

Improved parametrized test of general relativity using the IMRPHENOMX waveform family: Including higher harmonics and precession

Soumen Roy^{1,2,3,4,*} Maria Haney³ Geraint Pratten⁵ Peter T. H. Pang^{3,4} and Chris Van Den Broeck^{3,4}

¹*Centre for Cosmology, Particle Physics and Phenomenology—CP3, Université Catholique de Louvain, Louvain-La-Neuve, B-1348, Belgium*

²*Royal Observatory of Belgium, Avenue Circulaire, 3, 1180 Uccle, Belgium*

³*Nikhef—National Institute for Subatomic Physics, Science Park 105, 1098 XG Amsterdam, The Netherlands*

⁴*Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University, Princetonplein 1, 3584 CC Utrecht, The Netherlands*

⁵*School of Physics and Astronomy and Institute for Gravitational Wave Astronomy, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom*



(Received 4 May 2025; accepted 1 December 2025; published 8 January 2026)

When testing general relativity (GR) with gravitational wave observations, parametrized tests of deviations from the expected strong-field source dynamics are one of the most widely used techniques. We present an updated version of the parametrized framework with the state-of-art IMRPHENOMX waveform family. Our new framework incorporates deviations in the dominant mode as well as in the higher-order modes of the waveform. We demonstrate that the missing physics of either higher-order modes or precession in the parametrized model can lead to a biased conclusion of false deviation from GR. Our new implementation mitigates this issue and enables us to perform the tests for highly asymmetric and precessing binaries without being subject to systematic biases due to missing physics. Finally, we apply the improved test to analyze events observed during the second half of the third observing run of LIGO and Virgo (O3b). We provide constraints on GR deviations by combining O3b results with those from previous observation runs. Our findings show no evidence for violations of GR.

DOI: [10.1103/855k-sys5](https://doi.org/10.1103/855k-sys5)

I. INTRODUCTION

The detections of gravitational waves (GWs) from compact binary coalescences by Advanced LIGO [1] and Advanced Virgo [2] have opened up a new laboratory for testing general relativity (GR). This includes exploring the two-body dynamics in strong gravitational fields with velocities approaching the speed of light, the nature of compact objects, the properties of GW propagation, and the presence of extra dimensions. In this context, one of the most popular techniques are parametrized tests of deviations from the general-relativistic source dynamics of compact binary mergers.

The GW signals from inspiralling compact binaries are modeled using post-Newtonian (PN) expansion, expressed as a series expansion in the orbital velocity. The postinspiral part of the waveform is modeled using a set of phenomenological coefficients, which are determined by calibrating with numerical relativity and black hole perturbation theory for the ringdown part. Developing an accurate GW waveform model that describes the entire evolution of compact binaries

is essential for determining source properties, validating the predictions of GR, and exploring the potential alternative theories of gravity. However, the development of waveforms within specific alternative theories has not yet reached a level of maturity sufficient to perform model comparisons with GR. While numerical-relativity simulations in beyond-GR theories are becoming more prevalent [3–13], there are not yet sufficient simulations in one specific theory to calibrate a semianalytic inspiral-merger-ringdown model in that theory. Moreover, the true theory of gravity may be an alternative to GR that we are not yet theoretically aware of. This has led to the development of theory-agnostic approaches, which are classified into two categories: either verifying the consistency between the prediction of GR and data or introducing a deviation parameter to quantify the degree to which the GR prediction agrees with observed data.

Alternative theories of gravity often suggest deviation in the numerical values of the coefficients compared to those derived from Einstein's GR. This has prompted the development of theory-agnostic inspiral parametrized tests of GR [14–21], which look for the departure from GR by introducing parametric deformations to the PN phasing coefficients. The same framework can be used in search

*Contact author: soumen.roy@uclouvain.be

pipelines to detect non-GR signals [22,23]. In the inspiral parametrized test, we introduce a deviation parameter (also referred to as the non-GR or beyond-GR parameter) to the PN phasing coefficients, allowing for a generic departure from GR. A similar approach can be applied to the post-inspiral coefficients defined in phenomenological inspiral-merger-ringdown (IMR) waveform models [21]. For each coefficient, we measure that deviation parameter by performing Bayesian parameter estimation analysis. If the resulting posterior of the deviation parameter is consistent with zero, we then provide a quantitative indication of the degree to which GR describes the data. In contrast, finding inconsistency with zero may indicate that GR does not agree with the observed data, which will lay the foundation for a new theoretical model. However, claiming a violation of GR would require extensive studies to rule out false identifications of GR deviations. Potential sources of false deviations include noise systematics, waveform systematics (missing physics in the source model such as spin precession and higher-order modes [24–26], eccentricity [27,28]), astrophysical environmental effects [29,30], and gravitational lensing [31–33], among others. A comprehensive list of these potential sources is provided in Ref. [34].

Parametrized tests of the phasing coefficients within a Bayesian framework have been extensively developed for the LIGO-Virgo-KAGRA (LVK) analyses of testing GR with GW transient catalogs [35–37]. The results have been reported for two implementations: the test infrastructure for general relativity (TIGER) approach [19–21], which is based on a frequency-domain phenomenological waveform family, and the flexible theory-independent (FTI) approach [38,39], which can be applied to any aligned-spin frequency-domain waveform model.

In this work, we develop a TIGER framework for the parametrized tests using the IMRPHENOMX waveform family with high-order modes and precession. We incorporate the deviation parameter into the phasing coefficients of the dominant quadrupole waveform IMRPHENOMXAS model. In the inspiral phase, we introduce a fractional deviation to each PN coefficient, starting from 0PN up to 3.5PN order. We also introduce absolute deviation parameters at -1 PN and 0.5 PN, as many alternative theories predict nonzero value at those PN orders, but they are uniquely zero in Einstein’s theory. In the postinspiral part, we introduce the fractional deviation to the phenomenological coefficients of the model itself. During the inspiral phase, the phasing coefficients in the higher-order modes are connected through a scaling relation, which enables us to propagate the same deviation into the high-order modes of the waveform.

The earlier version of TIGER was developed by modifying the phasing coefficients of the IMRPHENOMD approximant [21,40,41]. Since IMRPHENOMD forms the baseline of IMRPHENOMPv3HM [42], this allowed the corrections to propagate into the higher-order modes. However, IMRPHENOMPv3HM is not well suited for highly

asymmetric or strongly precessing binaries, and it becomes comparatively slow when precession and higher-order modes are enabled. In contrast, the new TIGER framework employs the more accurate and computationally efficient IMRPHENOMXPHM waveform model [43]. The IMRPHENOMX family was calibrated to a substantially larger set of numerical-relativity simulations [43–45], improving accuracy across a broader region of the parameter space. The IMRPHENOMX family also incorporates several optimizations and an efficient twisting-up procedure, making IMRPHENOMXPHM significantly faster to evaluate than IMRPHENOMPv3HM and, therefore, better suited for the large-scale Bayesian analyses required for parameterized tests with TIGER. Our new implementation, thus, enables computationally efficient and more accurate tests of GR for highly asymmetric and precessing binaries, without incurring systematic biases from missing physical effects.

The rest of this paper is organized as follows: Sec. II describes the various waveform models in the IMRPHENOMX family, phasing coefficients in the IMRPHENOMXAS model, parametrized waveform in the TIGER framework, and the setup of a Bayesian analysis framework to perform the parametrized test; Sec. III demonstrates how missing physics like spin precession or higher-order modes in the TIGER can lead to false deviation from GR and our new implementation alleviates this issue; Sec. IV presents the analyses of the GWTC-3 events and includes a separate discussion of two notable asymmetric binary mergers, GW190412 and GW190814. Finally, we conclude in Sec. V.

II. WAVEFORM MODEL AND PARAMETRIZED TEST OF GENERAL RELATIVITY

We develop the improved TIGER framework based on the quasicircular IMRPHENOMX waveform family. This new framework includes two classes of waveform models: one for binary black holes (IMRPHENOMXAS [44], IMRPHENOMXP [43], IMRPHENOMXHM [45], and IMRPHENOMXPHM [43]) and another for binary neutron stars (IMRPHENOMXAS_NRTIDALV2 [46], IMRPHENOMXP_NRTIDALV2 [46], IMRPHENOMXAS_NRTIDALV3 [47], and IMRPHENOMXP_NRTIDALV3 [47]). The IMRPHENOMXAS model describes the dominant $\ell = 2$, $|m| = 2$ spherical harmonic mode of nonprecessing binary. The extension of that model is IMRPHENOMXHM, which contains all the potential higher harmonics $(\ell, |m|) = (2, 1), (3, 3), (3, 2), (4, 4)$, and mode mixing effects for the $(3, 2)$ spherical harmonic. The IMRPHENOMXPHM model is an extension of IMRPHENOMXHM to describe the precessing binaries based on the technique of “twisting up” the nonprecessing waveform. This technique maps the aligned-spin waveform modes in the coprecessing frame to the precessing waveform modes in the inertial frame. The IMRPHENOMXP model is a particular case of IMRPHENOMXPHM in

that it allows the dominant quadrupole contribution in the coprecessing frame. For binary neutron star models, the tidal corrections are added to the aligned-spin model IMRPHENOMXAS, which is known as IMRPHENOMXAS_NRTIDALV2model. Its precession version IMRPHENOMXP_NRTIDALV2 describes the matter effects in the twisting-up framework. The most recent model NRTIDALV3 improves upon previous semianalytical binary neutron star model NRTIDALV2 by calibrating with a larger set of NR data and including dynamical tidal effects, thereby providing more accurate modeling for high-mass-ratio systems and covering a wider range of equations of state.

