



# CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

## Potential observations of false deviations from general relativity in gravitational wave signals from binary black holes

Peter T. H. Pang, Juan Calderón Bustillo, Yifan Wang, and Tjonnie G. F. Li

Phys. Rev. D **98**, 024019 — Published 9 July 2018

DOI: [10.1103/PhysRevD.98.024019](https://doi.org/10.1103/PhysRevD.98.024019)

# On the Potential Observation of False Deviations from General Relativity in Gravitational Wave Signals from Binary Black Holes

Peter T. H. Pang,<sup>1,\*</sup> Juan Calderón Bustillo,<sup>2,†</sup> Yifan Wang,<sup>1</sup> and Tjonnie G. F. Li<sup>1</sup>

<sup>1</sup>*Department of Physics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong*

<sup>2</sup>*Center for Relativistic Astrophysics and School of Physics,  
Georgia Institute of Technology, Atlanta, GA 30332*

Detections of gravitational waves emitted by binary black holes allow for tests of General Relativity in the strong-field regime. In particular, deviations from General Relativity can be observed by comparing incoming signals to waveform templates that include parametrized deviations from General Relativity. However, it is essential that the General Relativity sector of these templates accounts for all predictable physics. Otherwise, missing physics might be mimicked by the “beyond General Relativity” sector of the templates, leading the analysis to report apparent deviations from General Relativity. Current parametrized tests implement templates that omit physical phenomena such as orbital eccentricity and higher-order modes. In this paper, we show how the omission of higher modes can lead to false deviations from General Relativity when these effects are strong enough. We study the extent of these deviations as a function of the mass ratio and the orbital orientation. We find that significant false deviations can arise when current tests are performed on signals emitted by asymmetric binaries whose orbital angular momentum is orthogonal to the line-of-sight. We estimate that the Advanced LIGO-Virgo network operating at its design sensitivity can observe false violations with a significance above  $5\sigma$  as often as once per year. Similar results are expected for other tests of General Relativity that compare the data to waveforms where some physical effects are omitted.

## I. INTRODUCTION

Since its postulation in 1916 [1], General Relativity (GR) has successfully passed all consistency tests performed. Most of them have been testing gravity in its so called weak-field regime based on astrophysical observations [2–4]. It was not until the recent detection of gravitational waves (GWs) from binary black holes (BBH) [5–7] and a binary neutron star (BNS) [8] that GR has started to be tested in its strong field regime. Since then, a variety of tests have been performed. As a result, bounds on the mass of the graviton improved by a factor of two previous existing ones [9]. Tests of the dynamics predicted by GR have been performed by comparing the final black hole states predicted from the early and late inspiral emission. Moreover, the recent joint detection of GW170814 by the Advanced LIGO [10] and Virgo [11] detectors, has allowed for the first direct measurement of the GW polarization, enabling to test the hypothesis that GWs are purely tensorial [12, 13]. Finally, the detected signals have been compared to generalized, analytical inspiral-merger-ringdown waveform templates that include parametrized deformations with respect to GR. Although none of these tests have shown strong evidence for deviations from GR,  $2\sigma$  deviations have been observed for some of these events. These deviations can be ascribed to the stochastic nature of the noise or any physics that has not been captured by our waveform models.

Most tests of GR using GW from BBH involve the comparison of the signal to waveform templates. As de-

scribed above, in some cases these models explicitly include deviations from GR. In doing so, it is crucial that the GR piece (or “sector”) of these templates, to which non-GR parameters (or a “nonGR sector”) is added, captures all physics predicted by GR. Otherwise, it is possible that the omitted physics is mimicked by physics beyond GR, leading to apparent false deviations from GR. State-of-the-art tests of this kind use templates whose GR sector omits the impact of physical phenomena such as the orbital eccentricity of the binary and the higher modes of the GW emission. In particular, current tests of GR implement underlying waveform models, or GR sector, that only consider the dominant  $\ell = 2$  modes [14], omitting the impact of further higher-order modes (see Section II). This is partly due to the fact that no waveform model accounting for the effect of both higher modes and generic spins were available. However, higher modes are known to have a large impact on the GW signal emitted by BBH with mass ratios  $q = m_1/m_2 > 4$  and total mass  $M = m_1 + m_2 > 100M_\odot$  [15–21].

In this work, we investigate how the impact of omitting physical effects in the GW signal can be mimicked by physics beyond GR, leading to apparent deviations from GR. In particular, we investigate the impact of omitting higher modes in the waveform model as a proxy for a missing physical effect.<sup>1</sup>

We note that, to date, no source in the region of the parameter space described above has been found [22]. Moreover, it is known that matched-filter searches [23],

---

\* [thpang@phy.cuhk.edu.hk](mailto:thpang@phy.cuhk.edu.hk)

† [juan.bustillo@physics.gatech.edu](mailto:juan.bustillo@physics.gatech.edu)

<sup>1</sup> We note that although the spin of the binary has an impact on its higher mode content, this is dominated by the value of the mass ratio [16, 19].

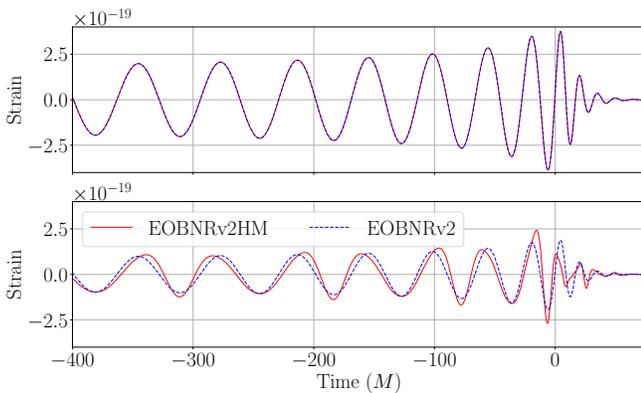


FIG. 1. **Impact of higher modes in a time domain waveform:** The last cycles of a waveform from a non-spinning,  $q = 8$  binary (in geometrical units) with only the  $(2, 2)$  mode (EOBNRv2, blue dashed) and with higher-order modes included (EOBNRv2HM, red solid), when the source is face-on (top panel) and edge-on (bottom panel). There are no visible differences when the source is face-on, but higher modes introduce strong features near the merger-ringdown regime in the edge-on case.

which are in principle optimal once the morphology of the signal is known, have a poor sensitivity in this part of the parameter space. This may be partly due to the fact that current templates used as filters omit the higher-order modes of the signals, and partly due to the strong background of non-Gaussian noise transients affecting the high mass region of the parameter space [24]. In fact, this region is currently better covered by morphology agnostic searches [25, 26]. However, the current development of enhanced matched-filter searches [27, 28] and the expected increase in sensitivity of the Advanced LIGO detectors for future observation runs might enable the first observation of the sources considered in this study.

The rest of this paper is structured as follows. Section II provides a brief overview of higher-order modes and the situations in which these produce strong effects. Sections III and IV introduce the parametrized test of GR and its implementation within the framework of Bayesian Inference. Section V gives an overview of the simulated signals and recovery templates we use. In section VI we show that omission of higher modes in the GR sector of our templates can lead to strong apparent deviations from GR. This is, in the case of highly asymmetric BBH whose orbital plane is not perpendicular to the line-of-sight. Finally, we provide an outlook for future gravitational-wave detections Sec. VII.

## II. GRAVITATIONAL WAVES FROM BINARY BLACK HOLES

In GR, the emission from a non-eccentric BBH is characterized by a set of 15 parameters. The set of 8 parameters

given by the masses  $m_{1,2}$  and spins  $\vec{s}_{1,2}$  of the individual black holes are known as the intrinsic parameters  $\Xi$  of the binary. Next, the sky location of the observer in the frame of the detector can be described in standard spherical coordinates  $(d_L, \iota, \psi)$  with origin in the center of mass of the BBH. The axis  $\iota = 0$  is chosen to be orthogonal to the orbital plane, defined by  $\iota = \pi/2$ . In this framework, a BBH is said to be face-on or edge-on if the observer is located at  $\iota = 0$  or  $\iota = \pi/2$ , respectively. Finally, the time of coalescence  $t_c$ , the sky-location of the binary in the sky of the detector  $(\theta, \varphi)$ , and the polarisation angle  $\psi_p$  of the signal complete the 15 parameters set.

