Effect of Type II Strong Gravitational Lensing on Tests of General Relativity

Purnima Narayan, Nathan K. Johnson-McDaniel, and Anuradha Gupta

Department of Physics and Astronomy, The University of Mississippi, University, Mississippi 38677, USA

(Dated: November 23, 2024)

Gravitational wave (GW) observations of binary black hole (BBH) coalescences provide a unique opportunity to test general relativity (GR) in the strong-field regime. To ensure the reliability of these tests, it is essential to identify and address potential sources of error, particularly those arising from missing physics in the waveform models used in GW data analysis. This paper investigates potential biases in these tests arising from strong gravitational lensing, an effect not currently incorporated into the standard framework for GR tests. In the geometric optics approximation, strong lensing produces three types of images: Type I, Type II, and Type III. While Type I and Type III images do not distort the signal, Type II images introduce a characteristic phase shift that can mimic GR deviations for signals with higher-order modes, precession, or eccentricity. We assess the response of four standard GR tests on simulated Type II lensed BBH signals, including the two parameterized tests (TIGER and FTI), the modified dispersion relation test and the inspiral-merger-ringdown consistency test. We focus on precessing waveforms for binaries with total masses of $20M_\odot$ and $80M_\odot$, and dimensionless spins of 0.5 and 0.95, considering a fixed signal-to-noise ratio of 25 using the design A+ sensitivity of the LIGO-Virgo network. Our findings indicate that more mass-asymmetric and higher-spin binaries show larger false deviations from GR in the TIGER and modified dispersion relation tests when applying GR tests to Type II lensed signals. These results highlight the risk of false GR violations as detector sensitivity improves in future observational runs. Therefore, it is crucial to consider the possibility of strong lensing before drawing conclusions about deviations from GR in GW signals.

I. INTRODUCTION

Einstein's theory of general relativity (GR) has been extensively tested and validated in weak-field, low-speed, and linear gravity regimes [1–5]. However, the detection of gravitational waves (GWs) from binary black hole (BBH) mergers provides a unique opportunity to probe GR in the highly nonlinear, strong-field regime, which is inaccessible through other means. Testing GR using GWs relies on waveform models that are compared with the detected GW data. Further, the current tests of GR carried out by the LIGO-Virgo-KAGRA (LVK) collaboration, are based on waveform models that neglect certain physical effects that could be present in the signal. One such effect is strong gravitational lensing, wherein a mass located between the source and observer bends the GW signal, potentially producing multiple images with different magnifications and arrival times (see, e.g., [6, 7]). Another effect that is not currently included in GW tests of GR is eccentricity which is found to produce biases [8–12].

We focus on the effect due to Type II lensing that causes nontrivial deformations of the signal, as discussed below. While searches for strong lensing in GW data have yet to detect a definitive signature [13–15], the likelihood of detecting such events will increase as detectors increase in sensitivity. For instance, for the LIGO-Virgo-KAGRA network with the LIGO detectors at A+ sensitivity and Virgo and KAGRA at design sensitivity, Ref. [16] finds that one could expect several strongly lensed events per year, while third-generation observatories such as Cosmic Explorer (CE) [17] and the Einstein Telescope (ET) [18] could each detect over 100 lensed events per year [19]. Thus, one has to consider the possibility of a Type II lensed signal potentially leading to a false deviation from GR: While one expects the LVK lensing analyses (as in [14, 15]) to be able to detect the effects of Type II lensing in many cases (since they specifically look for the dephasing present in Type II signals), as well as to potentially be able to identify companion lensed images to a Type II lensed signal, it is still important to check the response of standard tests of GR to such a signal, particularly since the tests of GR are carried out simultaneously with the lensing analyses.

One obtains Type II signals in the geometric optics treatment of gravitational lensing, which has been extensively studied in the literature (see, e.g., [20]). In the short-wavelength, geometric optics regime, such as the lensing of GWs by galaxies, the Kirchhoff diffraction integral describing the effects of lensing simplifies to a Gaussian integral around extremal points on the image plane, corresponding to three types of images: Type I (minimum), Type II (saddle), and Type III (maximum) [21]. Of these, Type II images introduce a non-trivial distortion to the waveform in all but specialized cases, such as highly symmetric binaries [21]. Given that Type II images are expected to occur in over 90% of strongly lensed events [19], it is crucial to investigate their effects in order to properly assess any potential deviations from GR.

Previous studies have observed biases in parameter estimation when Type II images are analyzed with unlensed waveform models [22]. Recent studies have also explored the connection between the phase shift introduced by Type II lensing and deviations from GR, showing that some GR deviations are flagged by pipelines designed to search for Type II lensed images [23, 24]. Our work extends these findings, demonstrating that the application of standard LVK tests of GR to Type II lensed signals can result in false positives, indicating strong GR deviations. Therefore, before any claims of a GR violation are made, it is essential to rule out the possibility of the signal originating from a strongly lensed binary.

Specifically, we consider the Test Infrastructure for GEneral Relativity (TIGER) [25, 26], Flexible-Theory-Independent (FTI) [27], Modified Dispersion Relation (MDR) [28] and Inspiral-Merger-Ringdown (IMR) consis-

tency [29, 30] tests and check their response to simulated Type II lensed BBH GW signals in the LIGO-Virgo network at its design O5 (A+) sensitivity [31]. We consider signals from precessing BBHs with redshifted masses of $20M_\odot$ and $80M_\odot$, each with mass ratios of $2, 5$ and spins of $0.5, 0.95$, and all having an inclination angle of $\pi/3$. Each binary is placed at a luminosity distance (with unit magnification) so that it has a network signal-to-noise ratio (SNR) of 25. For comparison, we also apply these tests to the unlensed versions of these signals for selected testing parameters.

We find TIGER to be the most sensitive to Type II lensing effects. The absence of precession in current FTI implementations results in notable systematics for both lensed and unlensed signals. For each binary configuration, the testing parameter corresponding to frequency independent phase shift shows the strongest support for MDR. And finally for IMR consistency test, we focus on $80M_\odot$ binaries due to SNR considerations, finding that while the lower mass ratio binaries remain consistent with GR, the higher mass ratio binaries show GR inconsistencies in both lensed and unlensed cases. We also investigate the nature of the GR parameters when analyzing lensed waveforms with unlensed waveform models and find biases in the source parameters for TIGER, FTI and MDR tests.

This paper is organized as follows: In Sec. II, we explain the effect of the strong lensing on GWs and describe the method followed to simulate Type II lensed signals. In Sec. III we give the specifics of our simulated observations and in Sec. IV, we give the details of the four tests of GR we consider. In Sec. V, we discuss the results from our analysis and we conclude in Sec. VI. We use geometrized ($G = c = 1$) units throughout.

II. THEORY

In this section we discuss how Type II lensed GW signals are simulated for our study and how the presence of higher modes or precession makes the Type II phase shift lead to a nontrivial distortion of the waveform.

In the frequency domain and geometric optics approximation, the strongly lensed GW signal is [21]

$$\tilde{h}^L_{+,\times}(f) = \sum_j |\mu_j|^{1/2} \exp\left(2\pi i f t_d - i n_j \frac{\pi}{2} \operatorname{sign}(f)\right) \tilde{h}_{+,\times}(f),$$

(1)

where $+$ and $\times$ represent the plus and cross polarizations; μ_j is the magnification of the jth image; f is the frequency of the GW; t_d is the geometric time delay caused by lensing; the Morse index n_j takes the values 0, 1, and 2 for Type I, II, and III images respectively; and $\tilde{h}_{+,\times}(f)$ is the unlensed frequency-domain GW signal.

For a Type I image, there is no extra phase shift, and the lensed waveform is equivalent to the unlensed one apart from the magnification and $2\pi f t_d$ term, which accounts for the travel time to the observer and cannot be measured in the GW signal. For a Type III image, the phase shift equals π for all

frequencies, resulting in a waveform that flips its sign. This sign flip can be compensated exactly by a $\pi/2$ shift in the polarization angle, as explained in Appendix D of [21]. A nontrivial effect occurs for a Type II image, where all positive-frequency components are shifted by a phase $-\pi/2$, and all negative-frequency components are shifted by $\pi/2$. This results in an overall phase shift of $-(\pi/2)\operatorname{sign}(f)$, making Type II images equivalent to a Hilbert transform of Type I images. Since the GW strain is real-valued, we have $\tilde{h}(-f) = \tilde{h}^*(f)$, where the star denotes the complex conjugate. The Type II phase shift maintains this relation, as it must, and we can thus restrict our attention to the positive frequencies, as is standard in GW data analysis, since the negative frequencies do not provide any additional information. The Type II lensed signal that we use to construct our lensed waveform then is

$$\tilde{h}^{L,\text{II}}_{+,\times}(f) = -i\tilde{h}_{+,\times}(f),$$

(2)

where we have taken the magnification to be unity. We make this assumption following studies that indicate that the distribution of magnifications of Type II lensed images peaks close to 1 [19]. Additionally, since the SNR and redshifted masses are fixed in our analysis, and we are considering BBHs, variations in magnification merely correspond to changes in the binary's luminosity distance.

We now consider the situations in which Type II lensing leads to a nontrivial deformation of the signal (as discussed in [21]). We start by considering a quasicircular, aligned-spin system, where the frequency domain Type II lensed signal with unit magnification is given (for $f > 0$) by:

$$\tilde{h}^{L,\text{II},\parallel}_{+,\times}(f) = \sum_{\ell \geq 2} \sum_{m \geq 0} \mathcal{A}_{\ell m}(f)\mathcal{Y}_{\ell m,+,\times}(\iota)\exp\left[i\left(m\phi_c - \frac{\pi}{2}\right)\right],$$

(3)

where $\parallel$ denotes that this just holds for aligned-spin systems, $\mathcal{A}_{\ell m}(f)$ represents the amplitude of each (spin-weighted spherical harmonic) mode, ϕ_c is the coalescence phase, and $\mathcal{Y}_{\ell m,+,\times}(\iota)$ encodes the dependence on the inclination angle ι.

