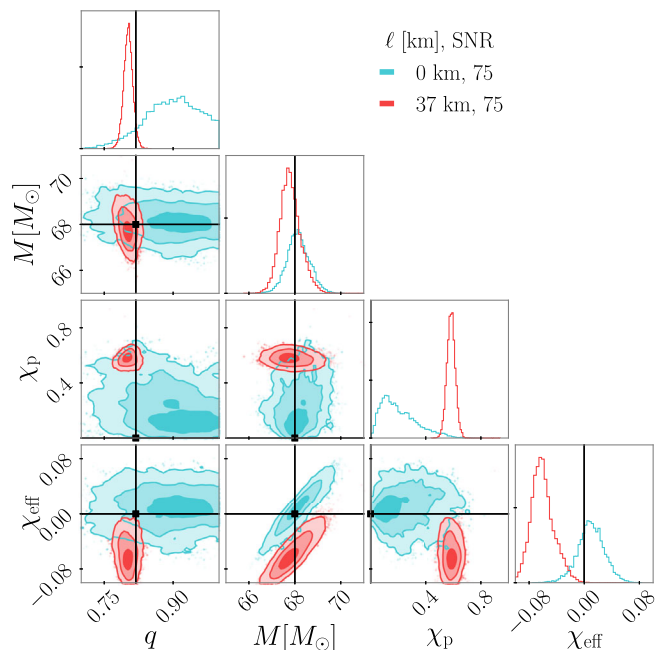

FIG. 6.   Marginalized one- and two-dimensional posteriors for various GR BBH parameters for a network SNR = 75 injection using a BILBY GR template-based analysis. We show q, the mass ratio, M, the detector frame total mass, χ_p, the precession spin parameter, and χ_{eff}, the effective spin. Blue corresponds to the posteriors for the GR ($\ell = 0$) injection, while red corresponds to the posteriors for a beyond-GR injection. The injected values are shown in black.

from spin-orbit coupling. The mass ratio q posterior remains consistent as ℓ changes, though it becomes increasingly peaked, a point to which we return later. Finally, a large χ_p results in large precession that can lead to a low inspiral amplitude compared to the merger amplitude. If the system precesses such that it becomes more face-on at merger compared to the inspiral, this increases the GW amplitude at merger.

Figures 6 and 7 also show that the beyond-GR posterior distributions are more sharply peaked and more tightly constrained than the GR posterior distributions. Indeed, the spread in q and χ_p in the $\ell = 37$ km case is 10 times smaller than the $\ell = 0$ case. While all posteriors narrow with increased SNR, the tighter constraints on the posterior distributions in the beyond-GR case can be explained by considering the intersection of isolikelihood contours with the signal manifold in the space of Fig. 1. The width of the likelihood distribution is controlled by the width of the intersection of such contours around the point in the manifold closest to the true signal (the maximum likelihood point); this is solely a function of the curvature of the signal manifold at that point, as would be evaluated by the Fisher matrix. When the maximum likelihood point corresponding to the beyond-GR signal lies in a region of parameter space with higher (such as low q) or lower (such as high q) curvature, then the width of the likelihood distribution in the
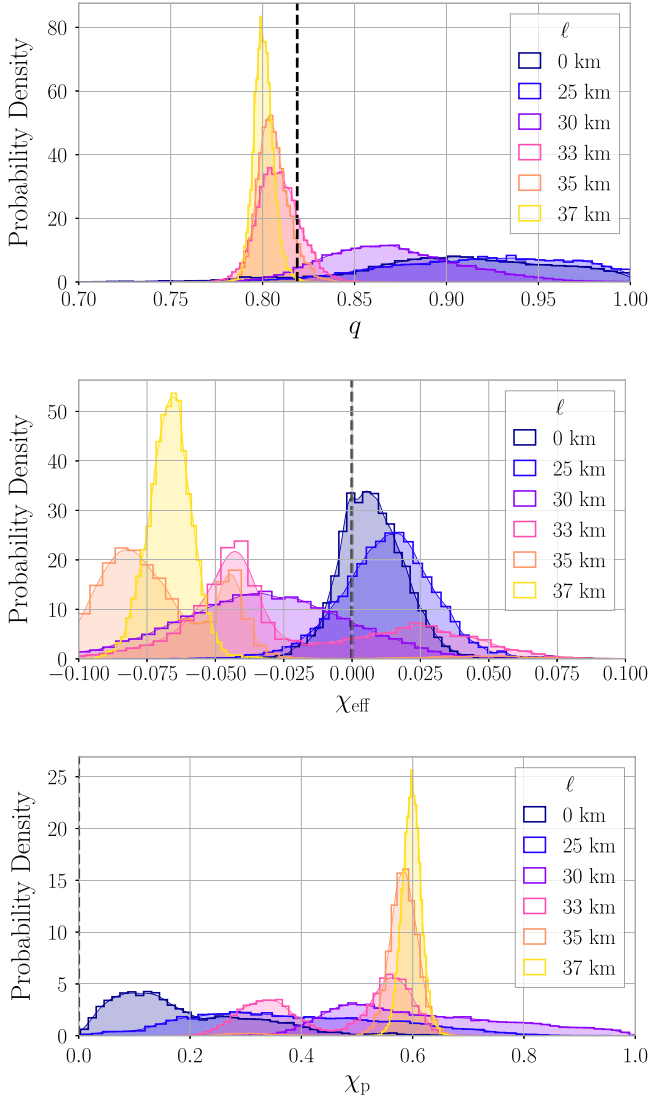
FIG. 7. One-dimensional marginalized posterior for the mass ratio q (top), the effective spin χ_{eff}, and the spin precession parameter χ_p (bottom), for injections with network SNR = 125. Each distribution corresponds to a different value of ℓ, and the dashed black line corresponds to the GR value ($\chi_p = 0$ in the GR case). As ℓ increases the posteriors are becoming increasingly biased and peaked, a behavior discussed in Sec. IV.