Since IMRPHENOMXAS serves as the baseline model for all the models in the IMRPHENOMX waveform family, the parametrized deviation parameters are integrated into the phase coefficients of this baseline model. We follow a similar treatment as introduced in the previous studies with IMRPHENOMP waveform family [21,48].

A. IMRPHENOMX phasing model

The frequency-domain phasing of the IMRPHENOMXAS model comprises three regimes [44]: inspiral, intermediate,

and merger-ringdown. In the phenomenological model, these regimes are implemented separately using different physical approaches: a PN formalism for the low-frequency inspiral, a quasinormal ringdown approach for high frequencies, and an intermediate regime that bridges the low- and high-frequency evolution by calibrating with NR simulations. Finally, these three segments are smoothly connected by ensuring the C^1 continuous condition in both phase and amplitude. For explicit details of the implementation, we refer the reader to [44,49]; here, we describe only the phase, which is primarily used to design the parametrized test.

(1) *Inspiral regime.* A low-frequency inspiral regime is described through PN expansion, with additional terms included at higher PN orders. The PN expansion is parametrized by PN coefficients $\{\varphi_0, \dots, \varphi_7\}$ and $\{\varphi_5^{(\ell)}, \varphi_6^{(\ell)}\}$. The higher-order additional terms are pseudo-PN coefficients $\{\sigma_0, \dots, \sigma_5\}$ that appear between the 4PN and 6PN orders. The full inspiral phase is written as

$$\Phi_{\text{Ins}} = 2\pi f t_c - \varphi_c - \frac{\pi}{4} + \underbrace{\frac{3}{128\eta} \sum_{i=0}^7 [\varphi_i(\vec{\lambda}) + \varphi_i^{(\ell)}(\vec{\lambda}) \log(\pi M f)] (\pi M f)^{(i-5)/3}}_{\text{standard PN terms}} + \underbrace{\frac{1}{\eta} \left(\sigma_0 + \sigma_1 f + \frac{3}{4} \sigma_2 f^{4/3} + \frac{3}{5} \sigma_3 f^{5/3} + \frac{1}{2} \sigma_4 f^{6/3} + \frac{3}{7} \sigma_5 f^{7/3} \right)}_{\text{pseudo-PN terms}}, \quad (2.1)$$

where t_c and φ_c reference time and reference phase, respectively. The parameter vector $\vec{\lambda}$ denotes the source parameters component masses (m_1, m_2) and spins (χ_1, χ_2). The inspiral phase ends at the minimum energy circular orbit (MECO) frequency f_{MECO} , the orbit at which the orbital energy is at its lowest value [50].

(2) *Intermediate regime.* An intermediate regime describes the dynamics of the binary during the merger phase and serves as a bridge between the inspiral and ringdown phases. In the phenomenological model, this regime is characterized by a polynomial ansatz along with an additional Lorentzian term. The polynomial term is parametrized by a set of phenomenological coefficients $\{b_0, \dots, b_4\}$. The intermediate phase is written as

$$\Phi_{\text{Int}} = \frac{1}{\eta} \left\{ b'_0 + b_0 f + b_1 \log f - b_2 f^{-1} - \frac{b_3}{2} f^{-2} - \frac{b_4}{3} f^{-3} - \frac{2c_L}{f_{\text{damp}}} \tan^{-1} \left(\frac{f - f_{\text{RD}}}{2f_{\text{damp}}} \right) \right\}, \quad (2.2)$$

where f_{RD} denotes the ringdown frequency, which approximately corresponds to the peak of the Lorentzian. The quantity f_{damp} denotes the damping frequency, which corresponds to the width of the Lorentzian.

(3) *Merger-ringdown regime.* A high-frequency regime where the waveform is dominated by quasinormal ringdown,

$$\Phi_{\text{MR}} = \frac{1}{\eta} \left\{ c'_0 + c_0 f + \frac{3c_1}{2} f^{2/3} - c_2 f^{-1} - \frac{c_4}{3} f^{-3} + \frac{c_L}{f_{\text{damp}}} \tan^{-1} \left(\frac{f - f_{\text{RD}}}{f_{\text{damp}}} \right) \right\}, \quad (2.3)$$

where the negative powers of frequency are added to account for the steep gradient in the inspiral phase.

For comparable-mass binaries, the standard termination frequency for the inspiral phase is at f_{MECO} . However, when generating waveforms, the transition frequency between the inspiral and intermediate regimes is slightly adjusted for numerical purposes. The termination frequency for the inspiral phase is given by

$$f_{\text{Ins}}^{\Phi} = f_{\text{MECO}} - \delta_R, \quad (2.4)$$

where $\delta_R = 0.03(0.3f_{\text{RD}} + 0.6f_{\text{ISCO}} - f_{\text{MECO}})$. Here, f_{ISCO} represents the frequency at the innermost stable circular orbit (ISCO).

The transition frequency between the intermediate and merger-ringdown phases is given by

$$f_{\text{Int}}^{\Phi} = 0.6(0.5f_{\text{RD}} + f_{\text{ISCO}}) + \frac{1}{2}\delta_R. \quad (2.5)$$

It is important to note that the transition frequencies for the amplitude differ from those for the phase [44].

B. Parametrized waveform in the TIGER approach

Within the TIGER framework, we construct the parametrized waveform model by incorporating a fractional deviation parameter ($\delta\hat{p}_i$) into the phasing coefficients:

$$\phi_i^{\text{GR}} \rightarrow (1 + \delta\hat{p}_i)\phi_i^{\text{GR}}, \quad (2.6)$$

where ϕ_i^{GR} represents the PhenomX phasing coefficients

$$\phi_i^{\text{GR}} \equiv \underbrace{\{\varphi_0, \dots, \varphi_7, \varphi_5^{(\ell)}, \varphi_6^{(\ell)}\}}_{\text{inspiral}}, \underbrace{\{b_1, b_2, b_3, b_4\}}_{\text{intermediate}}, \underbrace{\{c_1, c_2, c_4, c_L\}}_{\text{merger-ringdown}}. \quad (2.7)$$

We note that the deviations in the inspiral regime are introduced only in the nonspinning part of the PN coefficients, specifically for $\varphi_3, \varphi_4, \varphi_5^{(\ell)}, \varphi_6$, and φ_7 . However, our TIGER implementation in LALSuite software¹ supports deformations in either the spinning or nonspinning part of the PN coefficients or both equally [49].

We exclude the terms with unit power of frequency and the nonlogarithmic term at 2.5PN, since they can be absorbed in a redefinition of t_c and φ_c , respectively. We also exclude the pseudo-PN terms that enter beyond 3.5PN and are used for calibrating with the numerical waveform dataset. The pairs (b'_0, b_0) and (c'_0, c_0) are also excluded, as they are used to ensure the C^1 continuity condition of the waveform.

Many alternative theories of gravity include scalar fields alongside tensor fields, resulting in a leading dipolar contribution at -1PN and a tail-induced dipole contribution at 0.5PN [52–56]. In contrast, solely GR does not permit these parameters. To search for the non-GR dipolar radiation, we introduce the absolute deviation parameters at -1PN and 0.5PN. The correction to the GW phase is given by

$$\Delta\Phi = \delta\hat{\varphi}_i \frac{3}{128\eta} (\pi M f)^{(i-5)/3}, \quad (2.8)$$

¹There was an issue in the TIGER implementation in the LALSuite software for the -1PN and -0.5PN cases. Please use the checkout [51] or the master branch until a new tag is released.

where $i = -2$ and $i = 1$ correspond to the -1PN and 0.5PN terms, respectively.

1. Parametrized deviation into higher-order modes

The GW wave signal $h(t)$, propagating along an arbitrary direction (ι, ϕ_0) in the source frame, can be decomposed over the spin-weighted spherical harmonic basis (with spin-weight -2) as

$$h(t; \vec{\lambda}, \iota, \phi_0) = \sum_{\ell=2}^{\infty} \sum_{m=-\ell}^{m=\ell} h_{\ell m}(t; \vec{\lambda}) {}_{-2}Y_{\ell m}(\iota, \phi_0), \quad (2.9)$$

where $h_{\ell m}(t; \vec{\lambda}) = A_{\ell m}(t; \vec{\lambda}) e^{i\Phi_{\ell m}(t; \vec{\lambda})}$ represents the (ℓ, m) mode described by the amplitude $A_{\ell m}(t; \vec{\lambda})$ and phase $\Phi_{\ell m}(t; \vec{\lambda})$. In particular, for nonprecessing spinning BHs, the inspiral phase of an arbitrary (ℓ, m) mode can be expressed in terms of the phase of the (2,2) mode alone: $\Phi_{\ell m}(t) \simeq (m/2)\Phi_{22}(t)$. At a given PN order, if a phase correction is required for the (2,2) mode to accurately represent the signal, the same correction must be included in the (ℓ, m) mode phase, scaled by a factor of $m/2$. This suggests that the deviation introduced to the dominant mode should also be propagated to all higher-order modes. As we construct our parametrized model in the frequency domain, it is convenient to express the multimode as

$$\tilde{h}_{\ell m}(f) = A_{\ell m}(f) e^{i\Phi_{\ell m}(f)}. \quad (2.10)$$

For the inspiral phase, the relation among the modes in the time domain can be expressed in the Fourier domain as

$$\Phi_{\ell m}^{\text{Ins}}(f) \approx \frac{m}{2} \Phi_{22}^{\text{Ins}}\left(\frac{2}{m}f\right), \quad (2.11)$$

where Φ_{22}^{Ins} is given in Eq. (2.1). This relation allows us to propagate the same deviation in PN coefficients into the higher-order modes. The non-GR correction to the inspiral phase for (ℓ, m) mode is expressed as

$$\Delta\Phi_{\ell m} = \frac{3}{128\eta} \left(\frac{2\pi M f}{m}\right)^{(k-5)/3} \times \begin{cases} \delta\hat{\varphi}_k^{(\ell)} \varphi_k^{(\ell)}(\vec{\lambda}) \log\left(\frac{2\pi M f}{m}\right), & \text{for log terms,} \\ \delta\hat{\varphi}_k, & \text{for } k = -2, 1, \\ \delta\hat{\varphi}_k \varphi_k(\vec{\lambda}), & \text{otherwise} \end{cases} \quad (2.12)$$

where the superscript (ℓ) denotes the coefficients of the log terms in PN expansion, not related to the (ℓ, m) mode.