For a simple quasicircular, aligned-spin binary dominated by the $\ell = 2, m = 2$ mode, the $\pi/2$ phase shift from Type II lensing can be absorbed into the coalescence phase ϕ_c via the shift $\phi_c \to \phi_c + \pi/4$. However, the phase shift required to absorb the lensing effect depends on the mode number m, so if there is a nonnegligible contribution from a mode with $m \neq 2$, the ϕ_c shift for the $m = 2$ mode will not provide the shift of $\phi_c \to \phi_c + \pi/(2m)$ necessary to account for the lensing phase shift seen in the $m \neq 2$ mode. This mode-dependent phase shift causes a shift in the phase evolution across different modes, resulting in a nontrivial distortion of the overall waveform, illustrated in the time domain in, e.g., [19, 22]. A binary exhibiting observable precession will have modes with different $|m|$ values present in the inertial frame, so it will also experience a nontrivial distortion from Type II lensing. We thus will consider precessing, mass-asymmetric binaries that have significant higher mode content in this study. Eccentric waveforms also have nontrivial distortions from Type II lensing, as discussed in [21], but we do not consider them here.

III.   SIMULATED OBSERVATIONS AND PARAMETER ESTIMATION SETUP

We simulate Type II lensed BBH image strains in the LIGO-Virgo network by applying Eq. (2) to standard waveform models (discussed below) for unlensed signals from quasicircular, precessing BBHs. We consider the network at the sensitivity predicted for O5/A+ [31] (with the more sensitive Virgo noise curve) when analyzing the observations. We do not include KAGRA [32] owing to the significant uncertainties in its plus-era sensitivity (see [33]). To avoid biases from particular noise realizations, we do not incorporate noise in our simulated observations (i.e., we take the zero realization of Gaussian noise). We will also use the standard GW data analysis terminology of referring to these simulated GW observations as injections.

We select two (redshifted) total masses, $20M_\odot$ and $80M_\odot$, to examine the lensing effects on an inspiral-dominant binary and one where higher-order modes have a greater impact on the inference. For the lower mass ($20M_\odot$), the signal falls in the detectors' most sensitive range mainly during the inspiral phase with at most 6 precessional cycles in band. In contrast, for the $80M_\odot$ binary, the merger and ringdown phases lie in a more sensitive frequency range for the detectors with just 1 precessional cycle in band. Since the amplitudes of higher modes are larger in the merger-ringdown phase than earlier in the inspiral, they thus have a larger effect on the parameter estimation (PE) in this case, though they remain subdominant.

For each total mass, we study two mass ratios $q \in \{2, 5\}$ and two dimensionless spin magnitudes such that $\chi_1 = \chi_2 = \chi \in \{0.5, 0.95\}$. In total, we consider simulated observations of 8 Type II lensed binaries, each with an inclination angle of $\pi/3$ and spin tilt angles of $(2.03, 0.43)$ rad, where all spin directions are specified at 20 Hz. We choose the component of the spin of the primary in the orbital plane to be in the direction of the vector from the smaller to the larger black hole, and the difference between the azimuthal angles of the individual spin vectors to be $\phi_{12} = 2.59$ rad. For the extrinsic parameters, for each binary we choose the luminosity distance (with unit magnification) such that the network SNR is 25 for the fixed randomly chosen values for the other extrinsic parameters we use: right ascension (5.41 rad), declination (0.88 rad), polarization angle (2.59 rad), coalescence GPS time (1126259642.413 s), and coalescence phase (2.27 rad). We give the mass ratios, spins and luminosity distances of the simulations considered in Table I.

We analyze all injections starting from a lower Fourier frequency of 20 Hz, with upper limits of 896 Hz for the $20M_\odot$ cases and 448 Hz for the $80M_\odot$ cases, except where different frequency limits are applied in the inspiral and postinspiral analyses for the IMR consistency test. The upper frequencies are chosen to be sufficiently high that the SNR in the frequencies above the cutoff is negligible. Specifically, they come from the Nyquist frequencies associated with the sampling frequencies of 2048 Hz and 1024 Hz used for the $20M_\odot$ and $80M_\odot$ cases, respectively, after accounting for the roll-off factor of 0.875 that accounts for the effects of a window function (as discussed in Appendix E of [34]). Parameter es-

TABLE I. The lensed simulations we consider and their properties. All the binaries have a luminosity distance chosen such that the network SNR is 25. The dimensionless spins of the two black holes are the same.

mass ratio	dimensionless spin magnitude	luminosity distance (Gpc)
M = 20M$_\odot$		
2	0.5	0.58
2	0.95	0.62
5	0.5	0.54
5	0.95	0.43
M = 80M$_\odot$		
2	0.5	1.95
2	0.95	2.15
5	0.5	1.57
5	0.95	0.83

timation is performed using BILBY [35] with the nested sampler DYNESTY [36], and the BILBY_TGR [37] plugin for the TIGER, FTI, and MDR tests. We use uniform priors for the component masses, spin magnitudes and non-GR parameters; isotropic priors for spin directions, binary orientation, and sky location; and a luminosity distance prior corresponding to a uniform merger rate in the source's comoving frame, using the same Planck cosmology [38] as in the LVK catalog analysis [34].

IV.   TESTS OF GR

The tests we conduct are based on waveform models for unlensed GW signals from quasicircular BBHs in GR. For these tests, we use the IMRPhenomXPHM [39] model for both injection and recovery in the TIGER, MDR, and IMR consistency tests, while in the FTI test, we use the SEOBNRv4PHM [40] model for injection and SEOBNRv4HM_ROM [41, 42] for recovery. We use the aligned-spin model SEOBNRv4HM_ROM since it forms the basis of the only higher-mode version of the FTI test currently implemented, and thus use the closest precessing waveform model for the injections, SEOBNRv4PHM.

IMRPhenomXPHM is a frequency-domain phenomenological model for quasicircular BBH waveforms that enhances the accuracy of the baseline IMRPhenomXAS model [43] for the dominant $(2, \pm 2)$ modes of aligned-spin waveforms by incorporating spin precession and the subdominant modes $(2, \pm 1)$, $(3, \pm 3)$, $(3, \pm 2)$, and $(4, \pm 4)$ in the coprecessing frame. SEOBNRv4HM_ROM is a frequency-domain reduced-order model for quasicircular BBH waveforms based on a time-domain aligned-spin effective-one-body (EOB) model, including the $(2, \pm 1)$, $(3, \pm 3)$, $(4, \pm 4)$, and $(5, \pm 5)$ subdominant modes. However, we do not consider the $(5, \pm 5)$ modes in our FTI analysis, since they do not make a significant contribution to the signal. To evaluate the impact of excluding the $(5, \pm 5)$ modes, we computed the SNR difference using the SEOBNRv4PHM waveform model and found the difference to be less than 0.2 and 0.3 for $M = 20M_\odot$ and $M = 80M_\odot$ binaries respectively. This difference is small enough to not

impact our results significantly.

A. TIGER and FTI

Parameterized tests of GR have been developed to detect potential deviations from GR by introducing modifications to the frequency-domain phase coefficients of GR waveform models. One such framework is the TIGER test, which introduces deviations in the GR phase coefficients of `IMRPhenomXAS`, a non-precessing dominant-mode BBH waveform model. This model is then extended to higher modes [44] following which the aligned-spin higher-mode model is twisted up to account for precession, giving the final modified `IMRPhenomXPHM` model used in our study. In the inspiral portion of the signal, the deviations introduced in the post-Newtonian (PN) coefficients also affect the higher modes (in the coprecessing frame), as illustrated below. However, deviations in the phenomenological intermediate and merger-ringdown coefficients only affect the dominant mode of the waveform (in the coprecessing frame). The end of the inspiral portion of the signal is set around the frequency of the minimum energy circular orbit and is different for different modes [44].

The deviations to the GR phase are introduced into the PN coefficients (φ_k and $\varphi_{k(l)}$) within the Fourier-domain inspiral phase of the waveform. The inspiral phase $\Phi_{\ell m}(f)$ of the (ℓ, m) spin-(-2)-weighted spherical harmonic mode in the coprecessing frame is given by

$$\Phi_{\ell m}(f) = \frac{3}{128 \eta v^5} \frac{m}{2} \sum_{n=0}^{7} \left(\varphi_k + 3\varphi_{k(l)} \log v \right) v^k \quad (4)$$

(excluding additive constants and phase and time shifts), where the factor of 3 in the logarithmic term arises from the definition of PN coefficients used in TIGER. Here, $\eta = m_1 m_2 / M^2$ is the symmetric mass ratio, with m_1 and m_2 representing the binary's individual (redshifted) masses, $M = m_1 + m_2$ is the total (redshifted) mass of the binary and t_c, ϕ_c are the time and phase at coalescence. The variable v is defined as $v = (\pi M f)^{1/3}$, where f is the GW frequency. The logarithmic coefficients φ_{kl} are non-zero only for $k \in \{5, 6\}$. We also consider -1PN deviations that are not present in GR, but would appear in the above expression with $k = -2$.

The TIGER deviations are introduced in both the PN coefficients of the inspiral phase and in the phenomenological coefficients b_k and c_k during the intermediate and merger-ringdown phases of the waveform, respectively. The intermediate parameters b_k have a frequency dependence $\propto f^k$ and the merger-ringdown parameters c_1, c_2, c_4 have the dependence $f^{-2/3}, f^1, f^3$ respectively and c_l is the coefficient of the antiderivative of the Lorentzian that models the ringdown. In the `IMRPhenomXAS` model, the phase is constructed to be C^1 continuous, such that any modification in the lower-frequency (e.g., inspiral) phase also affects the higher-frequency portions of the phase (the intermediate and merger-ringdown regions for inspiral deviations). All the deviations are parameterized by introducing a deviation parameter $\delta \hat{p}_k$ through the substitution $p_k \to (1 + \delta \hat{p}_k) p_k$, where p_k denotes any of the PN and phenomenological coefficients. An exception is made for $\delta \hat{\varphi}_{-2}$ and $\delta \hat{\varphi}_1$, which are zero in GR, and thus are normalized by just the 0PN coefficient. The normalization of the PN coefficients uses the coefficient without the spin contributions to avoid degeneracies since the coefficients can become zero when including the spin contributions. In GR, all deviation parameters are zero by definition.

In the standard TIGER test, only one PN parameter is varied at a time to avoid generating uninformative results, as is illustrated for GW150914 in [45]. However, deviations affecting multiple PN coefficients can still be detected by individually varying the parameters, even if the specific coefficients are not directly modified, as is shown in [26, 46]. Nevertheless, with multiband observations of BBHs with space-based and ground-based GW detectors, the degeneracies between parameters will be removed and it will be possible to constrain all PN coefficients simultaneously [47, 48]. And even for observations in a single band, one can also obtain the well-measured combinations of parameters using principal component analysis [49–52].