The GW strain  $h$  of a binary located at a luminosity distance  $d_L$ , as observed by a GW detector with antenna pattern  $(F_+, F_\times)$  [29] can be expressed as a combination of the two GW polarizations  $(h_+, h_\times)$  as

$$h(\Xi, d_L, \iota, \psi, \theta, \varphi, \psi_p; t - t_c) = \sum_{i=+, \times} F_i(\theta, \varphi, \psi_p) h_i(d_L, \iota, \psi, \Xi; t - t_c). \quad (1)$$

Each GW polarization can be expressed as a superposition of gravitational wave modes  $h_{\ell, m}$

$$h_+ - ih_\times = \frac{1}{d_L} \sum_{\ell \geq 2} \sum_{m=-\ell}^{m=\ell} Y_{\ell, m}^{-2}(\iota, \psi) h_{\ell, m}(\Xi; t - t_c) \equiv h. \quad (2)$$

For the case of non-precessing binaries, the above sum is always dominated by the  $(\ell, m) = (2, \pm 2)$  modes during the inspiral part of the coalescence. The remaining modes, known as higher-order modes, have a subdominant effect and only contribute significantly to the GW signal during the last few cycles and merger of the binary. Notably, the amplitude of the higher-order modes grows with the mass ratio  $q$ , so that highly asymmetric binaries tend to have strong higher modes. Finally, the orientation of the binary also impacts the contribution of higher modes to the GW signal. The spherical harmonics  $Y_{2, \pm 2}$  have their maximum at  $\iota = (0, \pi)$  and minimum at  $\theta = \pi/2$ , while the rest of the harmonics behave in different ways, having their maxima close to  $\iota = \pi/2$ . This makes the  $(2, 2)$  mode especially dominant for face-on/off sources, while edge-on sources have a more important contribution from higher-order modes [15, 16, 30].

## III. PARAMETRIZED TESTS OF GENERAL RELATIVITY

In this study we will focus on a generic test of GR that focuses on measuring parameterized deviations deformation from the template waveform [31]. In particular, the method uses a closed-form expression for the waveform which is parameterized by a set of parameters

$\{p_i\}$ . For example, in the so-called IMRPhenomPv2 model these parameters are given by inspiral PN parameters  $\{\varphi_0 \dots \varphi_7\}$  and  $\{\varphi_{5l}, \varphi_{6l}\}$ , phenomenological inspiral-merger  $\{\beta_0 \dots \beta_3\}$  and merger-ringdown  $\{\alpha_0 \dots \alpha_5\}$ . Deviations from GR, or the non-GR sector, are represented by relative deviations  $p_i \rightarrow (1 + \delta\hat{p}_i)p_i$ . In this so-called parametrized test, GW data is compared to waveforms  $h$  with the goal of obtaining a probability distribution for  $\delta\hat{p}_i$ . Values of  $\delta\hat{p}_i \neq 0$  indicate that the GR sector of  $h$  can not recover the whole physics present in the incoming signal, therefore providing evidence that the signal contains physics beyond GR. In these sorts of analyses, it is essential that the GR sector of the sector  $h_{\text{GR}}(t)$  includes all the physics predicted by GR. Otherwise, it is possible that GR physics missing in GR sector of the template could be absorbed by the non-GR one. This would lead to an apparent false violation of GR.

Alternatively, the idea of the parameterized tests can also be re-expressed as

$$h(t) = h_{\text{GR}}^{\text{incomplete}}(t) + h_{\text{nonGR}}(t) \quad (3)$$

where, in general

$$h_{\text{GR}}^{\text{incomplete}}(t) = h_{\text{GR}}(t) - h_{\text{GR}}^{\text{missing}}. \quad (4)$$

Above,  $h_{\text{GR}}(t)$  denotes an ideal GR sector accounting for all physics predicted by GR while  $h_{\text{GR}}^{\text{missing}}(t)$  denotes the part of the full GR sector omitted by some incomplete GR sector  $h_{\text{GR}}^{\text{incomplete}}(t)$ . And  $h_{\text{nonGR}}(t)$  denotes nonGR sector accounting for physics beyond GR.

Current parametrized tests of GR implement a GR sector that omits the higher-order modes of the GW emission [14]. Denoting by  $\Xi$  the set of intrinsic parameters of a given BBH, we can express this as:

$$\begin{aligned} h_{\text{GR}}(t, \Xi) &= \sum_{(\ell \geq 2, |m| \leq \ell)} Y_{\ell, m}^{-2}(\iota, \psi) h_{\ell, m}(t, \Xi), \\ h_{\text{GR}}^{\text{incomplete}}(t, \Xi) &= \sum_{(\ell = 2, |m| \leq \ell)} Y_{\ell, m}^{-2}(\iota, \psi) h_{\ell, m}(t, \Xi), \quad (5) \\ h_{\text{GR}}^{\text{missing}}(t, \Xi) &= \sum_{(\ell \neq 2, |m| \leq \ell)} Y_{\ell, m}^{-2}(\iota, \psi) h_{\ell, m}(t, \Xi). \end{aligned}$$

In the following, we will investigate how the higher modes, or missing physics  $h_{\text{GR}}^{\text{missing}}(t, \Xi)$  omitted by the GR sectors  $h_{\text{GR}}^{\text{incomplete}}(t, \Xi)$  of the templates  $h$  can be mimicked by the physics beyond GR present in its nonGR sector  $h_{\text{nonGR}}(t)$ , leading to apparent violations of General Relativity. Our goal is to systematically study the strength of these violations as a function of the parameters  $\vec{\lambda} = (\Xi; \iota, \psi)$  of the binary. To this end, we will perform parametrized tests of GR on a large set of simulated signals with different physical parameters, which refer to as *injections*, including higher modes. In order to prove that any violations we find are caused the presence of higher modes in the injections, and not by other possible differences between the injections and the templates, we repeat our analysis using a second injection set with identical physical parameters in which higher modes are omitted.

#### IV. BAYESIAN INFERENCE

The inner product of two real functions  $a(t)$  and  $b(t)$  is defined as

$$(a|b) = 4\Re \int_{f_{\text{low}}}^{f_{\text{high}}} \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df. \quad (6)$$

Here,  $\tilde{a}(f)$  is the Fourier transform of  $a(t)$ ,  $*$  denotes complex conjugation and  $S_n(f)$  denotes the one sided power spectral density of the detector noise. Our analysis considers the predicted Advanced LIGO sensitivity curve, also known as the zero-detuned-high-power configuration [32]. We use lower and higher frequency cutoffs  $f_{\text{low}}$  and  $f_{\text{high}}$ , of 20Hz and 4096Hz respectively.

The posterior  $p(\vec{\lambda}|d)$  that a signal  $h(\vec{\lambda})$  with parameters  $\vec{\lambda}$  is embedded in a given signal  $d$ , is given by

$$p(\vec{\lambda}|d, \mathcal{H}) = \frac{p(d|\vec{\lambda}, \mathcal{H})p(\vec{\lambda}|\mathcal{H})}{p(d|\mathcal{H})}. \quad (7)$$

Assuming Gaussian noise, the likelihood  $p(d|\vec{\lambda})$  that a signal  $d(t)$  with unknown parameters is the result of burying some signal template  $h(\vec{\lambda})$  with parameters  $\vec{\lambda}$  in such noise, is given by

$$p(d|\vec{\lambda}) \propto \exp \left[ -\frac{1}{2} (d - h(\vec{\lambda})|d - h(\vec{\lambda})) \right]. \quad (8)$$

In general, the vector of parameters  $\vec{\lambda}$  consists on a combination of 15 physical parameters  $\{\Xi; d_L, \theta, \varphi; \iota, \psi, \psi_p; t - t_c\}$  and the parameters  $\delta\hat{p}_i$  accounting for deviations from GR. The shape and spread of the likelihood function along the parameter space follow from the signal present in the data and the power spectral density of the noise. In this study, our data  $d$  will consist in simulated signals  $h_{\text{inj}}$  injected in Gaussian noise  $h$ , generated according to some power spectral density  $S_n$ . In what follows, we will refer to our input data  $h_{\text{inj}}$  as *injections*, while we will refer to our recovery waveforms  $h(\lambda)$  as *templates*.