The scaling relation between the modes is valid only during the inspiral phase and does not extend to the intermediate and merger-ringdown stages. In the IMRPHENOMXHM model,

the phenomenological coefficients for the postinspiral part are modeled separately for individual modes. As a result, deviations in the postinspiral phase do not affect the higher-order modes. No postinspiral deviation parameters are introduced for the higher-order modes in this new TIGER framework.

2. Twisting up the parametrized deviation

In the context of modeling GWs from precessing spin binaries, “twisting up” is a technique to map between nonprecessing waveform modes in the coprecessing frame and precessing waveform modes in the inertial frame [57]. When the spin components of the BHs are misaligned with the orbital angular momentum, the binary plane experiences a general relativistic precession and nutation, which results in an observable modulation in the phase and amplitude of the GW signal [58,59]. The twisting-up technique approximates amplitude and phase modulation by performing a time-dependent rotation over the nonprecessing modes [60].

To outline the procedure, we consider the L frame (also known as the coprecessing frame), where the z axis is aligned with the orbital angular momentum, and the J frame (the inertial frame), where the z axis is aligned with the total angular momentum. The Euler angles (α, β, γ) are defined to represent an active rotation from the inertial J frame to the precessing L frame. The GW modes in these two frames are related through the transformation of a Weyl scalar under a rotation $\mathcal{R} \in \text{SO}(3)$:

$$h_{\ell m}^J = \sum_{m'=-\ell}^{\ell} \mathcal{D}_{mm'}^{\ell*}(\alpha, \beta, \gamma) h_{\ell m'}^L, \quad (2.13)$$

$$h_{\ell m'}^L = \sum_{m=-\ell}^{\ell} \mathcal{D}_{mm'}^{\ell}(\alpha, \beta, \gamma) h_{\ell m}^J, \quad (2.14)$$

where $\mathcal{D}_{mm'}^{\ell*}(\alpha, \beta, \gamma)$ are the Wigner D matrices:

$$\mathcal{D}_{mm'}^{\ell}(\alpha, \beta, \gamma) = e^{im\alpha} e^{im'\gamma} d_{mm'}^{\ell}(\beta), \quad (2.15)$$

where $d_{mm'}^{\ell}(\beta)$ are the real-valued Wigner-d matrices. The frequency-domain expressions for the GW polarizations in the inertial J frame $\tilde{h}_{+,\times}^J(f)$ in terms of spherical harmonic modes $\tilde{h}_{\ell m}^L(f)$ in the coprecessing L frame are given as

$$\begin{aligned} \tilde{h}_{+,\times}^J(f > 0) &= \frac{1}{2} \sum_{\ell \geq 2} \sum_{m' > 0}^{\ell} \tilde{h}_{\ell - m'}^L(f) e^{im'\gamma} \\ &\times \sum_{m=-\ell}^{\ell} [e^{-im\alpha} d_{m-m'}^{\ell}(\beta) {}_{-2}Y_{\ell m} \\ &\pm (-1)^{\ell} e^{-im\alpha} d_{mm'}^{\ell}(\beta) {}_{-2}Y_{\ell m}^*], \end{aligned} \quad (2.16)$$

where the plus-minus symbol $\pm$ corresponds to $+$ polarizations, respectively. The modes $\tilde{h}_{\ell m}$ and $\tilde{h}_{\ell - m}$ are connected by equatorial symmetry, expressed as $\tilde{h}_{\ell m}(f) =$

$(-1)^{\ell} \tilde{h}_{\ell - m}^*(-f)$, which is valid only for nonprecessing binaries. Here, the modes are $(\ell, |m|) = (2, 1), (3, 3), (3, 2), (4, 4)$, as implemented in the IMRPHEMOMXHM model [45]. For complete details on this twisting-up technique for the IMRPHEMOMXPHM model, we refer the reader to [43].

In the TIGER framework, we apply the phase correction in the L frame, which does not affect the spin dynamics of the binary. This implies that Euler angles are completely determined by the GR parameters and are not changed by the parametrized deviation. However, our beyond-GR correction in the L frame,

$$\Delta \tilde{h}_{\ell m} = \tilde{h}_{\ell m} (e^{i\Delta\Phi_{\ell m}} - 1), \quad (2.17)$$

is twisted up to obtain the waveforms in the J frame, where $\Delta\Phi_{\ell m}$ is given in Eq. (2.12). Therefore, the beyond-GR correction in the TIGER framework can be obtained by substituting $\Delta \tilde{h}_{\ell m}$ into Eq. (2.16), yielding

$$\begin{aligned} \Delta \tilde{h}_{+,\times}^J(f > 0) &= \frac{1}{2} \sum_{\ell \geq 2} \sum_{m' > 0}^{\ell} \Delta \tilde{h}_{\ell - m'}^L(f) e^{im'\gamma} \\ &\times \sum_{m=-\ell}^{\ell} [e^{-im\alpha} d_{m-m'}^{\ell}(\beta) {}_{-2}Y_{\ell m} \\ &\pm (-1)^{\ell} e^{-im\alpha} d_{mm'}^{\ell}(\beta) {}_{-2}Y_{\ell m}^*]. \end{aligned} \quad (2.18)$$

This expression shows how the parametrized deviations in the TIGER framework are twisted up when accounting for the frame transformation effects.

3. Connection with IMRPHEMOMP model

In the previous version of TIGER with the IMRPHEMOMP family [42,61,62], parametrized deviation terms were incorporated into the phasing coefficients of the IMRPHEMOMD model [40,41]. The inspiral phasing for both the IMRPHEMOMD and IMRPHEMOMXAS models is based on the TaylorF2 approximation, derived using the stationary phase approximation, and its functional form is provided in Eq. (2.1). For the postinspiral phase, the IMRPHEMOMX model employs a similar phasing approach as the IMRPHEMOMD model. However, there are differences in the phasing coefficients, their frequency powers, and the transition frequencies between the three frequency regimes.

The intermediate phasing of the IMRPHEMOMD model is given in Eq. (16) in Ref. [41] as

$$\Phi_{\text{Int}} = \frac{1}{\eta} \left\{ \beta_0 + \beta_1 f + \beta_2 \log(f) - \frac{\beta_3}{3} f^{-3} \right\}, \quad (2.19)$$

where $\beta_0, \dots, \beta_3$ are phenomenological coefficients. In the previous TIGER framework, the parametrized deviations were introduced into β_2 and β_3 . In the IMRPHEMOMXAS phasing, as shown in Eq. (2.2), there are two analogous

terms with the same frequency power, where $\beta_2 \leftrightarrow b_1$ and $\beta_3 \leftrightarrow b_4$.

The merger-ringdown phase of IMRPHENOMD model is given in Eq. (14) of Ref. [41] as

$$\Phi_{\text{MR}} = \frac{1}{\eta} \left\{ \alpha_0 + \alpha_1 f - \alpha_2 f^{-1} + \frac{4}{3} \alpha_3 f^{3/4} + \alpha_4 \tan^{-1} \left(\frac{f - \alpha_5 f_{\text{RD}}}{f_{\text{damp}}} \right) \right\}, \quad (2.20)$$

where $\alpha_0, \dots, \alpha_5$ are phenomenological coefficients. In the previous TIGER framework, the parametrized deviations were introduced into α_2, α_3 , and α_4 . In IMRPHENOMXAS phasing, shown in Eq. (2.3), corresponding terms appear at the same frequency power, with $\alpha_2 \leftrightarrow c_2$.

In both models, there is a subset of coefficients that enter at the same frequency power, but the modeling differs. The IMRPHENOMD coefficients are obtained by fitting a 2D parameter space of symmetric mass ratio $\eta \equiv m_1 m_2 / (m_1 + m_2)^2$ and effective spin $\chi_{\text{eff}} \equiv m_1 \chi_1 + m_2 \chi_2 / (m_1 + m_2)^2$, with different effective spins for different frequency regimes as needed. In contrast, IMRPHENOMXAS coefficients are derived by fitting a 3D parameter space, still expressed in terms of the symmetric mass ratio, effective spin, and spin difference $\Delta\chi = \chi_1 - \chi_2$. The additional term $\Delta\chi$ accounts for the unequal spin contribution. In Sec. IV, we did not include the IMRPHENOMP-based results of GWTC-1/2 events for the postinspiral parameters in our combined results with GWTC-3 events, though we did include them for the inspiral parameters.