recovered parameters decreases or increases accordingly with respect to the GR solution.

The specific point in parameter space which will maximize the likelihood will vary with the signal morphology and the preference for specific parameters (e.g., low q in Fig. 7) is not trivial to explain, and is left for future work.

For small deviations away from GR, we would expect parameter biases to appear as linear drifts in the multi-dimensional parameter space that are proportional to ℓ^4. This is indeed the observed behavior in q and χ_p in Fig. 7. The inferred value of χ_{eff}, however, varies with ℓ in a more complicated way and even displays multi-modal structure. We have verified that this χ_{eff} behavior is convergent with

increasing SNR, it is therefore well-resolved and not a sampling artifact. This suggests that some simulated signals are away from the regime of small deviations at least as far as χ_{eff} is concerned.

V. DISCUSSION AND CONCLUSIONS

We performed a GW data analysis injection study using beyond-GR BBH *merger* waveforms. Specifically, we focused on dCS gravity, a beyond-GR theory of gravity with motivations in string theory and loop quantum gravity [45–48]. Unlike previous work, this allows us to study how realistic analysis pipelines would respond to a full beyond-GR inspiral-merger-ringdown signal obtained by approximately solving the dynamical field equations of a concrete alternative to GR.

As detailed in Sec. II, we injected the numerical relativity merger waveforms from [22] that approximately solve the dCS equations through an order-reduction scheme, with parameters consistent with GW150914 into morphology-independent and template-based analysis pipelines for a variety of values of the dCS coupling constant ℓ and injection SNR. Though we used a specific waveform for the recovery, IMRPhenomXPHM [73], we found that it is faithful to the GR simulations at every SNR considered. We thus expect our main qualitative conclusions to be generic under other waveforms.

We showed that the morphology-independent analysis with BayesWave can successfully reconstruct beyond-GR signals for all injected values of ℓ. This analysis makes minimal assumptions, assuming only that the same signal be detected in different interferometers, come from one sky location, travel at the speed of light between detectors, and contain only tensor polarizations. Both GR and dCS fit into this class of assumptions. Should LIGO-Virgo observe a signal with significant beyond-GR effects, a morphology-independent analysis such as BayesWave will be able to fully reconstruct the signal. When compared to a BILBY analysis that is based on GR waveform models, comparing the two reconstructions will reveal a discrepancy and additional, unmodeled dynamics in the observed signal. Though not considered in this study, parametrized or residual tests might also be able to identify the beyond-GR effects [6–9], though the latter is only sensitive to large deviations [83].

We return to the picture of Fig. 1 to study the results of the GR-template analysis. The BILBY analysis used (all) solutions lying in $\mathcal{H}_{\text{GR}}$ to recover the injected beyond-GR waveforms. Figure 6 then shows that the recovered GR system parameters from beyond-GR data differ from their injected values, hence showing motion along $\mathcal{H}_{\text{GR}}$. However, we also showed in Fig. 3 that solutions in $\mathcal{H}_{\text{GR}}$ did *not* fully recover the beyond-GR signal, with the effect worsening for increasing ℓ. This in turn implies that when increasing ℓ from zero (which lies on $\mathcal{H}_{\text{GR}}$) there is also motion along $\hat{n}$, the normal to $\mathcal{H}_{\text{GR}}$. In turn, the beyond-GR solution does not lie on $\mathcal{H}_{\text{GR}}$.