The parametrized test of GR can be further extended to a model selection problem. In particular, one needs to decide which of two models is a better description of our GW data: a model that only accounts for GR physics, or another one including physics beyond it. Model selection is a problem that is now well established in gravitational wave data analysis in the framework of Bayesian inference. Given a simulated signal  $h_{\text{inj}}$  we want to compare two *hypothesis*:  $\mathcal{H}_0$  and  $\mathcal{H}_1$ . These assume respectively that data is well described by GR and that parametrized deviations from GR are needed for a good description of the data. The posterior probability of  $\mathcal{H}$  being true is given by

$$p(\mathcal{H}|d) = \frac{p(d|\mathcal{H})p(\mathcal{H})}{p(d)}, \quad (9)$$

where the marginal likelihood  $p(\mathcal{H})$  is given by

$$p(d|\mathcal{H}) = \int p(\vec{\lambda}, \mathcal{H}) p(d|\vec{\lambda}, \mathcal{H}) d\vec{\lambda}. \quad (10)$$

The two hypothesis can be then compared by means of their odds that, given by

$$\mathcal{O}_0^1 = \frac{p(\mathcal{H}_1|d)}{p(\mathcal{H}_0|d)} = \frac{p(d|\mathcal{H}_1) p(\mathcal{H}_1)}{p(d|\mathcal{H}_0) p(\mathcal{H}_0)} \equiv \mathcal{B}_0^1 \frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}. \quad (11)$$

Above,  $\mathcal{O}$  and  $\mathcal{B}$  are the so-called *odd ratio* and *Bayes factor*, respectively. A result of  $\mathcal{O}_0^1 > 1$  indicates that  $\mathcal{H}_1$  provides a better description of the data than  $\mathcal{H}_0$ , vice versa.

In this work, we want to show how inclusion of higher modes in the injections can lead to apparent deviations from GR when these are omitted in the templates. Since the effect of higher-order modes is only significant in the merger-ringdown stages of the binary [15, 17, 30, 33], we focus our analysis on these parts of the emission. Therefore, the set of extra GR parameters is reduced to those affecting these stages of the coalescence. These are commonly expressed as [9]:

$$\{\delta\hat{p}_n\} = \{\delta\hat{\alpha}_2, \delta\hat{\alpha}_3, \delta\hat{\alpha}_4\}. \quad (12)$$

The parametrised test will yield posterior distributions for each of the above parameters. We stress that values  $\delta\alpha_i \neq 0$  indicate deviations from GR. In this situation, we can define the strength of a violation of GR as the distance between the median of the posterior distribution of a given nonGR parameter  $\delta\alpha_i$  and 0 measured in number of standard deviations. Finally, the odds ratio  $O_{\text{GR}}^{\text{nonGR}}$  can be written as

$$\begin{aligned} O_{\text{GR}}^{\text{nonGR}} &= \frac{1}{7} \frac{P(\mathcal{H}_{\text{nonGR}})}{P(\mathcal{H}_{\text{GR}})} (B_{\text{GR}}^{\delta\hat{\alpha}_2 \neq 0} + B_{\text{GR}}^{\delta\hat{\alpha}_3 \neq 0} + B_{\text{GR}}^{\delta\hat{\alpha}_4 \neq 0} \\ &\quad + B_{\text{GR}}^{\delta\hat{\alpha}_2, \delta\hat{\alpha}_3 \neq 0} + B_{\text{GR}}^{\delta\hat{\alpha}_2, \delta\hat{\alpha}_4 \neq 0} + B_{\text{GR}}^{\delta\hat{\alpha}_3, \delta\hat{\alpha}_4 \neq 0} \\ &\quad + B_{\text{GR}}^{\delta\hat{\alpha}_2, \delta\hat{\alpha}_3, \delta\hat{\alpha}_4 \neq 0}) \end{aligned} \quad (13)$$

$$\equiv \mathcal{B}_{\text{GR}}^{\text{nonGR}} \frac{P(\mathcal{H}_{\text{nonGR}})}{P(\mathcal{H}_{\text{GR}})}. \quad (14)$$

We will say that evidence for deviations from GR have been found when either the posterior distributions for  $\{\delta\hat{\alpha}_2, \delta\hat{\alpha}_3, \delta\hat{\alpha}_4\}$  exclude the 0 value or when the odds ratio is such that  $O_{\text{GR}}^{\text{nonGR}} \gg 1$ . In our study, we take the the prior preference between hypothesis to be 1. Therefore  $O_{\text{GR}}^{\text{nonGR}} = \mathcal{B}_{\text{GR}}^{\text{nonGR}}$ . Finally, the sampling of the likelihood function over our 9 to 12-dimensional parameter space is carried using the Nested Sampling algorithm `lalinference_nest` [34] implemented in LAL [35].

## V. WAVEFORMS

The true GW emission from a BBH contains, in general, higher modes. Hence, for our injections, we need

a model that includes higher modes. By the time this study was performed, the available inspiral-merger-ringdown waveform models including higher modes were the effective-one-body models for non-spinning BBH presented in [36, 37], and we chose the former one as it includes more harmonic modes, namely,  $(\ell, m) = (2, \pm 1), (2, \pm 2), (3, \pm 3), (4, \pm 4)$  and  $(5, \pm 5)$  modes.<sup>2</sup> The reliability of this model has been tested recently in the appendix of [26]. However, we note that waveform models with the effects of higher modes have been published [38, 39].

We want to demonstrate how the effect of higher modes in the GW waveform can be mimicked by physics beyond GR. Therefore, we need to demonstrate that any violations we find are indeed produced by the effect of higher modes. Hence, we repeat our analysis using a target waveform model that omits higher modes and check that no violations are observed. To this, we choose as injections the restriction of the previous model to the  $(2, \pm 2)$  modes<sup>3</sup>. We note that there exists a large collection of models of the  $(2, \pm 2)$  modes emitted by a BBH that we could have used, in principle [40–45]. However, we want to be sure that the only differences in the two analyses is strictly due to the presence of higher modes in one of the models, and not due to different modelings of the dominant  $(2, 2)$  mode.

Following recent studies of this kind [9] our templates implement GR sectors computed by the phenomenological inspiral-merger-ringdown model for precessing coalescing binaries described in [14, 41], to which an extra nonGR sector is added in the parametrized way described in section III<sup>4</sup>. We note that this model aims to account for generic spin effects. In particular, an approximate modeling of precession effects is implemented via a precession parameter  $\chi_p$ . Finally, and most relevant for our purposes, this waveform model contains all  $\ell = 2$  modes, omitting the higher-order modes. Hence, our templates implement an incomplete GR sector  $h_{\text{GR}}^{\text{incomplete}}$  that can be expressed as in Eq. 5.

## VI. RESULTS

In this section, we show how the higher modes present in our injected signals, which are omitted by the GR sectors of our templates, can be interpreted as effects due to physics beyond GR. Whether these violations are observed or not will depend on the strength of the higher modes present in the injected signal, which in turn depends on the parameters of the corresponding source. Stronger higher modes will have a larger chance of being

<sup>2</sup> The former model is available in the LIGO LALSimulation library [35] and denoted as EOBNRv2HM.

<sup>3</sup> This model is known as EOBNRv2 in the LALSimulation library.

<sup>4</sup> This model is known as IMRPhenomPv2 in the LIGO Algorithm Library and we will refer to it as PhenomP throughout the text