The FTI test [27] is similar to TIGER but focuses exclusively on deviations in PN coefficients and can be applied to any aligned-spin waveform model. However, the current implementation of the higher-mode version is limited to `SEOBNRv4HM_ROM`. Additionally, FTI tapers the deviations to zero beyond a certain frequency instead of allowing them to affect the rest of the signal. The same taper frequency used in the TGR analysis [53] of GW230529 is being used here.

B. Modified dispersion relation

The MDR test constrains dispersive GW propagation, where different wavelengths travel at different velocities, leading to frequency-dependent modifications to the phase of the observed GW waveform. Specifically, the MDR test introduces a phenomenological dispersion relation $E^2 = p^2 + A_\alpha p^\alpha$, where E and p represent the energy and momentum of the GWs, respectively, while A_α and α are phenomenological parameters that determine the strength of the deviation from GR and the frequency dependence of the dispersion, respectively, following [28]. For $\alpha = 0$ and $A_0 > 0$, this corresponds to the dispersion relation of a massive graviton.

Following the discussion in [54], we assume that the waveform near the source remains very close to that predicted by GR, so that the only modifications we consider are those arising from the dispersive propagation. This dispersion relation leads to an additional term proportional to $A_\alpha f^{\alpha-1}$ in the waveform's frequency-domain phase; the explicit group velocity expression is given in [24]. As in the previous LVK analyses (e.g., [55]), we consider $\alpha \in \{0, 0.5, 1, 1.5, 2.5, 3, 3.5, 4\}$, omitting $\alpha = 2$ since there is no dispersion in this case. In this study, we parameterize the phase corrections in terms of A_{eff}, defined as:

$$A_{\alpha,\text{eff}} = \frac{D_\alpha}{D_L} (1 + z)^{\alpha-1} A_\alpha, \quad (5)$$

where, D_L is the luminosity distance, z is the redshift and D_α is a distance parameter, with an explicit expression given in [54].

C. IMR consistency test

The IMR consistency test [29, 30] checks the consistency between the low- and high-frequency portions of the frequency-domain GW signal (roughly corresponding to the inspiral and postinspiral regimes) of a BBH. The division between these portions is made at a cut-off frequency f_c corresponding to the frequency of the innermost stable circular orbit (ISCO) of the final Kerr black hole, obtained from the GR analysis of the full signal [56]. We follow the procedure used in the LVK analysis and calculate the cut-off frequency using the medians of the individual masses and spins. The test evaluates the consistency of the two portions by comparing the (redshifted) final mass M_f and spin χ_f from each portion, yielding deviation parameters

$$\frac{\Delta M_f}{\bar{M}_f} := 2\frac{M_f^{\text{insp}} - M_f^{\text{postinsp}}}{M_f^{\text{insp}} + M_f^{\text{postinsp}}}, \quad \frac{\Delta \chi_f}{\bar{\chi}_f} := 2\frac{\chi_f^{\text{insp}} - \chi_f^{\text{postinsp}}}{\chi_f^{\text{insp}} + \chi_f^{\text{postinsp}}},$$

where the "insp" and "postinsp" superscripts correspond to the low- and high-frequency portions of the signal. These deviation parameters should both be zero if the signal is consistent with the waveform model used in the analysis, which is a quasi-circular BBH merger in GR for all current applications. As in [55, 57], we reweight to a flat prior in the deviation parameters to obtain the final results. We obtain the final mass and spin using the following procedure: We perform PE separately on each portion of the signal using a standard BBH waveform model (in this case, `IMRPhenomXPHM`), parameterized by the binary's initial masses and spins. We then compute the final mass and spin using an average of fits to numerical relativity (NR) simulations [58–60] (where the aligned-spin final spin fits are augmented by the contribution from in-plane spins [61], though we do not evolve the spins, following the LVK applications of the test).

V. RESULTS

We now discuss the results obtained when performing the four tests of GR described in Sec. IV on the simulated Type II lensed signals discussed in Sec. III. Sometimes, results are expressed in terms of GR quantiles, which represent the quantile at which the GR value of the test is recovered. For the IMR consistency test, these quantiles are two-dimensional, with the GR quantile indicating the fraction of the posterior distribution enclosed by the isoprobability contour passing through the GR value. Larger GR quantile values correspond to larger deviations from GR. For all other tests, the GR quantiles are one-dimensional, where extreme values (either large or small) suggest that the posterior distribution does not peak near the GR value.

A. TIGER

We give the posterior probability distributions (henceforth posterior distributions or posteriors) of the TIGER testing parameters for all our low and high mass injections in Figs. 1 and 2 respectively. We also summarize the statistical level at which GR is excluded in Figs. 3 and 4.

1. $20M_\odot$ injections

All lensed binaries with total mass $20M_\odot$ are consistent with GR at 90% credibility for all testing parameters. Additionally, we found that the posteriors of the GR parameters are almost identical to those obtained when analyzing these Type II lensed signals with an unlensed GR waveform model, peaking at or close to the injected values. Additionally, as a check of these results, we consider one unlensed case for each injection. Here we select the testing parameter that yields the largest GR quantile. The testing parameters selected for the injections with $q = 2$ are φ_7 and φ_{5l}, while for $q = 5$ they are φ_3 and φ_{6l}; for both mass ratios the two cases correspond to $\chi = 0.5$ and 0.95, respectively. We plot the unlensed results in Fig. 1, and notice that all their posteriors are in good agreement with the lensed ones. Thus, Type II lensing does not lead to large GR deviations with TIGER in the lower mass cases we considered.

2. $80M_\odot$ injections

The TIGER analysis of the lensed binaries with total mass $80M_\odot$ we consider gives consistency with GR at the 90% credible level in all cases except for the most extreme case of $q = 5$ and $\chi = 0.95$, where GR is even excluded at $> 2\sigma$ for φ_3 through φ_7.

We find that the cases where GR is excluded at $> 2\sigma$ also give the largest differences in the recovery of the GR parameters, compared to the GR analysis. In particular, the posteriors peak at values larger than the injected ones for the chirp masses and luminosity distances, and at smaller values for θ_1 placing them just outside the 90% credible intervals. Here the chirp mass $(m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$ is a combination of the component masses (m_1, m_2) which gives the leading effect of the masses on the GW signal, while θ_1 is the the zenith angle between the Newtonian orbital angular momentum and the primary spin. In the GR analysis of the high-spinning binaries, the injected chirp mass falls within the 90% credible interval. However, the injected values for luminosity distance, θ_1 and spin magnitudes lie just outside the 90% credible interval, with the injected luminosity distance being smaller, θ_1 and spin magnitudes being larger than the recovered values, respectively.

As for the $20M_\odot$ injections, we apply the TIGER test on the unlensed injections by selecting the testing parameter that yields the largest GR quantile for each injection and compare their posteriors with the lensed counterparts. The cases with the largest GR quantile selected for unlensed injections are c_l,
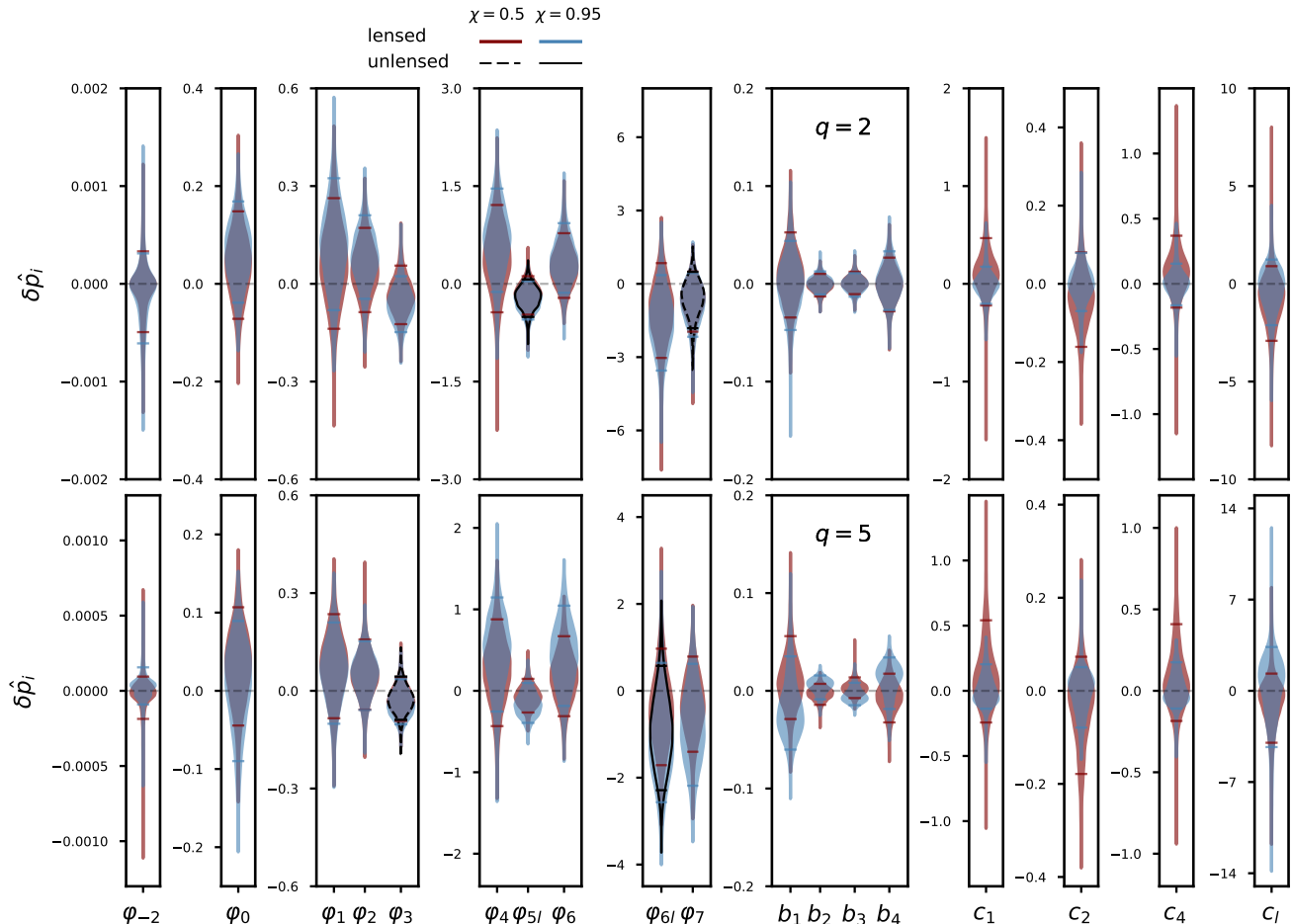
FIG. 1. The results of the TIGER test on the Type II lensed simulated $M = 20M_\odot$ injections, showing the results for mass ratios $q = 2$ and 5 in the top and bottom panels, respectively, and the results for dimensionless spins $\chi = 0.5$ and 0.95 with different colours. The posteriors of the testing parameters are presented as violin plots and the associated 90% credible intervals are shown using horizontal bars. We mark the GR value of zero with a dashed horizontal line. We also show the results for the unlensed injections we carried out for selected runs for each mass ratio and higher (lower) spin, as black solid (dashed) unfilled violins.

b_1 for $q = 2$, and c_1, φ_7 for $q = 5$, where for both mass ratios the two cases correspond to $\chi = 0.5$ and 0.95, respectively. We plot the unlensed results in Fig. 2 and observe that the lensed and unlensed posteriors are almost identical in all cases except for the injection with $q = 5$ and $\chi = 0.95$, where the lensed case shows a GR deviation of 2.8σ, while the unlensed posterior peaks at the GR value, as expected. Additionally, while we observe biases in the recovery of the GR parameters in the TIGER and GR analysis of the lensed cases, the corresponding unlensed injections recover the injected GR parameters around the median of the respective posteriors.