Attributing a discrepancy between the BAYESWAVE and BILBY reconstruction to beyond-GR effects comes, however, with additional complications. A number of effects could induce components along $\hat{n}$ besides modifications to GR, for example waveform systematics or unmodeled phenomena such as matter or orbital eccentricity. If such a discrepancy is detected in real data, extensive modeling and theory work will be required to reach a robust interpretation regarding the cause. Numerical simulations including beyond-GR dynamics could contribute to the theoretical understanding of such potential detected discrepancies. Additionally, combining constraints from multiple detections [84–87] could help disentangle between systematics and beyond-GR physics.

The code and documentation for performing and reproducing this analysis, including the numerical relativity waveforms used in this study, are available at [88].

ACKNOWLEDGMENTS

[1] T. Baker, D. Psaltis, and C. Skordis, Astrophys. J. **802,** 63 (2015).

[2] C. M. Will, Living Rev. Relativity **17,** 4 (2014).

[3] E. Berti *et al.*, Classical Quantum Gravity **32,** 243001 (2015).

[4] B. P. Abbott *et al.* (Virgo, LIGO Scientific Collaborations), Phys. Rev. Lett. **116,** 221101 (2016); **121,** 129902(E) (2018).

[5] N. Yunes, K. Yagi, and F. Pretorius, Phys. Rev. D **94,** 084002 (2016).

[6] B. P. Abbott *et al.* (LIGO Scientific, Virgo Collaborations), Phys. Rev. D **100,** 104036 (2019).

[7] R. Abbott *et al.* (LIGO Scientific, Virgo Collaborations), Phys. Rev. D **103,** 122002 (2021).

[8] R. Abbott *et al.* (LIGO Scientific, Virgo, KAGRA Collaborations), arXiv:2112.06861.

[9] B. P. Abbott *et al.* (LIGO Scientific, Virgo Collaborations), Phys. Rev. Lett. **116,** 221101 (2016); **121,** 129902(E) (2018).

[10] S. Ghonge, K. Chatziioannou, J. A. Clark, T. Littenberg, M. Millhouse, L. Cadonati, and N. Cornish, Phys. Rev. D **102,** 064056 (2020).

[11] A. Ghosh *et al.*, Phys. Rev. D **94,** 021101 (2016).

[12] A. Ghosh, N. K. Johnson-Mcdaniel, A. Ghosh, C. K. Mishra, P. Ajith, W. Del Pozzo, C. P. L. Berry, A. B. Nielsen, and L. London, Classical Quantum Gravity **35,** 014002 (2018).

[13] A. Maselli, P. Pani, L. Gualtieri, and E. Berti, Phys. Rev. D **101,** 024043 (2020).

[14] N. Yunes and F. Pretorius, Phys. Rev. D **80,** 122003 (2009).

[15] K. Chatziioannou, N. Yunes, and N. Cornish, Phys. Rev. D **86,** 022004 (2012); **95,** 129901(E) (2017).

[16] M. Agathos, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, J. Veitch, and S. Vitale, Phys. Rev. D **89,** 082001 (2014).

[17] J. Meidam, M. Agathos, C. Van Den Broeck, J. Veitch, and B. S. Sathyaprakash, Phys. Rev. D **90,** 064009 (2014).

[18] M. Isi, M. Giesler, W. M. Farr, M. A. Scheel, and S. A. Teukolsky, Phys. Rev. Lett. **123,** 111102 (2019).

[19] G. Carullo, W. Del Pozzo, and J. Veitch, Phys. Rev. D **99,** 123029 (2019); **100,** 089903(E) (2019).

[20] R. Brito, A. Buonanno, and V. Raymond, Phys. Rev. D **98,** 084038 (2018).

[21] M. Okounkova, L. C. Stein, M. A. Scheel, and S. A. Teukolsky, Phys. Rev. D **100,** 104026 (2019).

[22] M. Okounkova, L. C. Stein, J. Moxon, M. A. Scheel, and S. A. Teukolsky, Phys. Rev. D **101,** 104016 (2020).

[23] M. Okounkova, Phys. Rev. D **102,** 084046 (2020).