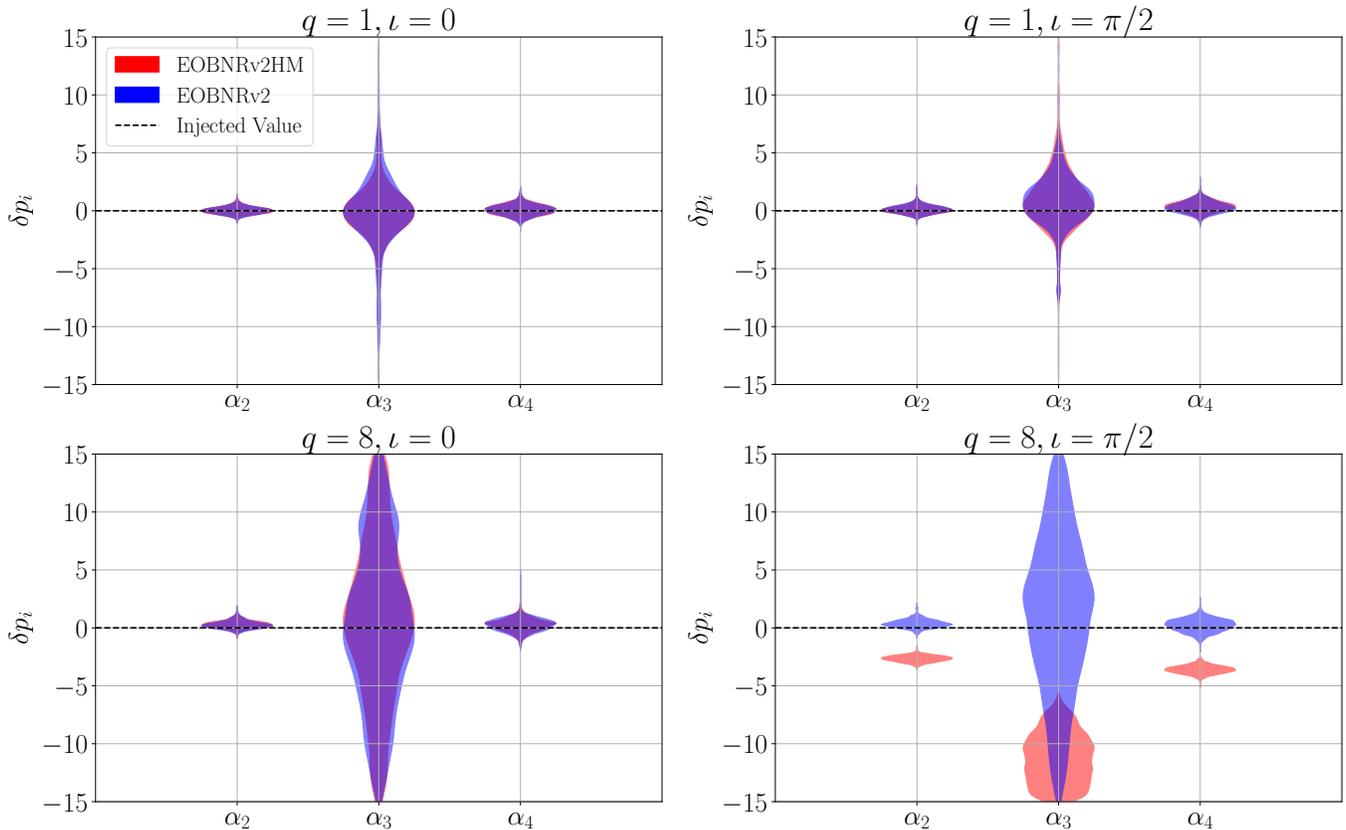


FIG. 2. **Impact of higher modes on parameterized deviations from GR:** Posterior distributions of the nonGR parameters affecting the merger-ringdown emission ( $\alpha_2$ ,  $\alpha_3$  and  $\alpha_4$ ) obtained for four different injections. These correspond to four BBH sources with total mass of  $M = 200M_\odot$ , but different inclination angle  $\iota$  and mass ratio  $q$ . In all cases, we inject signals including (red) and omitting (blue) higher modes. All injections have an optimal SNR  $\rho_{\text{opt}} = 15$ . Note that no significant deviations from GR are observed for  $\iota = 0$ ,  $q = 1$  (upper left panel),  $\iota = \pi/2$ ,  $q = 1$  (upper right panel) and  $\iota = 0$ ,  $q = 8$  (lower left panel). However, a strong deviation from GR is observed the case with  $\iota = \pi/2$ ,  $q = 8$  (lower right panel) when higher modes are included in the injected signal.

interpreted physics beyond GR, while signals for which higher modes are weak will be well recovered by the GR sectors of our templates.

### A. A First Example

The upper panels of Fig. 2 show the results of the parametrized test applied on signals emitted by an equal-mass binary with  $M = 200M_\odot$ , for inclinations  $\iota = 0$  and  $\iota = \pi/2$ . In particular, we show the posterior distributions obtained for each of the nonGR parameters  $\delta\alpha_i$ . In these cases the impact of higher modes in the signal is small and the GR sector of our templates, given by the PhenomP model, can recover most of the power of the signal. This leaves almost no power available to be absorbed by the nonGR sector of the templates. Therefore, no deviations from GR are reported by the test.

In the bottom case, we set the mass ratio of the binary to  $q = 8$ . In the face-on case, higher modes have a minimal effect and no deviations from GR are observed.

However, in the edge-on case, the strong features introduced by higher modes cannot be recovered by the incomplete GR sector. Instead, an important fraction of this is mimicked by the nonGR sector of the templates. As a consequence, none of the posterior distributions includes the GR value  $\delta\alpha_i = 0$ , and the test reports a strong deviation from GR. In all cases the simulated source has been placed at a distance such that its optimal signal-to-noise ratio (SNR) is  $\rho_{\text{opt}} = 15$  (see Eq. 15)

Fig. 3 illustrates the above effect. The left panel shows, in the Fourier domain, the injected signal (green), the GR PhenomP template that best recovers it (blue) and the best fitting waveform allowing for deviations from GR (orange). The GR template is unable to capture the high frequency components introduced by the higher modes of our injection. Instead, contributions from higher modes can be accounted for when one allows for deviations from GR. The right panel shows the same waveforms in the time-domain. In this case, it is particularly clear that the differences between the performance of the GR and beyond GR templates is caused by the strong impact that

higher modes have in the late merger-ringdown part of the signals. In particular, note how the blue GR waveform fails to capture the morphology of our injected signal (green) while this is quite well mimicked by when we allow for deviations from GR (orange).

## B. Predicting False Violations

As suggested in [46], the fractional amount of power in the injections that can not be recovered by the GR sector, and which could instead be picked up by the nonGR one, is the key ingredient to estimate the magnitude of potential deviations from GR. In our case, this can be given by the fraction of SNR that the best matching PhenomP template can recover from our injection [47].

The optimal SNR ratio of an injection  $h_{\text{inj}}$  is given by

$$\rho_{\text{opt}} = \sqrt{(h_{\text{inj}}|h_{\text{inj}})}. \quad (15)$$

Instead, the SNR that a template  $h$  can recover from  $h_{\text{inj}}$  is proportional to their overlap  $\mathcal{O}(h_{\text{inj}}|h)$ , as:

$$\begin{aligned} \rho(h_{\text{inj}}|h) &= \frac{(h_{\text{inj}}|h)}{\sqrt{(h|h)}} = \frac{(h_{\text{inj}}|h)}{\sqrt{(h_{\text{inj}}|h_{\text{inj}})(h|h)}} \times \rho_{\text{opt}} \\ &\equiv \mathcal{O}(h_{\text{inj}}|h) \times \rho_{\text{opt}} \end{aligned} \quad (16)$$

The *effectualness*  $\mathcal{E}$  of a template family (or a template bank) to an injection  $h_{\text{inj}}(\vec{\lambda}_{\text{inj}})$  is defined as the maximum overlap between  $h_{\text{inj}}(\vec{\lambda}_{\text{inj}})$  and any template  $h(\lambda)$  belonging to a given family, maximised over the parameters  $\lambda$  [47]. In our case:

$$\mathcal{E} = \max_{\lambda} \mathcal{O}(h_{\text{inj}}(\vec{\lambda}_{\text{inj}})|h(\vec{\lambda})) \quad (17)$$

The effectualness gives the maximum SNR (or power) from the signal that can be recovered by the non-modified PhenomP templates. The maximum SNR that can be picked by its nonGR sector is given by

$$\rho_{\perp} = \rho_{\text{opt}} \times \sqrt{1 - \mathcal{E}^2}. \quad (18)$$

Computing  $\mathcal{E}$  implies a rather expensive maximization of  $\mathcal{O}$  over all the possible parameters  $\vec{\lambda}_{\text{inj}}$ . Instead, we find that a good proxy for this quantity is given by the *faithfulness* of PhenomP to our injection. This is defined as the overlap of the PhenomP template having the same parameters  $\vec{\lambda}_{\text{inj}}$  as our injection :

$$F = \max_{\phi_c, t_c} \mathcal{O}(h_{\text{inj}}(\vec{\lambda}_{\text{inj}})|h(\vec{\lambda}_{\text{inj}})). \quad (19)$$

Fig. 4, shows the distribution of the strength of deviations from GR as a function of  $F$ . Since injections with low  $\rho_{\text{opt}}$  will have relatively low  $\rho_{\perp}$ , we expect such injections to reveal no clear violations of GR. For this

reason, we choose to distribute our injections uniformly in co-moving volume with  $\rho_{\text{opt}} \in [15, 25]$ .<sup>5</sup>