B. FTI

We give the posterior distributions of the FTI testing parameters for all our low and high mass injections in Figs. 5 and 6 respectively. We also summarize the statistical level at which

GR is excluded in Fig. 3.

1. $20M_\odot$ injections

Looking at Fig. 5, we observe significant deviations from GR when applying FTI to all the $M = 20M_\odot$ binaries except the one with $q = 2$, $\chi = 0.5$, which is consistent with GR at the 90% credible level for all testing parameters. While all cases with $q = 2$ and $\chi = 0.5$ are consistent with GR at the 90% credible level, all the $q = 2$, $\chi = 0.95$ cases exclude GR (again at the 90% credible level). GR is excluded at the 90% credible level for all other binaries and testing parameters, except for the parameters φ_{-2} and φ_7 for the $q = 5$, $\chi = 0.95$ binary.

In particular, the parameters φ_1 and φ_4 for $q = 5$, $\chi = 0.5$ show extreme deviations, with posteriors railing against the upper prior range of 60 we used for the deviation parameters.
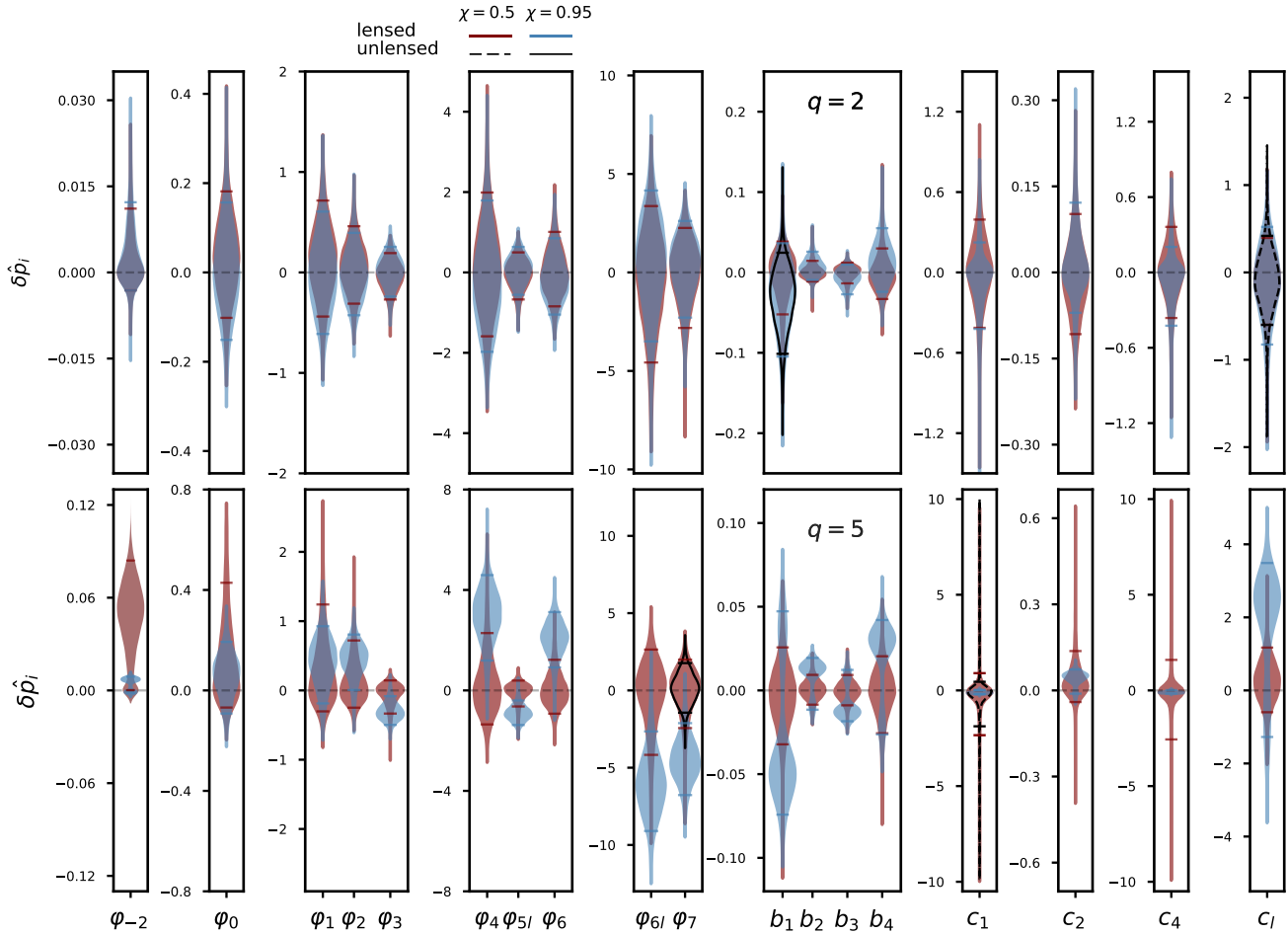
FIG. 2. The results of the TIGER test on the Type II lensed simulated $M = 80 M_\odot$ injections. The design of the plot is the same as Fig. 1.

We chose not to extend the prior range on these parameters further, as we find that all these large deviations are due to systematics due to the lack of precession in the current FTI implementation, rather than Type II lensing. For these testing parameters, we also notice a railing against the prior bounds for the chirp mass and luminosity distance which we similarly do not correct by extending the ranges for the same reason.

To confirm that these deviations are indeed not due to lensing, we performed FTI analysis on unlensed injections for the cases with the largest GR quantiles, as we did for TIGER, and also plot these in Fig. 5. The testing parameters chosen for the unlensed cases are φ_4, φ_{6l} for $q = 2$ and φ_{5l}, φ_7 for $q = 5$, where the two cases for each mass ratio correspond to $\chi = 0.5$ and 0.95, respectively. We find that the unlensed posteriors exhibit similar biases to their respective lensed counterparts, indicating that the observed biases are due to the absence of precession in the waveform currently used for the FTI test. Thus, we conclude that for the binaries considered here, precession can cause significant biases, but Type II lensing does not lead to a significant additional bias beyond that due to precession. In order to futher investigate the effects of precession, we apply the FTI test on an aligned-

spin (SEOBNRv4HM_ROM) injection for the binary with $q = 2$ and $\chi = 0.95$ and testing parameter φ_{5l} since this gives us the highest GR quantile. We find that the posterior exactly peaks at the GR value of zero, thus verifying that the biases observed were indeed due to precession.

When we plot the FTI significances in Fig. 3, we find that most $q = 5$ binaries exclude GR at such high credible levels that the estimation of the GR quantile becomes unreliable with the order of 10^4 posterior samples we obtain. Thus, as in [10], we conservatively report a lower bound of 3σ on the significance of the deviations, since that study found that larger significances are not accurately obtained with this number of posterior samples.

The binary with $q = 2$ and $\chi = 0.5$ is consistent with GR at $< 1.4\sigma$ across all testing parameters. However, all other binaries exclude GR at $> 2\sigma$, except for the parameters φ_{-2} and φ_7 in the $q = 5$, $\chi = 0.95$ configuration, which are consistent with GR at well below 1σ. Notably, all testing parameters for the $q = 5$, $\chi = 0.5$ case, and all but the aforementioned two for the $q = 5$, $\chi = 0.95$, show deviations from GR exceeding 3σ.

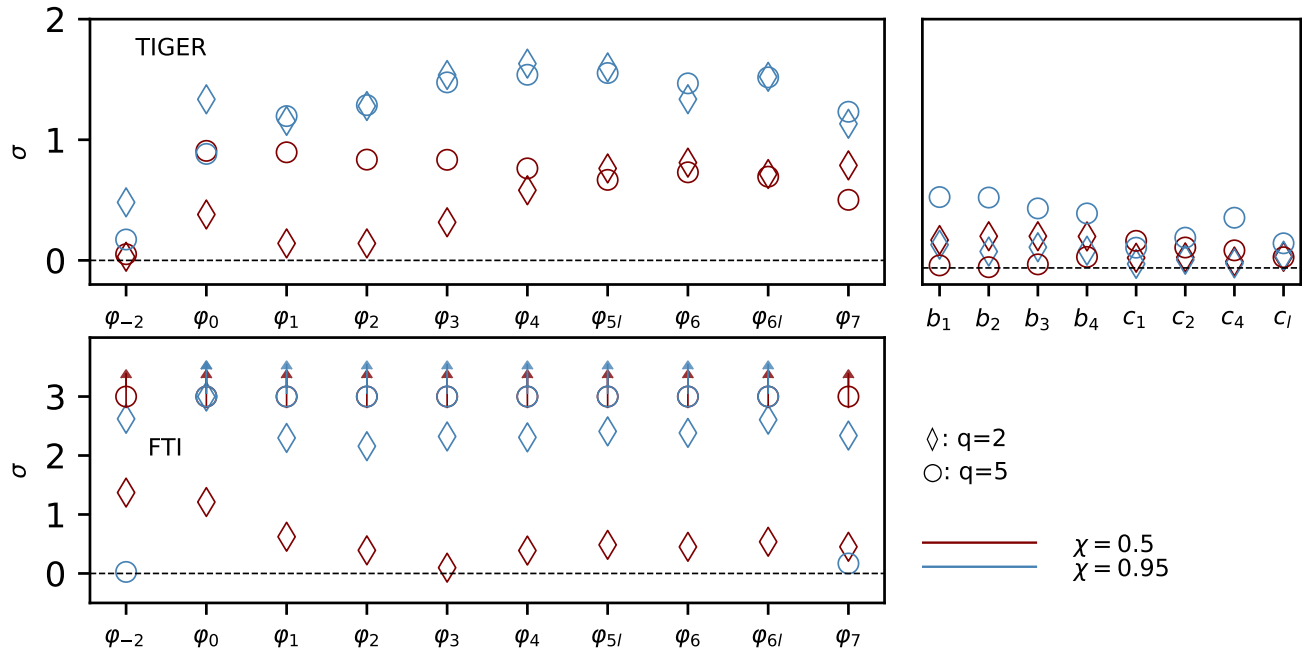For the $q = 2$ injections, we recover smaller values for the