[24] W. E. East and J. L. Ripley, Phys. Rev. D **103,** 044040 (2021).

[25] H. Witek, L. Gualtieri, P. Pani, and T. P. Sotiriou, Phys. Rev. D **99,** 064035 (2019).

[26] J. Cayuso, N. Ortiz, and L. Lehner, Phys. Rev. D **96,** 084043 (2017).

[27] R. Cayuso and L. Lehner, Phys. Rev. D **102,** 084008 (2020).

[28] G. Allwright and L. Lehner, Classical Quantum Gravity **36,** 084001 (2019).

[29] J. L. Ripley, Int. J. Mod. Phys. D **31,** 2230017 (2022).

[30] H. O. Silva, H. Witek, M. Elley, and N. Yunes, Phys. Rev. Lett. **127,** 031101 (2021).

[31] H. Witek, L. Gualtieri, and P. Pani, Phys. Rev. D **101,** 124055 (2020).

[32] E. W. Hirschmann, L. Lehner, S. L. Liebling, and C. Palenzuela, Phys. Rev. D **97,** 064032 (2018).

[33] E. Berti, V. Cardoso, L. Gualtieri, M. Horbatsch, and U. Sperhake, Phys. Rev. D **87,** 124020 (2013).

[34] J. Healy, T. Bode, R. Haas, E. Pazos, P. Laguna, D. M. Shoemaker, and N. Yunes, Classical Quantum Gravity **29,** 232002 (2012).

[35] J. Aasi *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **32,** 074001 (2015).

[36] F. Acernese *et al.* (Virgo Collaboration), Classical Quantum Gravity **32,** 024001 (2015).

[37] N. J. Cornish and T. B. Littenberg, Classical Quantum Gravity **32,** 135012 (2015).

[38] T. B. Littenberg and N. J. Cornish, Phys. Rev. D **91,** 084034 (2015).

[39] N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, Phys. Rev. D **103,** 044006 (2021).

[40] G. Ashton *et al.*, Astrophys. J. Suppl. Ser. **241,** 27 (2019).

[41] I. M. Romero-Shaw *et al.*, Mon. Not. R. Astron. Soc. **499,** 3295 (2020).

[42] K. Chatziioannou, M. Isi, C.-J. Haster, and T. B. Littenberg, Phys. Rev. D **104,** 044005 (2021).

[43] S. Vitale and W. Del Pozzo, Phys. Rev. D **89,** 022002 (2014).

[44] M. Vallisneri and N. Yunes, Phys. Rev. D **87,** 102002 (2013).

[45] S. Alexander and N. Yunes, Phys. Rep. **480,** 1 (2009).

[46] M. B. Green and J. H. Schwarz, Phys. Lett. **149B,** 117 (1984).

[47] V. Taveras and N. Yunes, Phys. Rev. D **78,** 064070 (2008).

[48] S. Mercuri and V. Taveras, Phys. Rev. D **80,** 104007 (2009).

[49] B. P. Abbott *et al.* (Virgo, LIGO Scientific Collaborations), Phys. Rev. Lett. **116,** 061102 (2016).

[50] M. Pürrer and C.-J. Haster, Phys. Rev. Res. **2,** 023151 (2020).

[51] T. Delsate, D. Hilditch, and H. Witek, Phys. Rev. D **91,** 024027 (2015).

[52] M. Okounkova, L. C. Stein, M. A. Scheel, and D. A. Hemberger, Phys. Rev. D **96,** 044020 (2017).

[53] B. P. Abbott *et al.* (LIGO Scientific, Virgo Collaborations), Phys. Rev. Lett. **116,** 241102 (2016).

[54] P. Kumar, J. Blackman, S. E. Field, M. Scheel, C. R. Galley, M. Boyle, L. E. Kidder, H. P. Pfeiffer, B. Szilagyi, and S. A. Teukolsky, Phys. Rev. D **99,** 124005 (2019).

[55] G. Lovelace *et al.*, Classical Quantum Gravity **33,** 244002 (2016).

[56] J. T. Gálvez Ghersi and L. C. Stein, Phys. Rev. E **104,** 034219 (2021).

[57] S. Bhagwat, M. Okounkova, S. W. Ballmer, D. A. Brown, M. Giesler, M. A. Scheel, and S. A. Teukolsky, Phys. Rev. D **97,** 104065 (2018).

[58] M. Giesler, M. Isi, M. A. Scheel, and S. Teukolsky, Phys. Rev. X **9,** 041060 (2019).