Red dots correspond to cases in which higher modes are included in the injections. Blue dots correspond to injections having the same parameters as the previous ones but omitting the higher modes. For the latter, we always get  $F \sim 1$ , and all cases can be well recovered by PhenomP, leading to no significant deviations from GR. Addition of higher modes causes  $F$  to deviate from 1 for those cases in which higher modes are strong, leaving a larger power in the injection that can be absorbed by the nonGR sector of our templates. This leads to deviations from GR whose magnitude increases, in average, as  $F$  decreases. Of course, not all the power that is not recovered by PhenomP will be recovered the nonGR sector of the template. Instead, this will be proportional to the overlap between the portion of the injected waveform not recovered by the GR sector and the corresponding nonGR sector, which will vary depending on the specific effect that higher-order modes have on the injection. As a result one may write the signal portion causing the apparent violation of GR as

$$h_{\text{nonGR}}^{\text{apparent}} = \rho(h_{\text{inj}} - h_{\text{GR}}^{\text{incomplete}}|h_{\text{nonGR}})\hat{h}_{\text{nonGR}}, \quad (20)$$

with  $\hat{h}_{\text{nonGR}} = h_{\text{nonGR}}/\sqrt{(h_{\text{nonGR}}|h_{\text{nonGR}})}$ . For some of the cases we considered, the portion of signal not recovered by PhenomP is almost orthogonal to the nonGR sector of the templates, thus yielding no evidence for violations of GR. This, together with the fact that signals are injected in different realizations of Gaussian noise, is partly responsible for the spread in the strength of the violations we see for a given value of  $F$ .

## C. Landscape of False Violations

Fig. 5 shows the density of sources showing violations of GR stronger than  $5\sigma$  as a function of the inclination  $\iota$  and mass-ratio  $q$  of the binary. Not surprisingly, these are more likely to occur for the cases for which higher modes are known to be stronger. This is, highly asymmetric (large  $q$ ) binaries with an orientation far from face-on ( $\iota = 0$ ). For these cases, the strong higher-order mode content of the signals causes the GR sector of the recovery templates (described by the PhenomP model) not being able to recover all the power from the injection, leaving a lot of power available to be picked up by the nonGR sector of the templates [15, 17, 19]. Conversely, we never observe any violation of GR for the case of face-on/off binaries or for binaries with  $q < 3$ .

<sup>5</sup> In principle, we could have extended our range down to  $\rho_{\text{opt}} = 0$ . However, due to the high computational cost of this study, we preferred to avoid performing runs for which the result is fairly predictable.

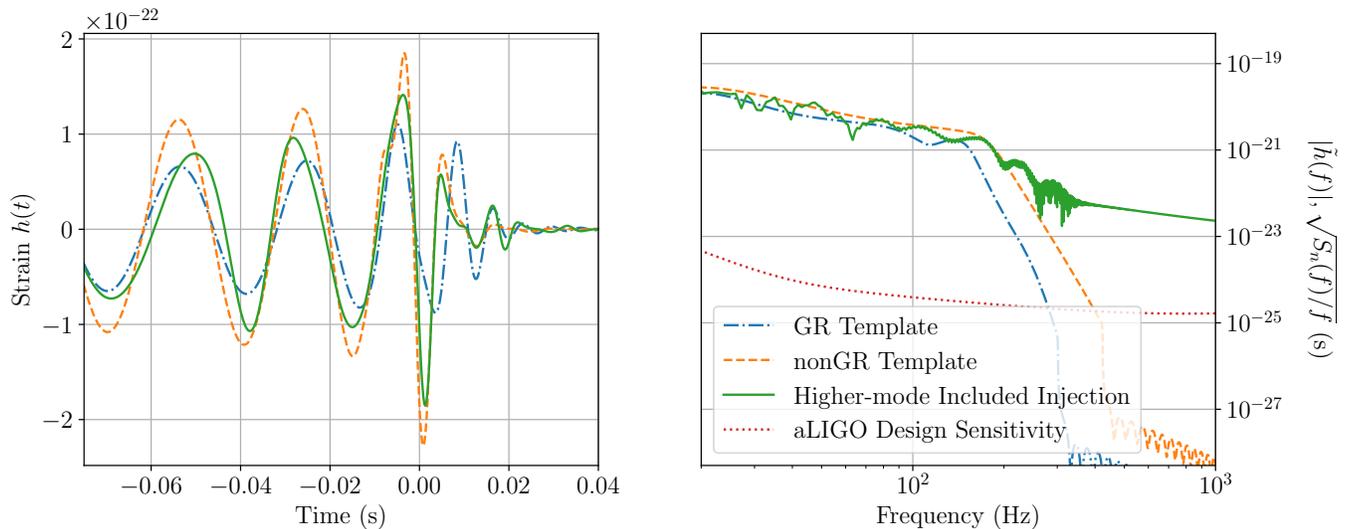


FIG. 3. **Physics beyond GR mimicking the effect of higher modes:** We show the best-fit waveform with (orange) and without (blue) testing GR parameters on top of the injected waveform (green). The latter corresponds to a BBH with  $q \approx 5.5$ ,  $\iota \approx \pi/2$ . The waveforms are plotted in frequency domain (*left panel*) and time domain (*right panel*). Note, in the right panel, how at low frequencies both the GR and beyond GR templates show a similar performance. However, the high frequency content introduced by higher modes can only be recovered by the template enabling physics beyond GR. The right panel shows that most of these differences come from the strong effects that higher modes have in in the late merger and ringdown stages.

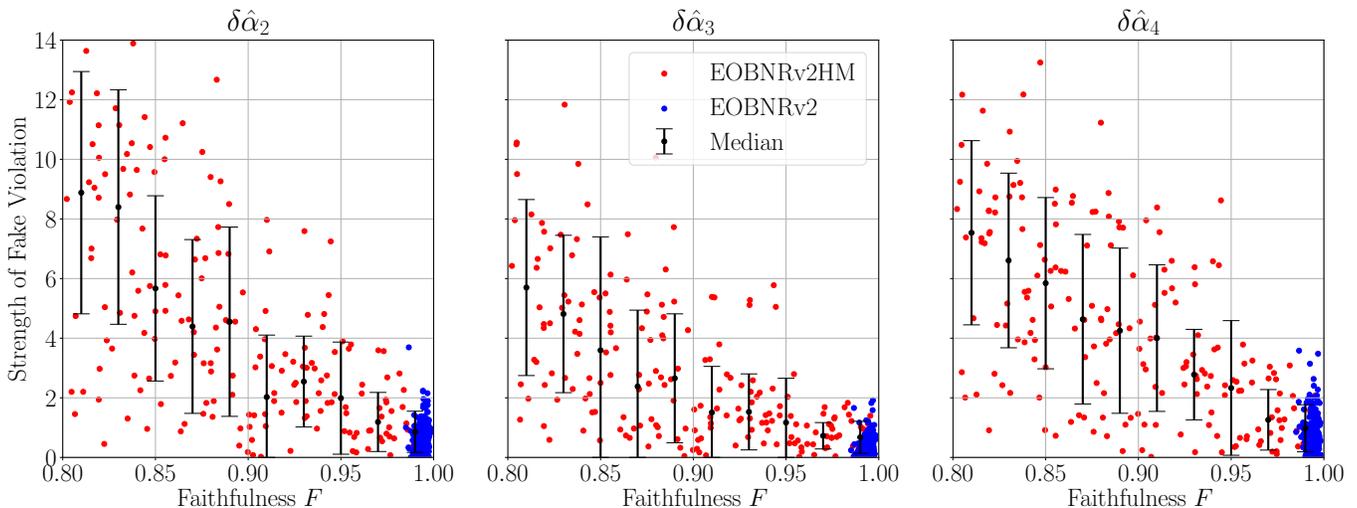


FIG. 4. **Strength of deviations from GR as a function of faithfulness:** The panels show the strength of deviations from GR as a function of the faithfulness between our injections and the recovery PhenomP templates. The injections are distributed uniformly in co-moving volume with their optimal SNRs  $\rho_{\text{opt}}$  distributed between 15 and 25. Note how when higher modes are included in the injections (red dots), the strength of the deviations from GR grows as the faithfulness decreases. When we omit the higher modes in the injections (blue dots) we always obtain a high faithfulness and hence, we don't observe any strong evidence for deviations from GR.