C. Bayesian analysis framework

The parametrized test with IMRPHENOMX family described above introduces an additional set of non-GR parameters:

$$\delta\hat{p}_i \equiv \underbrace{\{\delta\hat{\phi}_{-2}, \delta\hat{\phi}_0, \dots, \delta\hat{\phi}_7, \delta\hat{\phi}_5^{(\ell)}, \delta\hat{\phi}_6^{(\ell)}\}}_{\text{inspiral}} \underbrace{\{\delta\hat{b}_1, \delta\hat{b}_2, \delta\hat{b}_3, \delta\hat{b}_4, \delta\hat{c}_1, \delta\hat{c}_2, \delta\hat{c}_4, \delta\hat{c}_L\}}_{\text{merger-ringdown}}, \quad (2.21)$$

which correspond to the phase coefficients of the IMRPHENOMXAS model as given in Eq. (2.7). Our parametrized test involves a nested hypothesis where the GR is a special case. The beyond-GR model is characterized by all the standard GR signal parameters ($\vec{\theta}_{\text{GR}}$) plus one additional non-GR parameter $\delta\hat{p}_i$, such that $\vec{\theta} \equiv \{\vec{\theta}_{\text{GR}}, \delta\hat{p}_i\}$. For given data d and hypothesis $\mathcal{H}$, we obtain the posterior probability density $p(\vec{\theta}|d, \mathcal{H})$ using Bayes' theorem:

$$p(\vec{\theta}|d, \mathcal{H}) \propto p(d|\vec{\theta}, \mathcal{H}) p(\vec{\theta}|\mathcal{H}), \quad (2.22)$$

where $p(\vec{\theta}|\mathcal{H})$ is the prior probability distribution and $p(d|\vec{\theta}, \mathcal{H})$ is the likelihood. We use the standard likelihood

function assuming additive noise that is stationary and Gaussian [63]:

$$p(d|\vec{\theta}, \mathcal{H}) \propto \exp \left[\frac{1}{2} \langle d - h(\vec{\theta}) | d - h(\vec{\theta}) \rangle \right], \quad (2.23)$$

where the angular brackets, $\langle \cdot | \cdot \rangle$, represent the noise-weighted inner product, defined as

$$\langle a | b \rangle \equiv 4 \text{Re} \int_{f_{\text{low}}}^{f_{\text{high}}} df \frac{\tilde{a}^*(f) \tilde{b}(f)}{S_n(f)}, \quad (2.24)$$

where $S_n(f)$ is the one-sided power spectral density of the detector noise and $\tilde{a}(f)$ and $\tilde{b}(f)$ denote the Fourier transform of two time-series signals $a(t)$ and $b(t)$, respectively. The integration limits f_{low} and f_{high} represent the lowest and highest cutoff frequencies, respectively, of the sensitivity bandwidth. We analyzed all LVK events with a lower cutoff frequency of 20 Hz. This specific configuration was selected based on information available in the LVK public data. We estimate the beyond-GR effects by computing the marginalized posterior probability distribution of the $\delta\hat{p}_i$, integrating the posterior distribution $p(\vec{\theta}|d, \mathcal{H})$ over the nuisance parameters:

$$p(\delta\hat{p}_i|d) = \int \left(\prod_{\vec{\theta} \setminus \{\delta\hat{p}_i\}} d\theta_i \right) p(\vec{\theta}|d, \mathcal{H}). \quad (2.25)$$

To obtain the posterior probability distribution $p(\vec{\theta}|d, \mathcal{H})$ in Eq. (2.22), we typically use the algorithms such as Markov-chain Monte Carlo or nested sampling. Throughout this work, we use the Bilby TGR package [64] based on Bilby [65], with the nested sampling algorithm Dynesty sampler [66].

For the prior probability distribution $p(\vec{\theta}|\mathcal{H})$, we follow the setup provided in the LVK public data. For the GR parameters, we assume a uniform prior on component masses, uniform prior on individual spin magnitudes with isotropic orientations, and a luminosity distance d_L prior that is uniform in comoving volume, given by $p(d_L) \propto d_L^2$. For the non-GR parameter, we assume a uniform prior symmetric around zero.

To assess the effectiveness of the different PN deviation coefficients at spotting deviation from GR or level of consistency with GR, we compute the fraction of the posterior enclosed by the smallest highest posterior density interval that contains the GR value. We refer to this quantity as the GR quantile, denoted as $\mathcal{Q}_{\text{GR}}$. A GR quantile value of 0% indicates that the GR value lies exactly at the peak of the posterior (perfect consistency with GR), whereas a value of 100% means that the GR value lies at the edge or outside the 100% credible region, implying a strong deviation from GR. An intermediate value of $\mathcal{Q}_{\text{GR}}$ quantifies the level of consistency between the GR prediction and the recovered posterior of the deviation parameter.

To compare two competing hypotheses, we compute the ratio of their evidences, referred to as the Bayes factor. Assuming we compare the GR hypothesis ($\mathcal{H}_{\text{GR}}$) against one incorporating non-GR effects ($\mathcal{H}_{\text{nonGR}}$), the Bayes factor is defined as

$$\mathcal{B}_{\text{GR}}^{\text{nonGR}} = \frac{p(d|\mathcal{H}_{\text{nonGR}})}{p(d|\mathcal{H}_{\text{GR}})}, \quad (2.26)$$

where $p(d|\mathcal{H}_{\text{nonGR}})$ and $p(d|\mathcal{H}_{\text{GR}})$ are the evidences for $\mathcal{H}_{\text{nonGR}}$ and $\mathcal{H}_{\text{GR}}$, respectively. We typically compute the evidence by integrating the likelihood $p(d|\vec{\theta}, \mathcal{H})$ weighted by the prior $p(\vec{\theta}|\mathcal{H})$ over the entire parameter space:

$$p(d|\mathcal{H}) = \int p(d|\vec{\theta}, \mathcal{H}) p(\vec{\theta}|\mathcal{H}) d\vec{\theta}. \quad (2.27)$$

To ensure that the posterior distribution in Eq. (2.22) is a valid probability distribution, we normalize the posterior by evidence.

III. IMPACT OF HIGHER HARMONICS AND PRECESSION

In this section, we demonstrate the impact of higher harmonics and precession when one of these components is omitted from the baseline GR waveform model in TIGER. We consider a set of injections that includes significant contributions from higher-order modes and binary precession effects. The parametrized tests are then performed using the IMRPHENOMXHM and IMRPHENOMXP models. The IMRPHENOMXHM model includes higher-order modes but excludes precession, while the IMRPHENOMXP model incorporates precession but omits higher-order modes.

To carry out this study, we select a debatable event in GWTC-3, GW200129_065458 (hereafter GW200129), which was observed as a highly precessing binary merger [67] and recorded the highest SNR among the events in the catalog [68]. With this event, some testing GR pipelines reported false indications of beyond-GR effects [25]. These were attributed either to the use of a nonprecessing baseline GR waveform model or data quality issues in the Livingston detector [69].

Here, we focus on only the waveform systematics due to missing physics in the template waveform model. We choose the maximum likelihood point from the posterior samples generated with the IMRPHENOMXPHM model. We consider three cases by varying the parameter values: (i) using the exact maximum likelihood point (case I), (ii) assuming an edge-on orientation of the binary to achieve the maximum precession effect and a little impact of higher-order modes (case II), and (iii) a highly asymmetric, nonspinning binary at edge-on orientation to achieve the significant impact of higher-order modes (case III). The parameter values of these cases are detailed in Table I. To avoid the statistical errors due to noise systematics, we generate the injections

TABLE I. Parameter values for GW200129-like injections, where case I corresponds to the maximum likelihood point from event parameter estimation with IMRPHENOMXPHM model. Details on the spin parametrization are available in the LALSimulation documentation [72]. The other extrinsic parameters are fixed for all the cases as follows: polarization angle of 2.3, geocentric GPS time of 1264316116.4 seconds, sky location at right ascension 5.6 and declination 0.04, and a coalescence phase of 0.41. The bottom panel shows the logarithmic Bayes factor values for three different models, where the injections were generated using the IMRPHENOMXPHM model.

Parameters	Case I	Case II	Case III
Total mass [$M_{\odot}$]	78.5	78.5	78.5
Mass ratio (m_2/m_1)	0.42	0.42	1/9
Primary spin (a_1, θ_1)	(0.99, 1.42)	(0.99, 1.42)	(0, 0)
Secondary spin (a_2, θ_2)	(0.6, 2.0)	(0.6, 2.0)	(0, 0)
Azimuthal angles ($\phi_{12}, \phi_{\text{JL}}$)	(5.3, 1.4)	(5.3, 1.4)	...
Zenith angle (θ_{IN})	0.4	$\pi/2$	$\pi/2$
Luminosity distance (d_L [Mpc])	910	910	364.0
Waveform model	Log Bayes factor values ($\log_{10}\mathcal{B}$) for GR template		
IMRPHENOMXPHM ($\log_{10}\mathcal{B}^{\text{XPHM}}$)	417.9	101.1	119.2
IMRPHENOMXP ($\log_{10}\mathcal{B}^{\text{XP}}$)	413.6	94.9	83.0
IMRPHENOMXHM ($\log_{10}\mathcal{B}^{\text{XHM}}$)	405.1	57.1	119.6

assuming a zero-noise realization with a three-detector network of Hanford (H1), Livingston (L1), and Virgo (V1). When performing the parameter estimation runs, we weight the likelihood inner product by the power spectral density corresponding to their advanced designed sensitivities: aLIGOZeroDetHighPower for H1 and L1 [70] and AdvVirgo for V1 [71], as implemented in LALSuite. With this setup, we find injection optimal SNR of 45.1, 23.9, and 25.0 for case I, II, and III, respectively.

In Fig. 1, we present the TIGER results for GW200129-like injections listed in Table I. In all the cases, it is evident that neglecting either higher harmonics or precession in the TIGER waveform model can lead to biased conclusions of deviations from GR. In the following paragraph, we report the Bayes factor values ($\log_{10}\mathcal{B}$) for GR runs with three different models, as listed in the bottom panel in Table I. This will help us assess how the absence of either higher harmonics or precession can lead to a biased conclusion.