FIG. 3. The Gaussian sigma value at which GR is excluded when applying the TIGER and FTI tests to the Type II lensed injections with $M = 20M_\odot$.

primary spin magnitudes. More pronounced biases are observed in the $q = 5$ injections, particularly in the recovered mass ratios. Testing parameters φ_{5l}, φ_3, φ_2 recover more asymmetric mass ratios while the others recover more symmetric values, with an exception of φ_7 where the injected value is close to the median of the recovered posterior. Additionally, these higher mass ratio injections yield significantly larger chirp masses, total masses, and luminosity distances than the injected values. The FTI analysis of the $q = 5$, $\chi = 0.95$ injection also recovers lower values for the component spin magnitudes.

2. $80M_\odot$ injections

FTI finds that all lensed binaries with total mass $80M_\odot$ are consistent with GR at 90% credibility for all testing parameters, though GR is excluded at up to 1.5σ in some cases. In particular, as illustrated in Fig. 4, GR is excluded at $\gtrsim 1\sigma$ for all testing parameters of the $q = 2$, $\chi = 0.95$ binary and for all except φ_6 and φ_7 for the $q = 5$, $\chi = 0.5$ binary (which gives the 1.5σ exclusion for φ_0). The cases where GR is recovered in the tail of the distribution are likely due to the lack of precession in the current implementation of FTI, similar to the much larger biases found for most of the $M = 20M_\odot$ injections. The lack of such significant bias in the $80M_\odot$ is presumably due to the shorter signals in the sensitive band of the detectors making the effects of precession on the signal less important.

When applying FTI to binaries with $\chi = 0.95$ for both mass ratios $q = 2, 5$, we observe that the posteriors peak at more symmetric mass ratios and smaller spin magnitudes, and the posteriors for the binary with $\chi = 0.95$ and $q = 5$ peak at larger luminosity distances. Additionally, for $q = 2$, $\chi = 0.95$, binary, we find a tendency for lower recovered chirp masses for testing parameters $\varphi_{-2}, \varphi_0, \varphi_1, \varphi_2$.

In order to study the effects of systematics, we analyze unlensed injections for the testing parameters that give the largest GR quantiles and make a comparison with the corresponding lensed cases, shown in Fig. 6. Specifically, we consider φ_{-2}, φ_{6l} for $q = 2$ and φ_0, φ_6 for $q = 5$, where the two values for each mass ratio correspond to $\chi = 0.5$ and 0.95, respectively. We observe that all the unlensed posteriors overlap significantly with the lensed posteriors, indicating that there are no significant additional biases due to Type II lensing in these cases beyond the ones from precession.

C. MDR

For MDR, we report the results of the test in Fig. 7 as Bayes factors B_{GR}^{MDR} comparing the MDR and GR models. The Bayes factor is the ratio of the evidences (i.e., the marginal likelihoods) of the two models. Larger values of $\log_{10} B_{GR}^{MDR}$ correspond to more support for MDR, while negative values correspond to support for GR. We quote Bayes factors instead of plotting the posteriors and quoting the GR quantiles because we find that the posteriors for A_{eff} has multiple modes with no support at the GR value in the cases where GR is excluded at a high credible level. Thus, it would be misleading to quantify the statistical level at which GR is excluded from such a posterior using quantiles. An example of such a multimodal
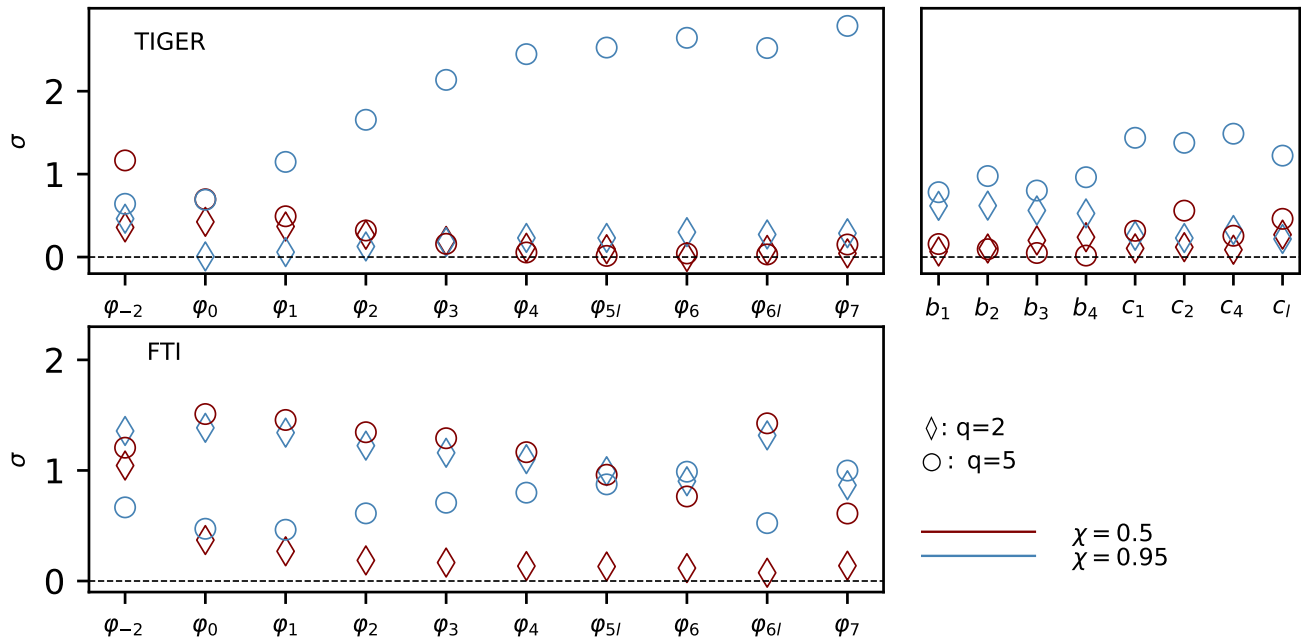
FIG. 4. The Gaussian sigma value at which GR is excluded when applying the TIGER and FTI tests to the Type II lensed injections with $M = 80 M_\odot$.

posterior is given in Fig. 8 for the binary with $M = 80 M_\odot$, $q = 5$, and $\chi = 0.95$ analyzed with $\alpha = 1$.

The posterior of A_{eff} is bimodal for all α values except $\alpha = 1$ with an infinite $\pm\pi$ degeneracy around the Type II lensing phase shift value of $\pi/2$. Infinite values of A_{eff} reproduce the modified GW signal by frequency independent phases $(k + 1/2)\pi$ for $k = (0, 1, 2, ...)$. We observe this from the Fig. 8 where, the first mode of the A_{eff} posterior at negative values, corresponding to a phase shift of $\pi/2$, peaks close to the injected value of -2.4×10^{-20} eV^2. We also find that the two α values with the next largest support for MDR for a given binary, at least for the $q = 5$, $\chi = 0.95$ cases, are 0.5 and 1.5, as would be expected, by continuity, and the support decreases as α becomes further from 1.

For the $80 M_\odot$ injections, the high spin binaries recover biased individual spin magnitudes with both GR and MDR analysis recovering lower values than those injected.

D. IMR Consistency Test

We perform the IMR consistency test only on the higher mass injection since the LVK only applies the test to signals with an SNR of at least 6 in both the inspiral and postinspiral portions, since both need to be informative for the test to make sense (see [54, 55, 57]). The cutoff frequencies demarcating the inspiral and postinspiral phases for $M = 20 M_\odot$ are 415, 418 Hz and 330, 399 Hz for the $q = 2$ and $q = 5$ binaries, respectively, where the two numbers for each mass ratio give the results for $\chi = 0.5, 0.95$. With these cutoffs, the largest SNR in the postinspiral regime is 5.76 for the $q = 2$, $\chi = 0.95$ case,

just below the threshold. For the $M = 80 M_\odot$ binaries, the cutoff frequencies are 104, 108 Hz and 80, 110 Hz for $q = 2$ and $q = 5$, respectively, where again the two values correspond to $\chi = 0.5, 0.95$. The SNR in the post-inspiral regime range from the $80 M_\odot$ binaries under consideration range from ~ 10 to ~ 15.

In Fig. 9, we show our results for the IMR consistency test on both lensed and unlensed injections. We find that the $q = 2$ lensed (unlensed) cases are consistent with GR (GR quantiles of $\sim 17\% (\sim 19\%)$, $\sim 10\% (\sim 13\%)$ for $\chi = 0.5$ and 0.95, respectively). The $q = 5$ cases show notable inconsistency (GR quantiles of $\sim 93\%$, $\sim 95\%$ for $\chi = 0.5$ and 0.95), even without lensing (GR quantiles of $\sim 94\%$, $\sim 92\%$).