Although the detection of a single violation from GR stronger than  $5\sigma$  might be sufficient to question the correctness of GR, it is natural to ask how this evidence can accumulate with several detections, even if these are weaker than  $5\sigma$ . To address this, we build a set of injections with more astrophysically motivated parameters and perform parametrized tests of GR on them. These

distribute uniformly in  $M$ ,  $q$  and  $\cos(\iota)$ . Fig. 6 shows the distribution of the logarithm of the Bayes factor  $B_{\text{GR}}^{\text{nonGR}}$  for each individual parameterized deviation for our injection set. As before, the red histogram corresponds to the case in which the injections contain higher modes while these are omitted in the blue histograms. As expected, the latter histogram shows no evidence for deviations

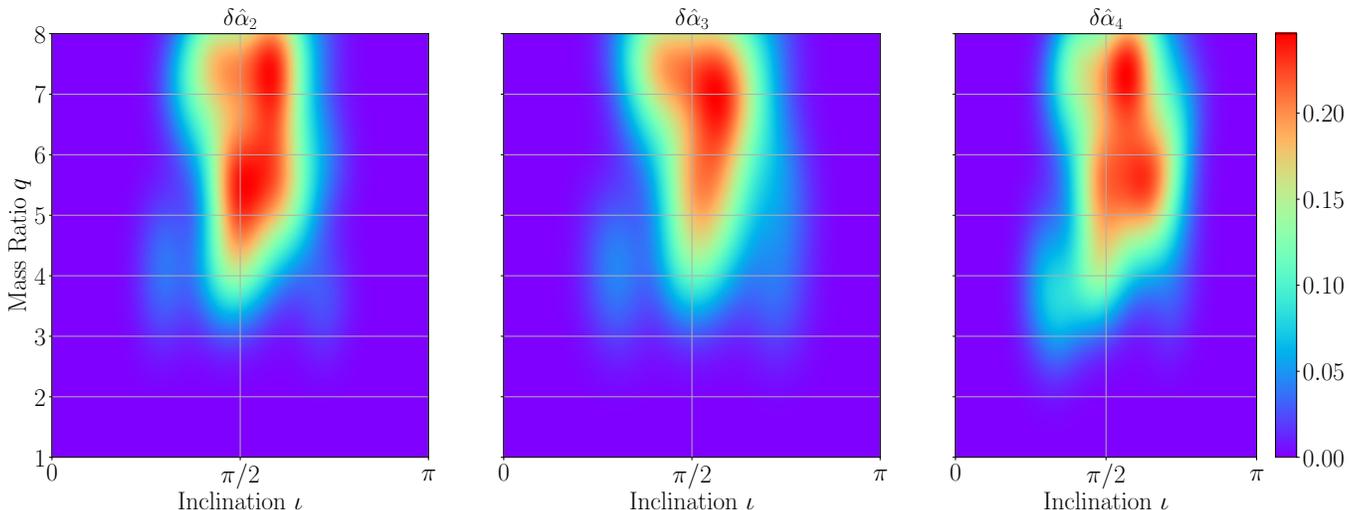


FIG. 5. **Strength of deviations from GR as a function of source parameters:** The three panels show density of sources showing violations of GR stronger than  $5\sigma$  as a function of the mass ratio  $q$  and the inclination  $\iota$  of the source when our injections include higher modes. Significant deviations are more likely to happen for sources with mass ratio  $q > 4$  and inclination close to  $\iota = \pi/2$ , for which higher modes are the strongest.

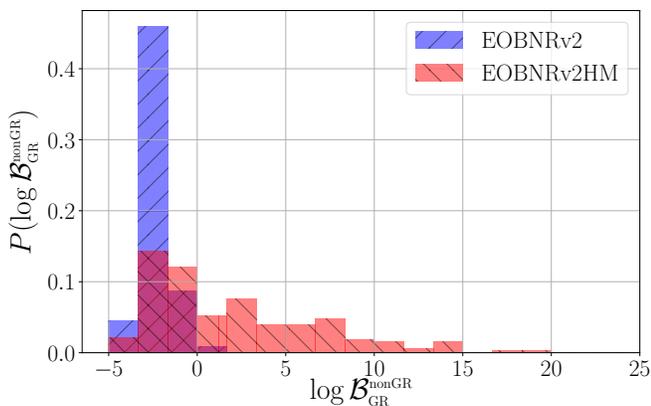


FIG. 6. **Impact of higher modes on model selection:** Distribution of the deviation Bayes factors for merger-ringdown parameters  $\delta\alpha_i$  from 200 samples drawn from a realistic astrophysical distribution (see text for details). Injections with (EOBNRv2HM) and without (EOBNRv2) higher modes are shown in blue and red, respectively. Injections with higher modes included have the nonGR hypothesis  $\mathcal{H}_{\text{nonGR}}$  outrank the GR hypotheses  $\mathcal{H}_{\text{GR}}$  in a significant fraction of injections.

from GR (i.e,  $\log B < 0$ ) while clear excursions toward positive values are obtained when our injections contain higher modes.

#### D. Rate of False Violations of GR

To close this section, we address the question of how long would it take to observe a fake violation of GR using current parametrized tests. To assess this we con-

sider a source distribution more realistic than the one in the previous paragraph. This is driven by the recent population distributions and merger rates of BBH recently obtained by the LSC. In particular, we consider the same standard BBH population used by the LSC in [49]: a power law distribution  $m_1$  and uniform in  $m_2$   $p(m_1, m_2) \propto m_1^{-\alpha}/(m_1 - 5M_\odot)$  and consider the merging rates derived from the LIGO First Observation Run [48] and GW170104[7]. We distribute our injections uniformly in co-moving volume with their optimal SNRs  $\rho_{\text{opt}}$  ranging between 8 and 30. With these ingredients, we can compute the time that it would take to “detect” a signal  $h_{\text{inj}}$  such that performing a parametrized test on it would lead to a false violation stronger than  $N\sigma$ . For our purposes, “detect” means that the corresponding source would be observed with an SNR  $\rho_{\text{obs}} \geq 8$ , which is the standard threshold used in the literature. Moreover, we assume that the search observing the event can recover its full optimal SNR  $\rho_{\text{opt}} = \sqrt{\langle h_{\text{inj}} | h_{\text{inj}} \rangle}$ . Hence, we only need to compute how many sources show strong fake deviations and coalesce close enough to us to produce  $\rho_{\text{opt}} \geq 8$  per unit time.

We note that assuming  $\rho_{\text{obs}} = \rho_{\text{opt}}$  may be too optimistic. In fact, matched filter searches are designed to recover  $\rho_{\text{obs}} \geq 0.956\rho_{\text{opt}}$  in the regions of the parameter space that are well covered by their template banks and, more importantly, these are not currently sensitive enough to signals with strong higher modes [23, 50]. However, promising progress has been made to make these searches sensitive to such signals [27]. Furthermore, more generic searches which remain agnostic to the morphology of the signals [51] should ideally be able to recover all of the optimal SNR of any kind of signal.