A. Case I

$\log_{10}\mathcal{B}^{\text{XP}}$ is slightly higher than $\log_{10}\mathcal{B}^{\text{XHM}}$, while $\log_{10}\mathcal{B}^{\text{XPHM}}$ is marginally greater than both, suggesting that the impact of higher harmonics is slightly less than that of precession but not inconsiderable. The top row in Fig. 1 presents the results for this injection. In all cases, the injection value lies within the 90% credible interval, except at -1PN and 0PN for the IMRPHENOMXHM model. The noninclusion of precession for a GW200129-like event can lead to false deviation from GR. While the impact of

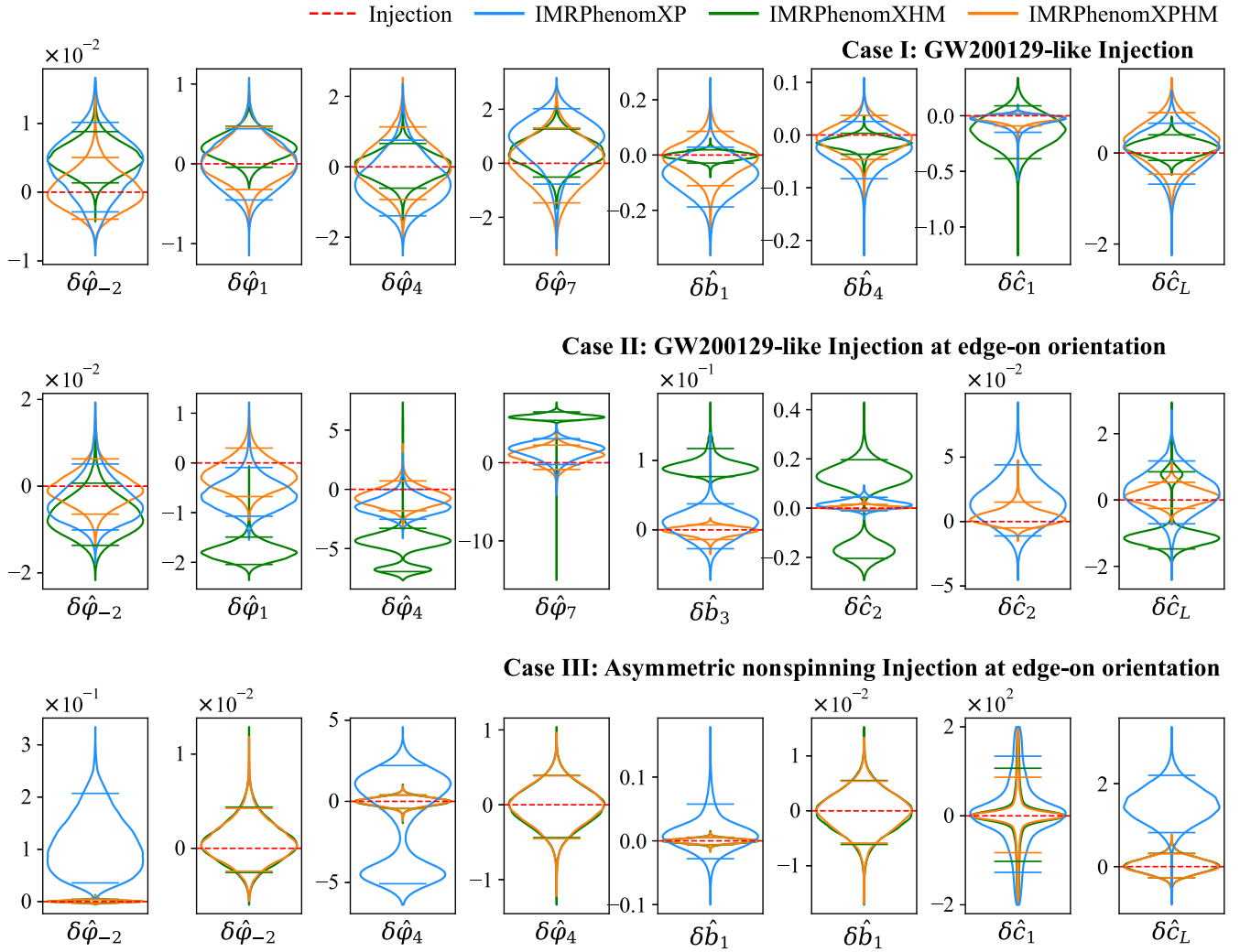


FIG. 1. Systematic errors and false deviation in the parametrized test of GR with TIGER due to missing physics, shown for GW200129-like injections; corresponding parameter values are listed in Table I. The GR injection (shown in red) was generated using a precessing spin with higher-order modes model IMRPHENOMXPHM and recovered with precessing spin without higher-order modes model IMRPHENOMXP (blue), nonprecessing spin with higher-order modes model IMRPHENOMXHM (green), and IMRPHENOMXPHM (orange) model, which includes both. The small horizontal lines represent the 90% credible interval of the posteriors. In the middle and bottom rows, for some deviation parameters, the posteriors obtained using IMRPHENOMXP are significantly broader than those from the other two models. These cases are shown in two consecutive subplots to highlight the difference and to clearly display the narrower posteriors. The full results are shown in Fig. 6.

higher harmonics is not significant, in some PN terms (2–3.5PN), the IMRPHENOMXP posteriors are slightly offset from the injected value.

B. Case II

The $\log_{10} \mathcal{B}^{\text{XP}}$ is significantly larger than $\log_{10} \mathcal{B}^{\text{XHM}}$ and $\log_{10} \mathcal{B}^{\text{XPHM}}$ is slightly larger than the former one, which implies the contribution of precession considerably higher than higher harmonics, but the impact of higher harmonics cannot be ignored. The middle row in Fig. 1 presents the results for this injection. The results obtained using IMRPHENOMXHM indicate that the absence of spin

precession can lead to a significantly biased conclusion, suggesting false evidence of deviations from GR. For all inspiral parameters ($\delta\hat{\phi}_i$) except the -1 PN, and for all parameters in the intermediate regime ($\delta\hat{b}_i$), the GR prediction lies entirely outside the 99% credible interval. The exception at -1 PN term could be due to its non-correlation with precession, because the leading precession effect enters at 2PN. In the IMRPHENOMXHM results, strong bimodality is seen in the posteriors of the merger-ringdown parameters, with almost no support at the injection value. In the IMRPHENOMXP results, for some lower PN terms (0–1PN), the GR prediction falls outside the 90% credible interval. For higher PN terms, the

posteriors are slightly shifted, with the GR prediction close to the 90% credible interval. For the postinspiral parameters, the distributions are sufficiently consistent with the GR prediction but significantly broader compared to the IMRPHENOMXPHM model.

C. Case III

The values of $\log_{10} \mathcal{B}^{\text{XHM}}$ and $\log_{10} \mathcal{B}^{\text{XPHM}}$ are nearly identical and significantly higher than $\log_{10} \mathcal{B}^{\text{XP}}$, indicating a substantial contribution from higher-order modes. The bottom row in Fig. 1 shows the results for this injection. In IMRPHENOMXP results, we see strong deviation at -1PN and one merger-ringdown parameter $\delta\hat{c}_\ell$. For all the remaining inspiral parameters, the posteriors are quite broader and have a long tail and bimodality in higher PN terms starting from 1.5PN . The posteriors of IMRPHENOMXHM and IMRPHENOMXPHM are nearly identical, since the injection does not have the precession contribution. This implies that the deviation parameters are not correlated with the precessing spin parameters for nonspinning systems. We note that the merger-ringdown parameters $\delta\hat{c}_1$ and $\delta\hat{c}_4$ show broad posterior support, indicating that they are not well constrained for nonspinning asymmetric binaries. A comparable behavior is seen for the GW190814 event, as shown in Fig. 5.

IV. ANALYSIS OF GWTC-3 EVENTS

A. Event selection

A deviation parameter in the inspiral regime might not be well measurable, and the resulting posterior could be uninformative if the observed inspiral SNR of the signal is low and vice versa for the postinspiral parameters. Therefore, we analyze only those events that meet specific selection criteria. Following the event selection criteria considered in the previous analyses by the LVK Collaboration [35,36], we consider the events that meet the significance threshold of $\text{FAR} < 10^{-3} \text{ yr}^{-1}$ and impose an additional requirement that the $\text{SNR} > 6$ in the inspiral regime. This criterion determines whether an event is included in the analyses with inspiral deviation parameters. Similarly, we apply the same SNR criterion in the postinspiral regime to decide if an event qualifies for the analysis with postinspiral parameters. As prescribed in the treatment of IMRPHENOMXAS model, we apply the (2,2) mode frequency at MECO to divide the waveform between inspiral and postinspiral regimes. Within the standard PN framework, the description of MECO with arbitrary mass ratio and spins is ill defined. A previous study proposed a hybrid MECO approach to alleviate this issue by combining the information of PN theory with the exact Kerr solution [50]. The corresponding function is implemented in LAL [49] and is labeled as XLALSimIMRPhenomXfMECO. We use that function to compute the cutoff frequency (f_c^{PAR}) between inspiral and postinspiral regimes.

In Table II, we present the optimal SNRs and cutoff frequencies for all selected events. The values listed for GWTC-1 and GWTC-2 events were obtained from analyses conducted by the LVK Collaboration [36]. Parametrized tests for these events were performed using the IMRPHENOMP model, where the cutoff frequency f_c^{PAR} is determined by the condition $GMf/c^3 = 0.018$, with M representing the binary's total mass in the detector frame.

For GWTC-3 events, we used public posterior samples with the tag C01:IMRPHENOMXPHM, provided by the LVK Collaboration and accessible on Zenodo [73]. The bottom section in Table II reports the median values of cutoff frequencies and SNRs for GWTC-3 events. We exclude the event GW191109_010717 from the combined analysis, even though it satisfies the event selection criteria for postinspiral analysis. We found apparent deviations for this event, similar to those reported in the tests of GR paper with GWTC-3 by LVK [37]. Data quality issues in both LIGO detectors are considered to be the explanation for the apparent tension rather than genuine departures from GR. Out of the 15 events, ten meet the criteria for the analysis of inspiral parameters and another nine for postinspiral parameters.