The waveform model used for injection and recovery in the IMR consistency test is `IMRPhenomXPHM` which includes higher order modes. Since we found a significant bias (GR quantiles of $> 95\%$) in the IMR consistency test for (quasi-circular) unequal-mass nonspinning binaries observed face-on in [10], we thus suspect that the bias observed for the $q = 5$ unlensed injections here could also be due to the presence of higher modes, even though we are considering an inclination angle of $\pi/3$, while ongoing studies [62] have found that these biases are only significant for binaries very close to face-on (or face-off). We thus applied the IMR consistency test to the unlensed $q = 5$, $\chi = 0.95$ injection including only the $(2, \pm 2)$ modes in the coprecessing frame (i.e., using the `IMRPhenomXP` waveform model) for both the injection and recovery. We use the same cutoff frequencies as in the `IMRPhenomXPHM` analysis. We indeed find that the unlensed injection agrees with GR in this case (GR quantile of 38%). Similar results were obtained in [10] wherein GR deviations
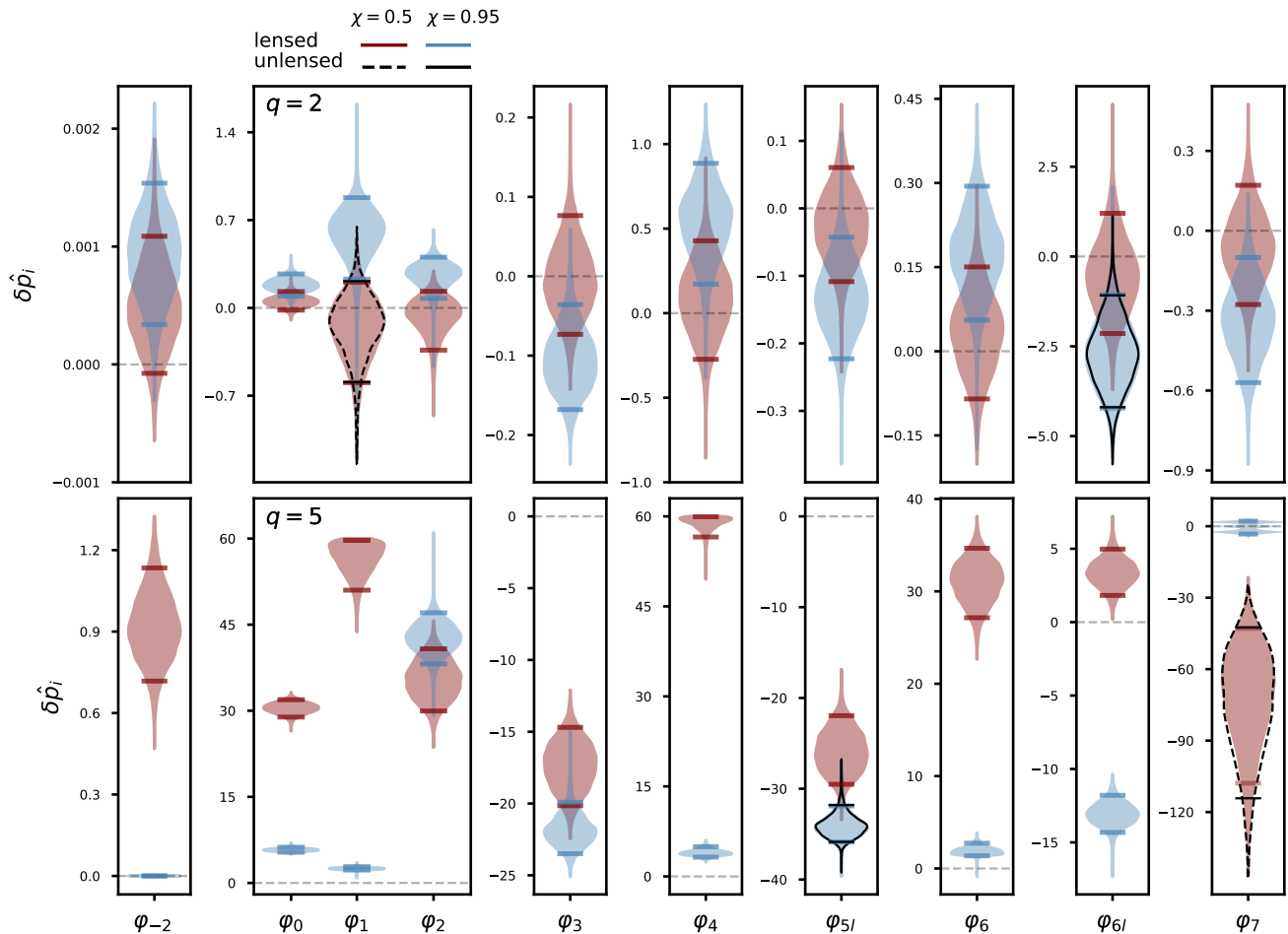
FIG. 5. Violin plots for FTI testing parameters for a total mass of $20M_\odot$. The color scheme and the layout of the subplots are similar to Fig. 1.

were found when a face-on quasi-circular NR injection was recovered using the quasi-circular `IMRPhenomXPHM` model with GR quantiles $> 95\%$.

VI. SUMMARY AND CONCLUSIONS

In thus study, we investigate the impact of neglecting Type II strong gravitational lensing when performing some of the LVK's standard GR tests on lensed BBH signals. Specifically, we examine the response of the TIGER, FTI, MDR, and IMR consistency tests to simulated Type II lensed BBH GW signals within the LIGO-Virgo network, assuming O5 design sensitivity. We simulate Type II lensed signals for the precessing binaries under consideration. These binaries have redshifted total masses of $20M_\odot$ and $80M_\odot$, each with mass ratios of 2 and 5. For each mass ratio, we consider equal dimensionless spins (effective precession spin parameter, χ_p) of 0.5 (0.45) and 0.95 (0.85), a total of eight cases, all with the same random spin angles that will lead to significant precession. Each binary is oriented to have an inclination angle of $\pi/3$ and

placed at a luminosity distance so that it has a network SNR of 25 with the magnification of 1 that we consider. For each parameter combination, we analyze an unlensed BBH signal for comparison with the corresponding lensed case. This is done for a single testing parameter for TIGER/FTI and a specific α value for MDR. All the injections are made in zero noise.

For the TIGER test, all lensed binaries with $M = 20M_\odot$ were consistent with GR at the 90% credible level. However, for higher mass binaries ($M = 80M_\odot$), significant deviations were found for the configuration with $q = 5$ and $\chi = 0.95$, where GR was excluded at $> 2\sigma$ for some testing parameters. For the FTI test, very significant deviations from GR were found for binaries with $M = 20M_\odot$ and $q = 5$. However, these deviations are also found for the unlensed injections and are thus due to systematics from analyzing strongly precessing injections with the current FTI analysis that does not include precession. All binaries with $M = 80M_\odot$ are consistent with GR at 90% credibility for all FTI testing paremeters. For the MDR test, we find Bayes factors favoring the MDR model over GR for the $q = 5$, $\chi = 0.95$ binaries, with MDR being favored more strongly for $M = 80M_\odot$ binary. The
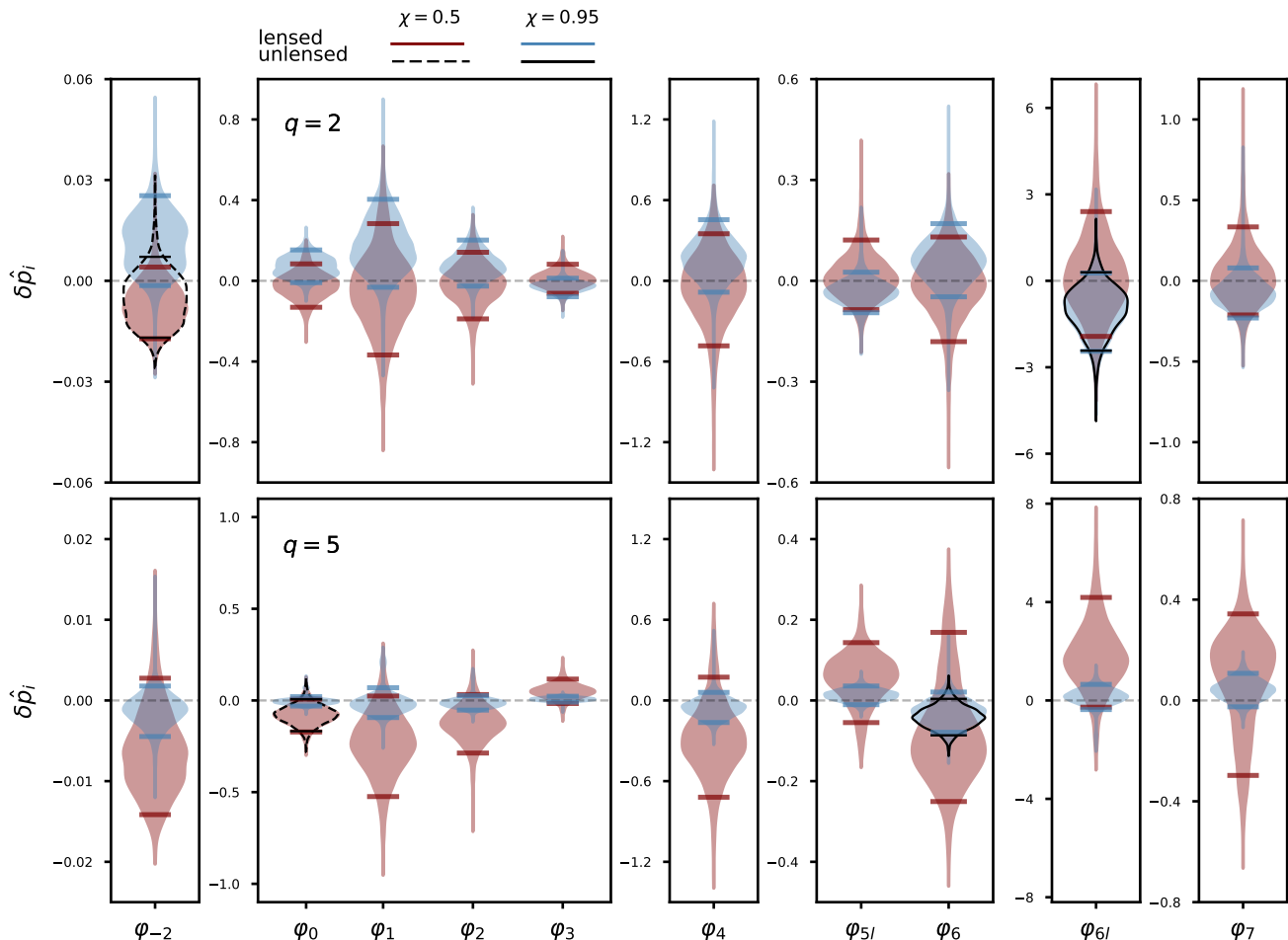
FIG. 6. Violin plots for FTI testing parameters for a total mass of $80M_\odot$. The color scheme and the layout of the subplots are similar to Fig. 1.

strongest support for MDR was for $\alpha = 1$, as expected, since this is where the MDR phase shift is able to reproduce the Type II lensing phase shift exactly. However, GR is favored slightly for some of the $q = 2$ and/or $\chi = 0.5$ cases. Finally for IMR consistency test, we only analyze the binaries with $M = 80M_\odot$, since the $20M_\odot$ binaries do not have sufficiently high postinspiral SNR. Here we observe that the $q = 2$ cases (with spins 0.5 and 0.95) are consistent with GR. However, the $q = 5$ cases exhibit significant inconsistencies, with GR quantiles as large as 95%. Furthermore, we also find these biases for the unlensed $q = 5$ injections, and find that they can be attributed to the influence of higher-order modes, as the quantiles reduce to $\sim 40\%$ when using a waveform model limited to the $(2, \pm2)$ modes in the coprecessing frame for both injection and recovery. These biases are likely related to similar biases found in [10] when including higher-order modes in the analysis of face-on (quasicircular) nonspinning unequal-mass binaries.