Fig. 7 shows the expected observation time for ob-

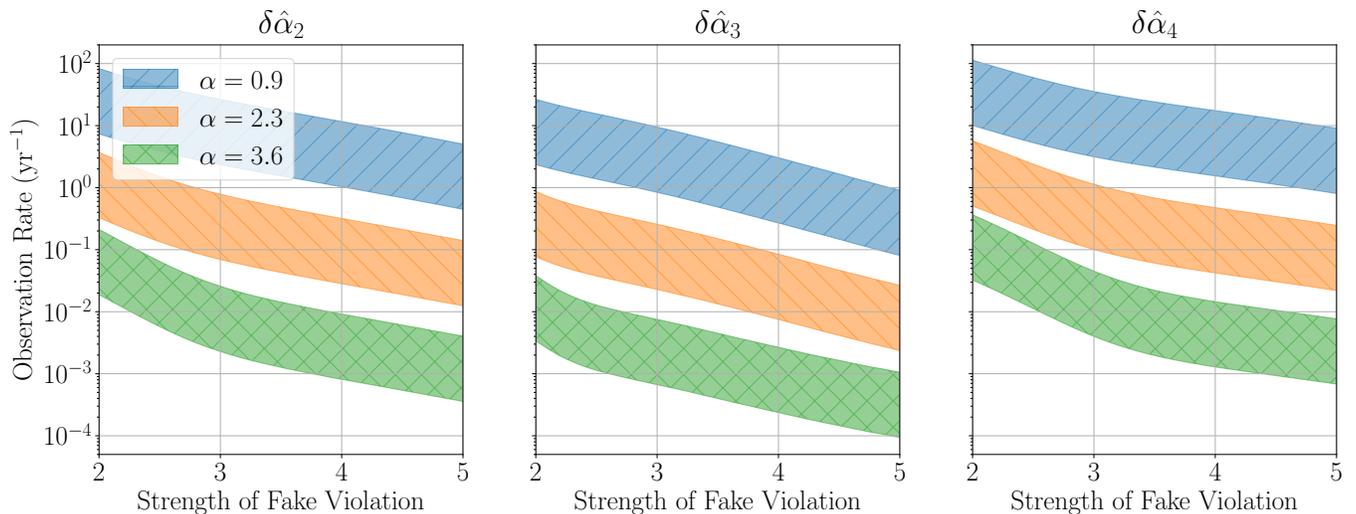


FIG. 7. **Expected observation rate of false violations of GR with a given strength:** The expected observation rate of a detection with a false violations of GR greater than a given strength for three source populations with different power law coefficients  $\alpha$ , where  $p(m_1, m_2) \propto m_1^{-\alpha}/(m_1 - 5M_\odot)$ , and with merger rates obtained from Ref. [48]. The injections are distributed uniformly in co-moving volume with their optimal SNRs distributed between 8 and 30. The strength is measured by the number of standard deviations  $\sigma$  between the median of the posterior distribution of the testing GR parameter and the GR value ( $\delta p_i = 0$ ). The plots report values for the testing GR parameters  $\delta\hat{\alpha}_2$  (left panel),  $\delta\hat{\alpha}_3$  (middle panel) and  $\delta\hat{\alpha}_4$  (right panel). For the flattest power law ( $\alpha = 0.9$ ), an observation of a  $5\sigma$  deviation from GR is expected to occur around once per year.

servicing a false violation of GR with given strength coming from each of the considered nonGR parameters, and assuming a three detector network formed by advanced LIGO and advanced Virgo operating at their design sensitivities. Results are shown for source populations corresponding to three different values of  $\alpha$ . In particular, these are chosen to be the 5th, 50th and 95th percentile of the  $\alpha$  distribution inferred in [49]. The width of the bands corresponds to the 90% credible interval of the event based merger rate.

We note again that for a signal to show strong false violations of GR, it requires a strong higher order mode content. Hence, the corresponding source should have a large mass ratio and be fairly edge-on. Signals from these sources are much weaker than signals from nearly equal-mass ones, or face-on sources. It is therefore reasonable that the vast majority of current and future BBH detections will mostly correspond to the latter group, reducing the chances of observing false violations from GR produced by higher modes.

Consistently, in the worst case, corresponding to the source distribution having  $\alpha = 0.9$ , we estimate that a fake violation with a significance of  $5\sigma$  can occur once per year (left and right panels) of observation. Although the probability of observing this kind of effect seems quite low, a single of these observations might be enough to put into question the correctness of the theory of General Relativity.

In this work we have only focused on biases produced on parameters accounting for deviations from GR due to the omission of higher modes in our templates. How-

ever, consistently with other studies like [16, 18, 19], we also obtain biased values for other parameters, like the total mass  $M$ , mass ratio  $q$  or inclination angle  $\iota$ . Existing work has shown how these biases are corrected once higher modes are included into the parameter estimation analyses [21] for the case of non-spinning binaries and how the accuracy of the measurement can be improved [39]. These improvements may then also reduce the bias in beyond-GR parameters.

## VII. CONCLUSIONS

The direct detection of gravitational waves from the coalescence of binary black holes has allowed for tests General Relativity in the strong gravitational field regime. So far, the results of all tests are consistent with General Relativity being correct. In this paper, we have focused on one of these tests, known as parametrized test, which seeks evidence for physics beyond General Relativity in the incoming signals. To this, gravitational wave signals are compared to waveform templates that account for possible deviations from General Relativity. In doing this, it is crucial that part ( or “sector”) of these waveforms that accounts for physics within General Relativity includes all of its effects. Otherwise, missing physics might be mimicked by the non-GR sector of these templates. This would lead the analysis to report false apparent violations of General Relativity.

Here, we have showcased this effect using the example of current parametrised tests of General Relativity

performed on signals from binary black holes, which currently omit physical phenomena like higher modes and eccentricity. In particular, we have demonstrated that when higher-order modes are present in the signal but are omitted in the templates, its effect can be mimicked by physics beyond General Relativity, leading to apparent violations. To this end, we have performed parametrized tests on simulated signals containing higher-order modes. We have done this for a large family of non-spinning binaries with total masses ranging in  $[100, 200]M_{\odot}$ , mass ratios  $q \leq 8$  and varying inclinations. Strong deviations from General Relativity, with significance larger than  $5\sigma$ , are reported for highly asymmetric binaries ( $q \geq 4$ ) whose orbital angular momentum is almost orthogonal to the line-of-sight (or edge-on oriented). This is consistent these sources being known to have a strong higher-order mode emission.

Finally, considering several astrophysically motivated source distributions and assuming an Advanced LIGO - Advanced Virgo network operating at design sensitivity, we have computed the rate at which false violations could arise if current parametrized test of GR are performed in future detections. In the most pessimistic scenario, we estimate that false observations from GR could be observed once per year. We note however that the low mass ratio of the binary black holes detected by Advanced LIGO and Virgo so far makes it unlikely that the effects we have described could affect the tests of GR based on these ob-

servations.

We stress that even though we have focused on parametrized tests of General Relativity and showcased the impact of omission of higher modes, there is a whole family of similar tests which involve the comparison of data with templates that are modified to account for given effects beyond General Relativity. Examples of these are searches for dipole radiation or dispersion effects. It is reasonable to expect that the effects described in this work do affect these and similar studies. Hence, we stress the importance of using complete waveform models (ideally including higher modes, eccentricity and precession) and quantify some of the systematics when this is not the case.

### VIII. ACKNOWLEDGMENTS

The authors are grateful for useful comments by Salvatore Vitale, Parameswaran Ajith, Chris Van Den Broeck and Abhay Ashtekar. JCB gratefully acknowledges support from the NSF grants 1505824, 1505524, 1333360. PTHP, YFW and TGFL were partially supported by a grant from the Research Grants Council of Hong Kong (Project No. CUHK 24304317), and the Direct Grant for Research from the Research Committee of the Chinese University of Hong Kong.