Two notable events, GW190412 and GW190814, were separately highlighted in the GWTC-2 testing GR analyses by the LVK due to the strong presence of higher-order modes and precession in their signals [36]. We also include these two events in our analysis alongside the GWTC-3 events. Both events satisfy the selection criteria for both inspiral and postinspiral deviation parameters. To evaluate the event selection criteria, we use the IMRPHENOMXPHM posterior samples provided in the GWTC-2.1 catalog by the LVK [74].

B. Combining multiple events

In the context of testing GR with multiple events, we typically summarize the findings with a single statement of agreement with GR by combining the results from individual events. This approach can enhance the detectability of potential deviations or constrain them more tightly. LVK analyses usually consider two approaches: (a) multiplying the posterior distributions from all events, assuming a common beyond-GR parameter across them, and (b) employing a hierarchical approach, which assumes that these parameters follow an underlying distribution.

1. Common parameter and multiplying posteriors

If we assume the deviation parameter value is the same across the events, we can combine the evidence to measure that parameter by multiplying individual posteriors for each event. For N events, let us assume the posterior distribution $p(\delta\hat{p}_i^{(j)}|d^{(j)})$ obtained from the j th data $d^{(j)}$ for the deviation parameter $\delta\hat{p}_i$, where $j = 1, \dots, N$. The joint posterior probability is given as

TABLE II. List of events selected for the parametrized test that satisfy the criterion of $\text{FAR} < 10^{-3} \text{ yr}^{-1}$. The quantity f_c^{PAR} represents the cutoff frequency that separates the inspiral and postinspiral regimes, which is used for calculating the optimal SNRs in those two regimes. For the events in GWTC-1 and GWTC-2, this separation frequency corresponds to the transition frequency from inspiral to intermediate regime and is given by the condition $GMf/c^3 = 0.018$, where M is the total mass of the binary in the detector frame. In GWTC-3 events, this separation frequency corresponds to the dominant GW frequency at MECO, the orbit at which the orbital energy is at its lowest value. The quantities ρ_{IMR} , ρ_{insp} , and ρ_{postinsp} are the optimal SNRs of the full signal, the inspiral, and postinspiral regions, respectively. All these listed values are the median of posterior samples. The last two columns specify whether the event is included in parametrized tests for the inspiral (PI), postinspiral (PPI), or both regimes. The listed values for GWTC-1 and GWTC-2 events were obtained from the public data provided by the LVK Collaboration [35,36]. The symbol ✓ denotes that the corresponding event was used in previous LVK analyses, but we excluded it from this work. The symbol ✓* denotes events that meet the selection criteria but are excluded from combined analysis to avoid false deviation due to data-quality issues.

Event	f_c^{PAR} [Hz]	ρ_{IMR}	ρ_{insp}	ρ_{postinsp}	PI	PPI	Event	f_c^{PAR} [Hz]	ρ_{IMR}	ρ_{insp}	ρ_{postinsp}	PI	PPI
GWTC-1													
GW150914	50	24.7	9.6	22.8	✓	✓	GW151226	153	12.3	11.1	5.3	✓	...
GW170104	60	13.4	7.9	11.3	✓	✓	GW170608	179	15.8	14.8	6.3	✓	✓
GW170809	54	12.0	5.8	10.9	...	✓	GW170814	58	16.3	9.1	13.6	✓	✓
GW170818	48	10.8	4.5	10.1	...	✓	GW170823	40	11.5	4.2	11.1	...	✓
GWTC-2													
GW190408_181802	68	15.0	8.3	12.5	✓	✓	GW190412	43	18.9	15.1	11.8	✓	✓
GW190421_213856	36	10.4	2.9	10.0	...	✓	GW190503_185404	39	13.7	4.3	13.0	...	✓
GW190512_180714	87	12.8	10.5	7.4	✓	✓	GW190513_205428	48	13.3	5.1	12.2	...	✓
GW190517_055101	41	11.1	3.4	10.5	...	✓	GW190519_153544	23	15.0	0.0	15.0	...	✓
GW190521	14	13.9	0.0	13.9	...	✓	GW190521_074359	40	25.4	9.7	23.5	✓	✓
GW190602_175927	22	13.1	0.0	13.1	...	✓	GW190630_185205	50	16.3	8.1	14.1	✓	✓
GW190706_222641	19	12.7	0.0	12.7	...	✓	GW190707_093326	161	13.4	12.2	5.5	✓	...
GW190708_232457	103	13.7	11.1	8.0	✓	✓	GW190720_000836	126	10.5	9.2	5.2	✓	...
GW190727_060333	35	12.3	2.0	12.2	...	✓	GW190728_064510	157	12.6	11.4	5.3	✓	...
GW190814	147	24.6	22.9	9.1	✓	✓	GW190828_063405	45	16.2	6.0	15.1	✓	✓
GW190828_065509	80	9.9	6.3	7.6	✓	✓	GW190910_112807	35	14.4	3.3	14.0	...	✓
GW190915_235702	46	13.1	3.7	12.6	...	✓	GW190924_021846	239	12.2	11.8	3.4	✓	...
GWTC-3													
GW191109_010717	26	16.1	1.2	16.0	...	✓*	GW191129_134029	222	12.5	11.9	3.9	✓	...
GW191204_171526	210	16.6	15.5	5.9	✓	...	GW191215_223052	73	10.5	6.1	8.4	✓	✓
GW191216_213338	213	18.0	16.8	6.4	✓	✓	GW191222_033537	36	11.8	2.8	11.5	...	✓
GW200115_042309	487	10.7	10.7	1.0	✓	...	GW200129_065458	62	25.8	12.3	22.6	✓	✓
GW200202_154313	233	10.3	9.8	3.0	✓	...	GW200208_130117	45	9.8	3.6	9.1	...	✓
GW200219_094415	40	9.5	2.3	9.2	...	✓	GW200224_222234	49	18.4	6.0	17.4	...	✓
GW200225_060421	98	11.8	8.0	8.7	✓	✓	GW200311_115853	56	16.5	6.9	15.0	✓	✓
GW200316_215756	182	9.8	9.0	3.7	✓	...							

$$p(\delta\hat{p}_i|\{d^{(j)}\}_{j=1}^N) \propto \prod_{j=1}^N p(\delta\hat{p}_i|d^{(j)}) \quad (4.1)$$

$$\propto \prod_{j=1}^N p(d^{(j)}|\delta\hat{p}_i)p(\delta\hat{p}_i), \quad (4.2)$$

where $p(d^{(j)}|\delta\hat{p}_i)$ is the likelihood and $p(\delta\hat{p}_i)$ is prior. In the parametrized test, we generally consider uniform prior distribution for $\delta\hat{p}_i$. In that case, the joint posterior is proportional to the product of the individual likelihoods and a single instance of the prior [21,75]. However, in practice, we obtain the marginalized posterior samples of a deviation parameter for each event. We use Gaussian kernel

density estimation (KDE) to approximate the posterior distribution for each set of samples. The individual KDEs are then multiplied pointwise to estimate the combined posterior. We evaluate all KDEs on a common grid using the `gaussian_kde` code in SciPy [76].

2. Hierarchical approach

Combining multiple events, assuming the deviation parameter is constant across the events, could lead to an incorrect conclusion. We expect their value to vary with different source properties and choice of alternative theory. As described in Refs. [75,77], we can relax the assumption of a common parameter across all events by using the hierarchical approach. Instead of a common parameter, this

method assumes the non-GR parameters are drawn from a common underlying distribution and then estimates the properties of that distribution using a population of events. Following the previous studies, we model the population distribution of non-GR parameters with a Gaussian distribution, $\delta\hat{p}_i \sim \mathcal{N}(\mu_i, \sigma_i)$, where μ_i and σ_i are the unknown mean and standard deviation, respectively. Setting $\sigma_i = 0$ amounts to assuming that all systems share the same beyond-GR parameter $\delta\hat{p}_i$. From the Bayes theorem, the joint posterior on μ_i and σ_i is given as

$$p(\mu_i, \sigma_i | \{d^{(j)}\}_{j=1}^N) \propto \prod_{j=1}^N p(d^{(j)} | \mu_i, \sigma_i) p(\mu_i, \sigma_i) \\ \propto p(\mu_i, \sigma_i) \prod_{j=1}^N p(d^{(j)} | \mu_i, \sigma_i), \quad (4.3)$$

where $p(\mu_i, \sigma_i)$ is the prior on hyperparameters (μ_i, σ_i) . We assume these to be uniformly distributed and the same for all events. The quantity $p(d^{(j)} | \mu_i, \sigma_i)$ is individual likelihood for the deviation parameter $\delta\hat{p}_i$:

$$p(d^{(j)} | \mu_i, \sigma_i) = \int p(d^{(j)} | \delta\hat{p}_i) p(\delta\hat{p}_i | \mu_i, \sigma_i) d[\delta\hat{p}_i], \quad (4.4)$$

where $p(\delta\hat{p}_i | \mu_i, \sigma_i) = \mathcal{N}(\mu_i, \sigma_i)$ by construction. The quantity $p(d^{(j)} | \delta\hat{p}_i)$ is the standard likelihood obtained from the standard parameter estimation analyses. In practice, we build a Gaussian KDE using the marginalized posterior samples of the deviation parameter as given in Eq. (2.25), which simplifies the evaluation of the integral in Eq. (4.4). After that, we evaluate the joint posterior on hyperparameters in Eq. (4.3) using a Hamiltonian

Monte Carlo algorithm implemented in the STAN package [78]. Finally, using the posterior of hyperparameters $p(\mu_i, \sigma_i | \{d^{(j)}\}_{j=1}^N)$, we infer the combined posterior of $\delta\hat{p}_i$ by marginalizing over μ_i and σ_i :

$$p(\delta\hat{p}_i | \{d^{(j)}\}_{j=1}^N) = \int p(\delta\hat{p}_i | \mu_i, \sigma_i) \\ \times p(\mu_i, \sigma_i | \{d^{(j)}\}_{j=1}^N) d\mu_i d\sigma_i. \quad (4.5)$$

If GR is correct, then both hyperparameters μ_i and σ_i are expected to be consistent with zero. If we find a nonzero μ_i , this is an obvious deviation from GR or a systematic error in the analysis of one or several of the events under consideration.