The results of this paper highlight the necessity of ruling out strongly lensed binaries as a source of false GR violations before claiming a potential deviation from GR. Specifically,

one must consider the potential for a Type II lensed signal to mimic a false deviation from GR since the LVK lensing analyses are generally expected to detect the dephasing characteristic of Type II signals and possibly identify companion lensed images. Addressing this requires incorporating the Type II lensing effect into baseline GR waveforms used in GR tests. A natural extension of this study is to evaluate the response of Type II lensing analyses to the signals considered here and confirm that including the lensing effect in baseline GR waveforms resolves the biases.
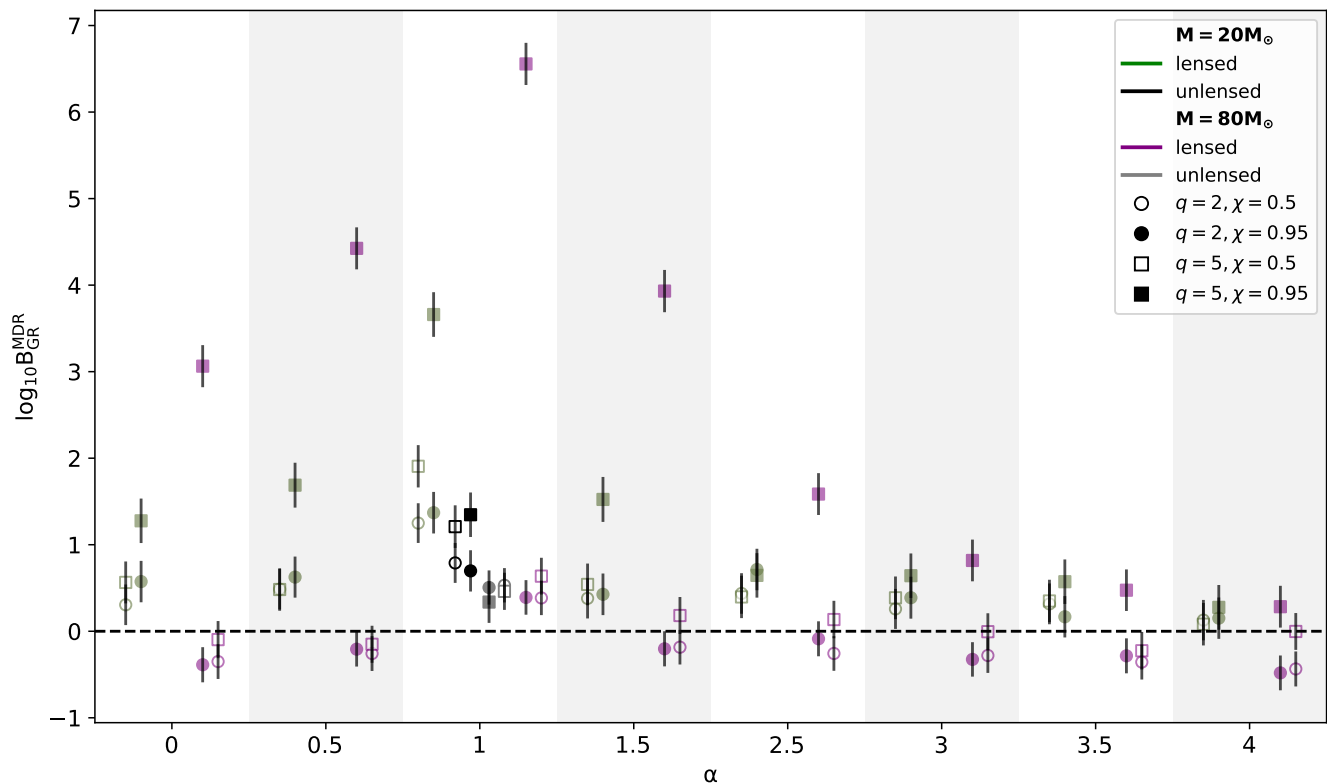
FIG. 7. Plot of Bayes factors comparing MDR and GR for the set of eight discrete values of the modified dispersion relation exponent α we consider, excluding $\alpha = 2$ where there is no dispersion. The horizontal positioning of the points for each α value is just done for clarity.

[1] C. M. Will, Living Rev. Relativity **17**, 4 (2014), arXiv:1403.7377 [gr-qc].
[2] N. Wex, "Testing Relativistic Gravity with Radio Pulsars," in *Frontiers in Relativistic Celestial Mechanics, Volume 2: Applications and Experiments*, edited by S. M. Kopeikin (De Gruyter, Berlin, 2014) arXiv:1402.5594 [gr-qc].
[3] J. M. Weisberg and J. H. Taylor, ASP Conf. Ser. **328**, 25 (2005), arXiv:astro-ph/0407149.
[4] G. Voisin, I. Cognard, P. C. C. Freire, N. Wex, L. Guillemot, G. Desvignes, M. Kramer, and G. Theureau, Astron. Astrophys. **638**, A24 (2020), arXiv:2005.01388 [gr-qc].
[5] M. Kramer *et al.*, Phys. Rev. X **11**, 041050 (2021), arXiv:2112.06795 [astro-ph.HE].
[6] L. Dai, T. Venumadhav, and K. Sigurdson, Phys. Rev. D **95**, 044011 (2017), arXiv:1605.09398 [astro-ph.CO].
[7] K. K. Y. Ng, K. W. K. Wong, T. Broadhurst, and T. G. F. Li, Phys. Rev. D **97**, 023012 (2018), arXiv:1703.06319 [astro-ph.CO].
[8] P. Saini, M. Favata, and K. G. Arun, Phys. Rev. D **106**, 084031

(2022), arXiv:2203.04634 [gr-qc].
[9] S. A. Bhat, P. Saini, M. Favata, and K. G. Arun, Phys. Rev. D **107**, 024009 (2023), arXiv:2207.13761 [gr-qc].
[10] P. Narayan, N. K. Johnson-McDaniel, and A. Gupta, Phys. Rev. D **108**, 064003 (2023), arXiv:2306.04068 [gr-qc].
[11] P. Saini, S. A. Bhat, M. Favata, and K. G. Arun, Phys. Rev. D **109**, 084056 (2024), arXiv:2311.08033 [gr-qc].
[12] M. A. Shaikh, S. A. Bhat, and S. J. Kapadia, Phys. Rev. D **110**, 024030 (2024), arXiv:2402.15110 [gr-qc].
[13] O. A. Hannuksela, K. Haris, K. K. Y. Ng, S. Kumar, A. K. Mehta, D. Keitel, T. G. F. Li, and P. Ajith, Astrophys. J. Lett. **874**, L2 (2019), arXiv:1901.02674 [gr-qc].
[14] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Astrophys. J. **923**, 14 (2021), arXiv:2105.06384 [gr-qc].
[15] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), Astrophys. J. **970**, 191 (2024), arXiv:2304.08393 [gr-qc].
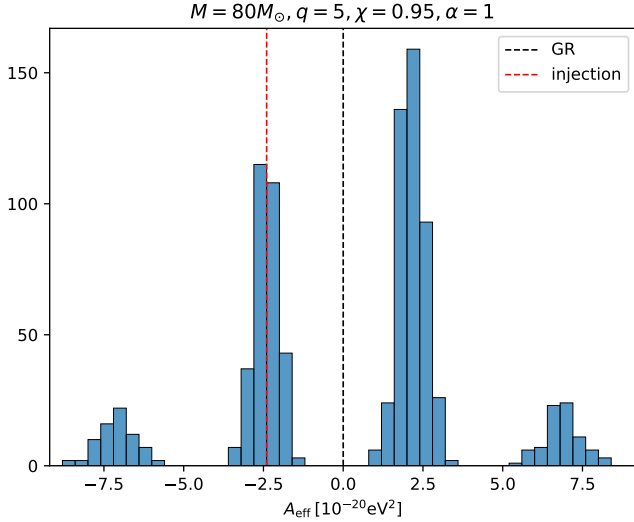[16] A. R. A. C. Wierda, E. Wempe, O. A. Hannuksela, L. V. E.