- 
- [1] A. Einstein, *Annalen Phys.* **49**, 769 (1916), [Annalen Phys.14,517(2005)].
  - [2] N. Wex, ArXiv e-prints (2014), arXiv:1402.5594 [gr-qc].
  - [3] C. M. Will, *Living Reviews in Relativity* **17**, 4 (2014), arXiv:1403.7377 [gr-qc].
  - [4] E. Berti, E. Barausse, V. Cardoso, L. Gualtieri, P. Pani, U. Sperhake, L. C. Stein, N. Wex, K. Yagi, T. Baker, C. P. Burgess, D. Doneva, A. D. Felice, P. G. Ferreira, P. C. C. Freire, J. Healy, C. Herdeiro, M. Horbatsch, B. Kleihaus, A. Klein, K. Kokkotas, J. Kunz, P. Laguna, R. N. Lang, T. G. F. Li, T. Littenberg, A. Matas, S. Mirshekari, H. Okawa, E. Radu, B. S. Sathyaprakash, C. V. D. Broeck, H. A. Winther, H. Witek, M. E. Aghili, J. Alsing, B. Bolen, L. Bombelli, S. Caudill, L. Chen, J. C. Degollado, R. Fujita, C. Gao, D. Gerosa, S. Kamali, H. O. Silva, L. Sadeghian, M. Sampaio, H. Sotani, and M. Zilhao, *Classical and Quantum Gravity* **32**, 243001 (2015).
  - [5] B. P. Abbott et al. (Virgo, LIGO Scientific), *Phys. Rev. Lett.* **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].
  - [6] B. Abbott et al. (Virgo, LIGO Scientific), *Phys. Rev. Lett.* **116**, 241103 (2016), arXiv:1606.04855 [gr-qc].
  - [7] B. Abbott et al. (VIRGO, LIGO Scientific), *Phys. Rev. Lett.* **118**, 221101 (2017).
  - [8] B. P. Abbott et al. (Virgo, LIGO Scientific), *Phys. Rev. Lett.* **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].
  - [9] B. P. Abbott et al. (Virgo, LIGO Scientific), *Phys. Rev. Lett.* **116**, 221101 (2016), arXiv:1602.03841 [gr-qc].
  - [10] J. Aasi et al. (LIGO Scientific), *Class. Quant. Grav.* **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].
  - [11] F. Acernese et al. (VIRGO), *Class. Quant. Grav.* **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].
  - [12] B. P. Abbott et al. (Virgo, LIGO Scientific), *Phys. Rev. Lett.* **119**, 141101 (2017), arXiv:1709.09660 [gr-qc].
  - [13] M. Isi and A. J. Weinstein, (2017), arXiv:1710.03794 [gr-qc].
  - [14] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, et al., *Phys.Rev.Lett.* **113**, 151101 (2014), arXiv:1308.3271 [gr-qc].
  - [15] L. Pekowsky, J. Healy, D. Shoemaker, and P. Laguna, *Phys.Rev.* **D87**, 084008 (2013), arXiv:1210.1891 [gr-qc].
  - [16] J. Calderón Bustillo, S. Husa, A. M. Sintes, and M. Pürrer, *Phys. Rev.* **D93**, 084019 (2016), arXiv:1511.02060 [gr-qc].
  - [17] J. Calderón Bustillo, P. Laguna, and D. Shoemaker, *Phys. Rev.* **D95**, 104038 (2017), arXiv:1612.02340 [gr-qc].
  - [18] V. Varma, P. Ajith, S. Husa, J. C. Bustillo, M. Hannam, et al., *Phys.Rev.* **D90**, 124004 (2014), arXiv:1409.2349 [gr-qc].
  - [19] V. Varma and P. Ajith, (2016), arXiv:1612.05608 [gr-qc].
  - [20] C. Capano, Y. Pan, and A. Buonanno, *Phys.Rev.* **D89**, 102003 (2014), arXiv:1311.1286 [gr-qc].
  - [21] P. B. Graff, A. Buonanno, and B. S. Sathyaprakash, *Phys. Rev.* **D92**, 022002 (2015), arXiv:1504.04766 [gr-qc].

- [22] B. P. Abbott et al. (Virgo, LIGO Scientific), *Phys. Rev. D* **96**, 022001 (2017), [arXiv:1704.04628 \[gr-qc\]](#).
- [23] S. A. Usman et al., *Class. Quant. Grav.* **33**, 215004 (2016), [arXiv:1508.02357 \[gr-qc\]](#).
- [24] T. Dal Canton and I. W. Harry, (2017), [arXiv:1705.01845 \[gr-qc\]](#).
- [25] S. Klimentko, G. Vedovato, M. Drago, F. Salemi, V. Tiwari, G. A. Prodi, C. Lazzaro, K. Ackley, S. Tiwari, C. F. Da Silva, and G. Mitselmakher, *Phys. Rev. D* **93**, 042004 (2016).
- [26] J. Calderón Bustillo, F. Salemi, T. Dal Canton, and K. Jani, (2017), [arXiv:1711.02009 \[gr-qc\]](#).
- [27] I. Harry, J. Calderón Bustillo, and A. Nitz, (2017), [arXiv:1709.09181 \[gr-qc\]](#).
- [28] A. H. Nitz, (2017), [arXiv:1709.08974 \[gr-qc\]](#).
- [29] P. Jaranowski, A. Krolak, and B. F. Schutz, *Phys. Rev. D* **58**, 063001 (1998), [arXiv:gr-qc/9804014 \[gr-qc\]](#).
- [30] E. Berti, V. Cardoso, J. A. Gonzalez, U. Sperhake, M. Hannam, S. Husa, and B. Bruggmann, *Phys. Rev. D* **76**, 064034 (2007), [arXiv:gr-qc/0703053 \[GR-QC\]](#).
- [31] T. G. F. Li, W. Del Pozzo, S. Vitale, C. Van Den Broeck, M. Agathos, J. Veitch, K. Grover, T. Sidery, R. Sturani, and A. Vecchio, *Phys. Rev. D* **85**, 082003 (2012).
- [32] D. Shoemaker et al. (LIGO Scientific Collaboration), LIGO-T0900288, <https://dcc.ligo.org/cgi-bin/DocDB/ShowDocument?docid=2974> (2010).
- [33] J. C. Bustillo, A. Bohé, S. Husa, A. M. Sintes, M. Hannam, et al., (2015), [arXiv:1501.00918 \[gr-qc\]](#).
- [34] 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), [arXiv:1409.7215 \[gr-qc\]](#).
- [35] “The LIGO algorithms library.”
- [36] Y. Pan, A. Buonanno, M. Boyle, L. T. Buchman, L. E. Kidder, H. P. Pfeiffer, and M. A. Scheel, *Phys. Rev. D* **84**, 124052 (2011), [arXiv:1106.1021 \[gr-qc\]](#).
- [37] T. Damour, A. Nagar, and S. Bernuzzi, *Phys. Rev. D* **87**, 084035 (2013), [arXiv:1212.4357 \[gr-qc\]](#).
- [38] R. Cotesta, A. Buonanno, A. Bohé, A. Taracchini, I. Hinder, and S. Ossokine, (2018), [arXiv:1803.10701 \[gr-qc\]](#).
- [39] L. London, S. Khan, E. Fauchon-Jones, X. J. Forteza, M. Hannam, S. Husa, C. Kalaghatgi, F. Ohme, and F. Pannarale, (2017), [arXiv:1708.00404 \[gr-qc\]](#).
- [40] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044006 (2016), [arXiv:1508.07250 \[gr-qc\]](#).
- [41] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044007 (2016), [arXiv:1508.07253 \[gr-qc\]](#).
- [42] A. Taracchini, Y. Pan, A. Buonanno, E. Barausse, M. Boyle, et al., *Phys. Rev. D* **86**, 024011 (2012), [arXiv:1202.0790 \[gr-qc\]](#).
- [43] A. Taracchini, A. Buonanno, Y. Pan, T. Hinderer, M. Boyle, et al., *Phys. Rev. D* **89**, 061502 (2014), [arXiv:1311.2544 \[gr-qc\]](#).
- [44] A. Buonanno, G. Faye, and T. Hinderer, *Phys. Rev. D* **87**, 044009 (2013), [arXiv:1209.6349 \[gr-qc\]](#).
- [45] L. Santamaria, F. Ohme, P. Ajith, B. Bruegmann, N. Dorband, et al., *Phys. Rev. D* **82**, 064016 (2010), [arXiv:1005.3306 \[gr-qc\]](#).
- [46] M. Vallisneri, *Phys. Rev. D* **86**, 082001 (2012), [arXiv:1207.4759 \[gr-qc\]](#).
- [47] T. A. Apostolatos, *Phys. Rev. D* **52**, 605 (1995).
- [48] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, B. Allen, A. Allocca, P. A. Altin, S. B. Anderson, W. G. Anderson, K. Arai, M. C. Araya, C. C. Arceneaux, J. S. Areeda, N. Arnaud, K. G. Arun, S. Ascenzi, G. Ashton, M. Ast, S. M. Aston, P. Astone, P. Aufmuth, C. Aulbert, S. Babak, P. Bacon, and M. K. M. Bader (LIGO Scientific Collaboration and Virgo Collaboration), *Phys. Rev. X* **6**, 041015 (2016).
- [49] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Afrough, B. Agarwal, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, B. Allen, G. Allen, A. Allocca, P. A. Altin, A. Amato, A. Ananyeva, S. B. Anderson, W. G. Anderson, S. Antier, S. Appert, K. Arai, M. C. Araya, J. S. Areeda, N. Arnaud, K. G. Arun, S. Ascenzi, G. Ashton, M. Ast, S. M. Aston, P. Astone, P. Aufmuth, C. Aulbert, K. AultO’Neal, and Avila-Alvarez (LIGO Scientific and Virgo Collaboration), *Phys. Rev. Lett.* **118**, 221101 (2017).
- [50] A. H. Nitz, T. Dent, T. Dal Canton, S. Fairhurst, and D. A. Brown, (2017), [arXiv:1705.01513 \[gr-qc\]](#).
- [51] S. Klimentko, I. Yakushin, A. Mercer, and G. Mitselmakher, *Classical and Quantum Gravity* **25**, 114029 (2008), [arXiv:0802.3232 \[gr-qc\]](#).