C. Results and discussion

The unfilled black solid violin plot in Figs. 2 and 3 shows the combined results obtained using the common parameter approach. Results with the hierarchical approach are shown in filled blue distributions. In both cases, all the distributions are consistent with GR, $\delta\hat{p}_i = 0$, and always within the 90% credible symmetric interval. Figure 4 shows the 90% credible region of the joint distribution for the hyperparameters μ and σ . We find a population that the population of all beyond-GR parameters is consistent with GR for both in μ_i and σ_i . All μ_i posteriors are consistent with 0 at the 0.5σ level or better, while all σ_i posteriors peak at 0.

We summarize the combined results in Table III. The hierarchical analysis yields the tightest bound with the -1PN term, $\delta\hat{b}_{-2} = -0.01^{+0.15}_{-0.13} \times 0.002$, within a 90% credible interval. The next tightest bounds are found with intermediate parameters, particularly $\delta\hat{b}_3 = 0.16^{+0.17}_{-0.17} \times 0.05$, within the

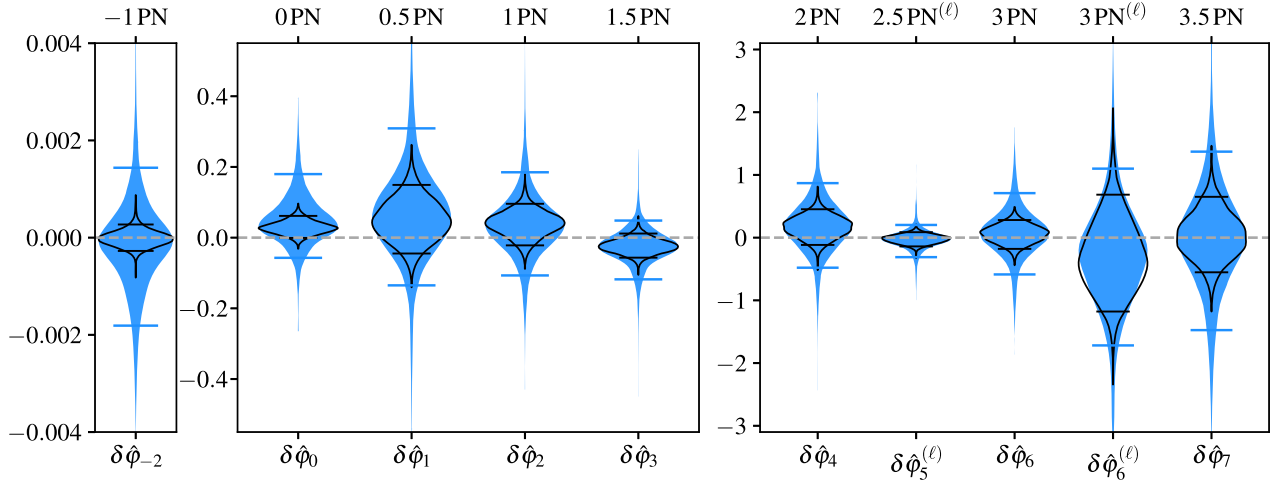


FIG. 2. Illustrating the combined results for the inspiral deviation parameters, obtained from the GWTC-1/2/3 events listed in Table II. The filled color probability distributions represent the results from the hierarchical approach, while the unfilled black solid lines show the combined results assuming a common value for each deviation parameter across all events. In contrast, the hierarchical approach allows for an independent value for each event. The horizontal blue ticks indicate the 90% credible intervals derived from the hierarchical analyses.

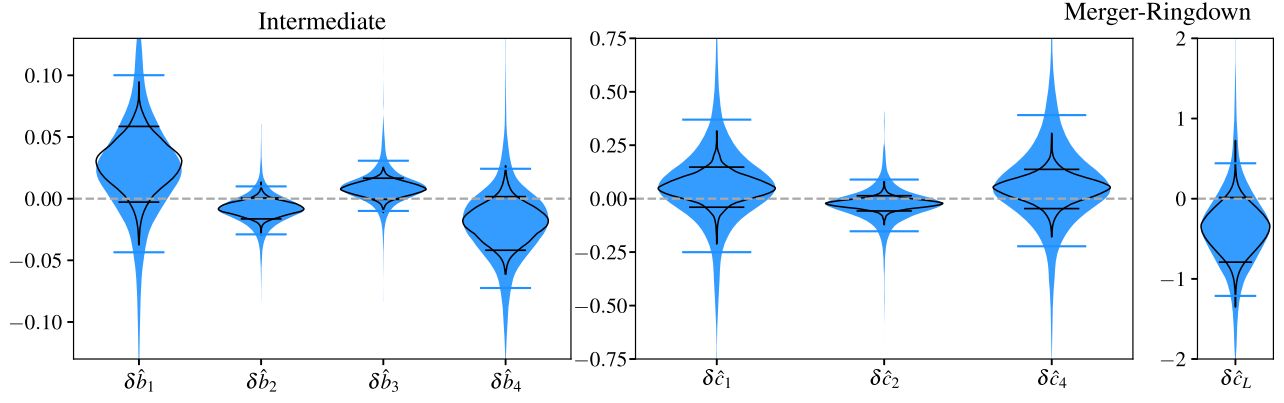


FIG. 3. The same as Fig. 2 but showing the combined results for the postinspiral deviation parameters, obtained from the GW190412, GW190814, and GWTC-3 events listed in Table II.

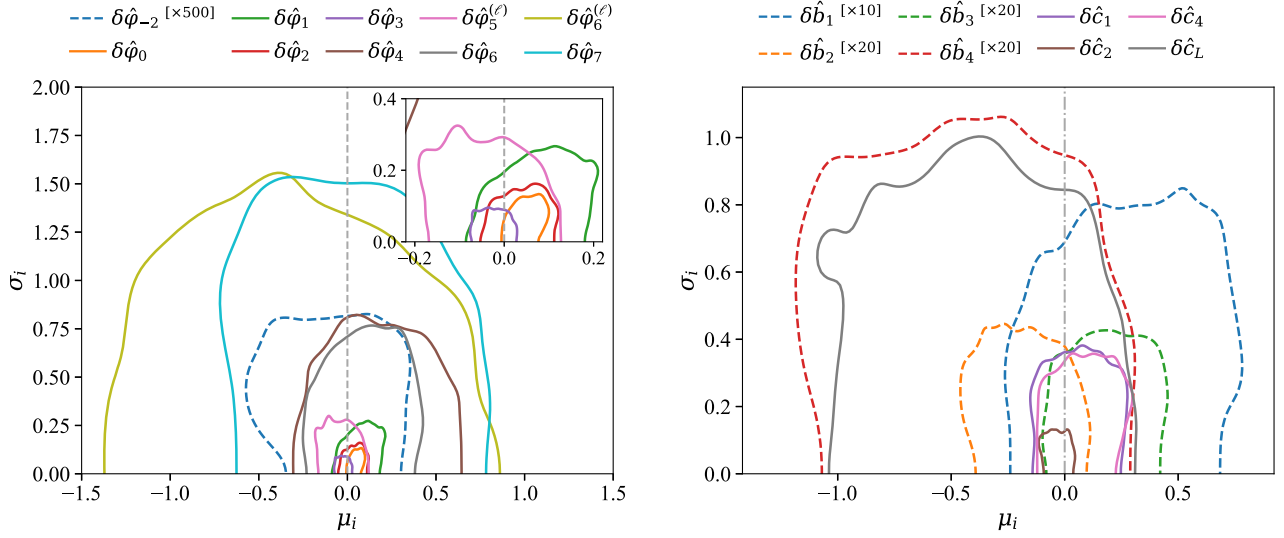


FIG. 4. Illustrating the joint distribution for the hyperparameters μ_i and σ_i of the beyond-GR parameters $\delta\hat{p}_i$ for the parametrized tests. The left panel shows the results for inspiral parameters obtained using the selected events in GWTC-1/2/3 as reported in Table II. The right panel shows the results for postinspiral parameters obtained using the selected events in GWTC-3, GW190412, and GW190814. The contours denote the 90% credible regions of the joint distribution. The contours for $\delta\hat{p}_{-2}$ and intermediate parameters ($\delta\hat{b}_i$) are rescaled by a factor of 1000 and 20, respectively, to improve visibility. All contours include $\mu_i = \sigma_i = 0$, consistent with the GR prediction.

same interval. The widest bound appears with the 3PN log term, $\delta\hat{\phi}_6^{(\ell)} = -0.33^{+1.02}_{-0.85}$. In Table III, we also report the results from the common parameter analysis under the “restricted” column, showing consistently tighter 90% credible intervals for $\delta\hat{p}_i$ compared to the hierarchical analysis. This is expected as common parameter analysis is a special case of hierarchical analysis works assuming $\sigma_i = 0$, which can explore only a subspace of the entire population of the non-GR parameters. We note that this can lead to a biased conclusion if that assumption is not correct. While comparing these two approaches for combining the events, we find significant changes in OPN, 3.5PN, and several postinspiral parameters, as reported in terms of $\mathcal{Q}_{\text{GR}}$ in Table III.