[59] P. Schmidt, I. W. Harry, and H. P. Pfeiffer, arXiv:1703 .01076.

[60] L. Barsotti, P. Fritschel, M. Evans, and S. Gras, Updated Advanced LIGO sensitivity design curve, Technical Report T1800044-v5, LIGO, 2018, https://dcc.ligo.org/LIGO-T1800044/public.

[61] R. Abbott *et al.* (LIGO Scientific, Virgo, KAGRA Collaborations), arXiv:2111.03606.

[62] B. Sathyaprakash *et al.*, Classical Quantum Gravity **29,** 124013 (2012); **30,** 079501(E) (2013).

[63] B. P. Abbott *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **34,** 044001 (2017).

[64] T. Apostolatos, Phys. Rev. D **52,** 605 (1995).

[65] J. Morlet, G. Arens, E. Fourgeau, and D. Glard, Geophysics **47,** 203 (1982).

[66] M. Wijngaarden, K. Chatziioannou, A. Bauswein, J. A. Clark, and N. J. Cornish, Phys. Rev. D **105,** 104019 (2022).

[67] P. J. Green, Biometrika **82,** 711 (1995).

[68] T. B. Littenberg, J. B. Kanner, N. J. Cornish, and M. Millhouse, Phys. Rev. D **94,** 044050 (2016).

[69] J. B. Kanner, T. B. Littenberg, N. Cornish, M. Millhouse, E. Xhakaj, F. Salemi, M. Drago, G. Vedovato, and S. Klimenko, Phys. Rev. D **93,** 022002 (2016).

[70] R. Abbott *et al.* (LIGO Scientific, Virgo Collaborations), Phys. Rev. X **11,** 021053 (2021).

[71] J. Skilling, Bayesian Anal. **1,** 833 (2006).

[72] J. Veitch and A. Vecchio, Phys. Rev. D **81,** 062003 (2010).

[73] G. Pratten *et al.*, Phys. Rev. D **103,** 104056 (2021).

[74] LIGO Scientific, Virgo Collaborations, LALSuite, https://git .ligo.org/lscsoft/lalsuite (2018).

[75] P. Ajith *et al.*, Phys. Rev. Lett. **106,** 241101 (2011).

[76] L. Santamaria *et al.*, Phys. Rev. D **82,** 064016 (2010).

[77] E. Racine, Phys. Rev. D **78,** 044021 (2008).

[78] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Phys. Rev. Lett. **113,** 151101 (2014).

[79] P. Schmidt, F. Ohme, and M. Hannam, Phys. Rev. D **91,** 024043 (2015).

[80] L. Lindblom, B. J. Owen, and D. A. Brown, Phys. Rev. D **78,** 124020 (2008).

[81] E. Baird, S. Fairhurst, M. Hannam, and P. Murphy, Phys. Rev. D **87,** 024035 (2013).

[82] K. Chatziioannou, A. Klein, N. Yunes, and N. Cornish, Phys. Rev. D **95,** 104004 (2017).

[83] N. K. Johnson-McDaniel, A. Ghosh, S. Ghonge, M. Saleem, N. V. Krishnendu, and J. A. Clark, Phys. Rev. D **105,** 044020 (2022).

[84] M. Isi, K. Chatziioannou, and W. M. Farr, Phys. Rev. Lett. **123,** 121101 (2019).

[85] M. Isi, W. M. Farr, and K. Chatziioannou, Phys. Rev. D **106,** 024048 (2022).

[86] M. Saleem, N. V. Krishnendu, A. Ghosh, A. Gupta, W. Del Pozzo, A. Ghosh, and K. G. Arun, Phys. Rev. D **105,** 104066 (2022).

[87] C. J. Moore, E. Finch, R. Buscicchio, and D. Gerosa, iScience **24,** 102577 (2021).

[88] M. Okounkova, Beyond gr analysis, github.com/mariaokounkova/BeyondGRAnalysis.

[89] C. M. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, Publ. Astron. Soc. Pac. **131,** 024503 (2019).

[90] A. Nitz *et al.*, GWASTRO/PYCBC: v2.0.2 release of PYCBC (2022).

[91] C. Hoy and V. Raymond, SoftwareX **15,** 100765 (2021).

[92] S. Ghonge, IGWN-WAVE-COMPARE library, https://git.ligo .org/sudarshan-ghonge/igwn-wave-compare.

[93] D. Foreman-Mackey, J. Open Source Software **1,** 24 (2016).