FIG. 8. A_{eff} posterior for $\alpha = 1$, $M = 80 M_\odot$, $q = 5$ and $\chi = 0.95$ case. We also mark the GR value of zero and the injected value due to Type II lensing phase shift of $\pi/2$. Here we have zoomed in around the GR value of zero where there is the largest posterior probability, but there are additional small peaks outside of this region.
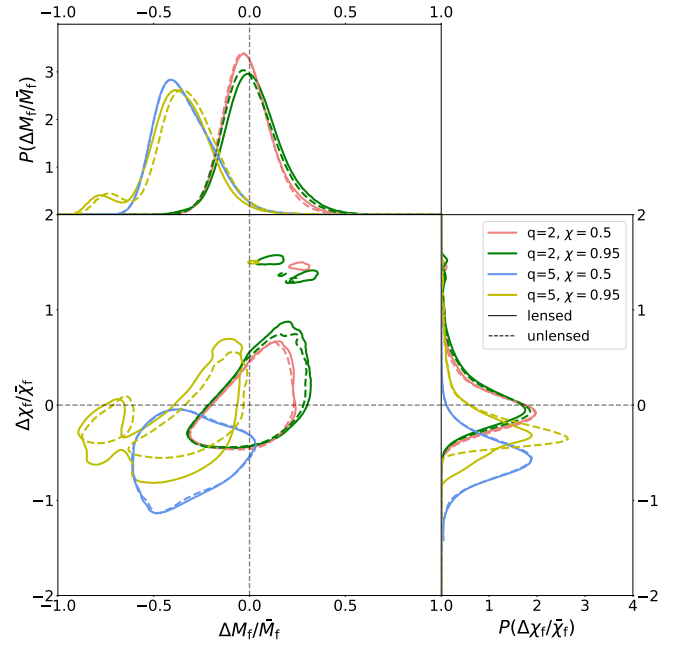

FIG. 9. The results from the IMR consistency test presented as the 90% credible regions of the joint posterior distributions of the recovered final mass and spin deviation parameters for lensed and unlensed injections for mass ratios 2 and 5 with dimensionless spins of 0.5 and 0.95. We also show the one-dimensional distributions for the marginalized deviation parameters.

Koopmans, and C. Van Den Broeck, Astrophys. J. **921**, 154 (2021), arXiv:2106.06303 [astro-ph.HE].

[17] D. Reitze *et al.*, Bull. Am. Astron. Soc. **51**, (7), 35 (2019), arXiv:1907.04833 [astro-ph.IM].

[18] M. Punturo *et al.*, Class. Quant. Grav. **27**, 194002 (2010).

[19] Y. Wang, R. K. L. Lo, A. K. Y. Li, and Y. Chen, Phys. Rev. D **103**, 104055 (2021), arXiv:2101.08264 [gr-qc].

[20] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses*, Astronomy and Astrophysics Library (Springer, 1992).

[21] J. M. Ezquiaga, D. E. Holz, W. Hu, M. Lagos, and R. M. Wald, Phys. Rev. D **103**, 064047 (2021), arXiv:2008.12814 [gr-qc].

[22] A. Vijaykumar, A. K. Mehta, and A. Ganguly, Phys. Rev. D **108**, 043036 (2023), arXiv:2202.06334 [gr-qc].

[23] M. Wright, J. Janquart, and N. K. Johnson-McDaniel, (2024), arXiv:2403.08957 [gr-qc].

[24] J. M. Ezquiaga, W. Hu, M. Lagos, M.-X. Lin, and F. Xu, J. Cosmol. Astropart. Phys. **08**, 016 (2022), arXiv:2203.13252 [gr-qc].

[25] M. Agathos, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, J. Veitch, and S. Vitale, Phys. Rev. D **89**, 082001 (2014), arXiv:1311.0420 [gr-qc].

[26] J. Meidam *et al.*, Phys. Rev. D **97**, 044033 (2018), arXiv:1712.08772 [gr-qc].

[27] A. K. Mehta, A. Buonanno, R. Cotesta, A. Ghosh, N. Sennett, and J. Steinhoff, Phys. Rev. D **107**, 044020 (2023), arXiv:2203.13937 [gr-qc].

[28] S. Mirshekari, N. Yunes, and C. M. Will, Phys. Rev. D **85**, 024041 (2012), arXiv:1110.2720 [gr-qc].

[29] A. Ghosh, A. Ghosh, N. K. Johnson-McDaniel, C. K. Mishra, P. Ajith, W. Del Pozzo, D. A. Nichols, Y. Chen, A. B. Nielsen, C. P. L. Berry, and L. London, Phys. Rev. D **94**, 021101(R) (2016), arXiv:1602.02453 [gr-qc].

[30] 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), arXiv:1704.06784 [gr-qc].

[31] B. P. Abbott *et al.* (KAGRA Collaboration, LIGO Scientific Collaboration, and Virgo Collaboration), Living Rev. Relativity **23**, 3 (2020), noise curves available from https://dcc.ligo.org/LIGO-T2000012/public, arXiv:1304.0670 [gr-qc].

[32] T. Akutsu *et al.* (KAGRA Collaboration), Nature Astron. **3**, 35 (2019), arXiv:1811.08079 [gr-qc].

[33] Gravitational wave detector observing timeline, https://dcc.ligo.org/G2002127-v26/public.

[34] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), Phys. Rev. X **13**, 041039 (2023), arXiv:2111.03606 [gr-qc].

[35] G. Ashton *et al.*, Astrophys. J. Suppl. **241**, 27 (2019), arXiv:1811.02042 [astro-ph.IM].

[36] J. S. Speagle, Mon. Not. R. Astron. Soc. **493**, 3132 (2020).

[37] G. Ashton, C. Talbot, S. Roy, G. Pratten, T.-H. Pang, M. Agathos, T. Baka, E. Sänger, A. Mehta, J. Steinhoff, E. Maggio, A. Ghosh, and A. Vijaykumar, "Bilby TGR," (2024).

[38] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **594**, A13 (2016), arXiv:1502.01589 [astro-ph.CO].

[39] G. Pratten *et al.*, Phys. Rev. D **103**, 104056 (2021), arXiv:2004.06503 [gr-qc].

[40] S. Ossokine, A. Buonanno, S. Marsat, R. Cotesta, S. Babak, T. Dietrich, R. Haas, I. Hinder, H. P. Pfeiffer, M. Pürrer, C. J. Woodford, M. Boyle, L. E. Kidder, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **102**, 044055 (2020), arXiv:2004.09442 [gr-qc].

[41] R. Cotesta, A. Buonanno, A. Bohé, A. Taracchini, I. Hinder, and S. Ossokine, Phys. Rev. D **98**, 084028 (2018), arXiv:1803.10701 [gr-qc].

[42] R. Cotesta, S. Marsat, and M. Pürrer, Phys. Rev. D **101**, 124040 (2020), arXiv:2003.12079 [gr-qc].

[43] G. Pratten, S. Husa, C. Garcia-Quiros, M. Colleoni, A. Ramos-Buades, H. Estelles, and R. Jaume, Phys. Rev. D **102**, 064001 (2020), arXiv:2001.11412 [gr-qc].

[44] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, Phys. Rev. D **102**, 064002 (2020), arXiv:2001.10914 [gr-qc].

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

[46] N. K. Johnson-McDaniel, A. Ghosh, S. Ghonge, M. Saleem, N. V. Krishnendu, and J. A. Clark, Phys. Rev. D **105**, 044020 (2022), arXiv:2109.06988 [gr-qc].

[47] A. Gupta, S. Datta, S. Kastha, S. Borhanian, K. G. Arun, and B. S. Sathyaprakash, Phys. Rev. Lett. **125**, 201101 (2020), arXiv:2005.09607 [gr-qc].

[48] S. Datta, A. Gupta, S. Kastha, K. G. Arun, and B. S. Sathyaprakash, Phys. Rev. D **103**, 024036 (2021), arXiv:2006.12137 [gr-qc].

[49] A. A. Shoom, P. K. Gupta, B. Krishnan, A. B. Nielsen, and C. D. Capano, Gen. Relativ. Gravit. **55**, 55 (2023), arXiv:2105.02191 [gr-qc].

[50] M. Saleem, S. Datta, K. G. Arun, and B. S. Sathyaprakash, Phys. Rev. D **105**, 084062 (2022), arXiv:2110.10147 [gr-qc].

[51] S. Datta, M. Saleem, K. G. Arun, and B. S. Sathyaprakash, Phys. Rev. D **109**, 044036 (2024), arXiv:2208.07757 [gr-qc].

[52] S. Datta, (2023), arXiv:2303.04399 [gr-qc].

[53] E. M. Sänger *et al.*, (2024), arXiv:2406.03568 [gr-qc].

[54] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. D **100**, 104036 (2019), arXiv:1903.04467 [gr-qc].

[55] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), (2021), arXiv:2112.06861 [gr-qc].

[56] J. M. Bardeen, W. H. Press, and S. A. Teukolsky, Astrophys. J. **178**, 347 (1972).

[57] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. D **103**, 122002 (2021), arXiv:2010.14529 [gr-qc].

[58] F. Hofmann, E. Barausse, and L. Rezzolla, Astrophys. J. Lett. **825**, L19 (2016), arXiv:1605.01938 [gr-qc].

[59] J. Healy and C. O. Lousto, Phys. Rev. D **95**, 024037 (2017), arXiv:1610.09713 [gr-qc].

[60] X. Jiménez-Forteza, D. Keitel, S. Husa, M. Hannam, S. Khan, and M. Pürrer, Phys. Rev. D **95**, 064024 (2017), arXiv:1611.00332 [gr-qc].

[61] N. K. Johnson-McDaniel, A. Gupta, P. Ajith, D. Keitel, O. Birnholtz, F. Ohme, and S. Husa, *Determining the final spin of a binary black hole system including in-plane spins: Method and checks of accuracy*, Tech. Rep. LIGO-T1600168 (LIGO Project, 2016) https://dcc.ligo.org/LIGO-T1600168/public/main.

[62] M. K. Singh (private communication).

[63] D. Williams, J. Veitch, M. L. Chiofalo, P. Schmidt, R. P. Udall, A. Vajpeji, and C. Hoy, J. Open Source Softw. **8**, 4170 (2023), arXiv:2207.01468 [gr-qc].

[64] I. M. Romero-Shaw *et al.*, Mon. Not. R. Astron. Soc. **499**, 3295 (2020), arXiv:2006.00714 [astro-ph.IM].

[65] R. K. L. Lo and I. Magaña Hernandez, Phys. Rev. D **107**, 123015 (2023), arXiv:2104.09339 [gr-qc].

[66] LVK Algorithm Library Suite (LALSuite), https://doi.org/10.7935/GT1W-FZ16.

[67] J. D. Hunter, Comput. Sci. Eng. **9**, 90 (2007).

[68] C. R. Harris *et al.*, Nature (London) **585**, 357 (2020), arXiv:2006.10256 [cs.MS].

[69] C. Hoy and V. Raymond, SoftwareX **15**, 100765 (2021), arXiv:2006.06639 [astro-ph.IM].

[70] L. London and E. Fauchon-Jones, Classical Quantum Gravity **36**, 235015 (2019), arXiv:1810.03550 [gr-qc].

[71] P. Virtanen *et al.*, Nat. Methods **17**, 261 (2020), arXiv:1907.10121 [cs.MS].

[72] M. Waskom, J. Open Source Softw. **6**, 3021 (2021).