Two widely used approaches in the context of parametrized tests of GR with GW signals are TIGER and FTI. In Appendix B, we compare TIGER results against FTI results reported by LVK. We highlight the results for two O3b events, GW200225_060421 and GW191204_171526, where the first one shows excellent agreement and the other has inconsistency in many of the PN terms. The baseline models of these approaches are not the same, and the FTI model neglects the contribution of higher-order modes and the precession effect, which could lead to disagreement. For GWTC-3 events, we have not noticed any significant inconsistency. The combined results reported in Fig. 2 are also comparable to FTI, except for the -1PN parameter.

TABLE III. Summary of the combined results for each deviation parameter $\delta\hat{p}_i$. All quantities represent the median and 90%-credible intervals except σ_i , for which we provide an upper limit. For both general and restricted results, $\mathcal{Q}_{\text{GR}}$ is the GR quantile associated with Figs. 2 and 3.

$\hat{p}_i$	Hierarchical				Restricted		$\hat{p}_i$	Hierarchical				Restricted	
	μ_i	σ_i	$\delta\hat{p}_i$	$\mathcal{Q}_{\text{GR}} [\%]$	$\delta\hat{p}_i$	$\mathcal{Q}_{\text{GR}} [\%]$		μ_i	σ_i	$\delta\hat{p}_i$	$\mathcal{Q}_{\text{GR}} [\%]$	$\delta\hat{p}_i$	$\mathcal{Q}_{\text{GR}} [\%]$
$\varphi_{-2}^{(\times 500)}$	$-0.06^{+0.34}_{-0.43}$	< 0.72	$-0.05^{+0.77}_{-0.86}$	16.1	$-0.01^{+0.15}_{-0.13}$	14.4	$b_{1[\times 20]}$	$0.49^{+0.82}_{-0.74}$	< 1.32	$0.48^{+1.52}_{-1.35}$	54.0	$0.57^{+0.6}_{-0.63}$	85.7
φ_0	$0.04^{+0.06}_{-0.04}$	< 0.11	$0.04^{+0.14}_{-0.1}$	42.3	$0.03^{+0.03}_{-0.03}$	84.4	$b_{2[\times 20]}$	$-0.18^{+0.22}_{-0.21}$	< 0.35	$-0.17^{+0.37}_{-0.41}$	55.1	$-0.16^{+0.17}_{-0.17}$	87.8
φ_1	$0.07^{+0.12}_{-0.11}$	< 0.21	$0.06^{+0.25}_{-0.2}$	43.7	$0.05^{+0.1}_{-0.09}$	42.1	$b_{3[\times 20]}$	$0.19^{+0.22}_{-0.23}$	< 0.35	$0.18^{+0.44}_{-0.38}$	63.6	$0.16^{+0.17}_{-0.17}$	86.3
φ_2	$0.04^{+0.07}_{-0.07}$	< 0.12	$0.03^{+0.15}_{-0.14}$	37.3	$0.04^{+0.06}_{-0.06}$	71.8	$b_{4[\times 20]}$	$-0.41^{+0.52}_{-0.63}$	< 0.83	$-0.39^{+0.87}_{-1.06}$	34.3	$-0.38^{+0.42}_{-0.45}$	87.8
φ_3	$-0.03^{+0.04}_{-0.04}$	< 0.08	$-0.02^{+0.07}_{-0.09}$	47.5	$-0.02^{+0.04}_{-0.03}$	77.9	c_1	$0.06^{+0.17}_{-0.16}$	< 0.3	$0.06^{+0.31}_{-0.31}$	33.1	$0.05^{+0.1}_{-0.09}$	62.6
φ_4	$0.18^{+0.39}_{-0.37}$	< 0.63	$0.18^{+0.69}_{-0.66}$	35.5	$0.17^{+0.28}_{-0.29}$	69.7	c_2	$-0.03^{+0.06}_{-0.07}$	< 0.11	$-0.03^{+0.12}_{-0.13}$	31.2	$-0.02^{+0.04}_{-0.04}$	74.3
$\varphi_5^{(\ell)}$	$-0.03^{+0.11}_{-0.13}$	< 0.24	$-0.03^{+0.23}_{-0.28}$	19.5	$-0.02^{+0.11}_{-0.12}$	6.0	c_4	$0.07^{+0.17}_{-0.14}$	< 0.28	$0.06^{+0.33}_{-0.28}$	23.6	$0.05^{+0.09}_{-0.1}$	63.1
φ_6	$0.09^{+0.25}_{-0.27}$	< 0.58	$0.07^{+0.64}_{-0.65}$	12.0	$0.06^{+0.22}_{-0.24}$	30.0	c_L	$-0.37^{+0.47}_{-0.49}$	< 0.69	$-0.36^{+0.81}_{-0.85}$	67.8	$-0.37^{+0.39}_{-0.42}$	88.0
$\varphi_6^{(\ell)}$	$-0.29^{+0.92}_{-0.87}$	< 1.19	$-0.28^{+1.38}_{-1.44}$	21.2	$-0.33^{+1.02}_{-0.85}$	53.6							
φ_7	$0.01^{+0.64}_{-0.66}$	< 1.28	$-0.0^{+1.37}_{-1.47}$	7.2	$0.03^{+0.63}_{-0.58}$	16.9							

D. Analysis of GW190412 and GW190814

Two notable events in GWTC-2, GW190412 and GW190814, originated from compact binary mergers with highly asymmetric component masses, which enabled the first observation of gravitational-wave radiation beyond the $(\ell, m) = (2, 2)$ mode [79–81]. The presence of higher harmonics has enabled many new tests of GR by examining how these higher-order modes are related to the primary signal [82–89]. Measuring these higher harmonics of the signal also provides a more accurate measurement of source parameters, which allows for tighter constraints on certain deviations from GR.

To assess the improvements due to higher harmonics, we perform the TIGER analysis using two different waveform models: IMRPHENOMXP and IMRPHENOMXPHM. Figure 5

highlights the biases and improvements in the posterior of deviation parameters when higher harmonics are included in the model compared to a model without them. The filled violins represent the posterior for the IMRPHENOMXPHM model, while the black unfilled violins correspond to the IMRPHENOMXP model. Figure 5 show the results for a part of the deviation parameters; violin plots for all the deviation parameters are shown in Fig. 7.

For both the events, the posteriors with IMRPHENOMXPHM model are more consistent with GR value compared to the case of IMRPHENOMXP model. We observe noticeable improvement in the postinspiral parameters. Another noticeable improvement is seen in the 2PN term for the GW190814 event. The $\delta\hat{\varphi}_4$ posterior exhibits bimodality, with the stronger peak deviating from the GR

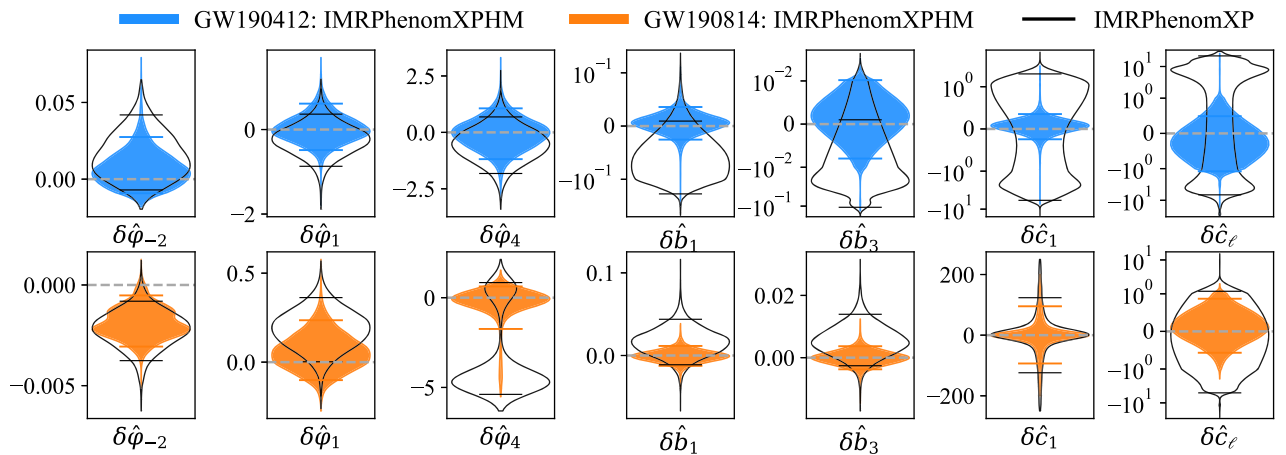


FIG. 5. Illustrating the posterior distributions of parametrized deviation parameters obtained from GW190412 (shown in blue) and GW190814 (shown in orange) events, inferred using the IMRPHENOMXPHM model. The black unfilled violin plots show the results inferred using the IMRPHENOMXP model, highlighting the systematic biases when higher-order modes are missing in the TIGER baseline waveform model. The horizontal solid lines indicate the 90% credible intervals, and the gray dashed line at $\delta\hat{p}_i = 0$ denotes the GR values. The full results are shown in Fig. 7.

value. This behavior is associated with the bimodality in the mass ratio and component masses. This significant improvement indicates the crucial role of higher harmonics in modeling TIGER waveforms to achieve unbiased conclusions in testing GR analysis.

We note that, in the analyses with GW190814, the GR value lies outside the 90% credible interval for several testing parameters (-1PN , 3PN , and 3.5PN). This is not necessarily surprising, as the GR value is found near the tail of the distribution. A similar trend was also observed in earlier analyses reported by the LVK in Fig. 19 of Ref. [36]. However, for the posterior obtained using the FTI framework, the GR value lies within the 90% credible interval, albeit near the edge, and a partially bimodal shape is also observed in several testing parameters, as shown in the bottom row in Fig. 8. We further note that the 0PN posterior for the GW230529 event was found to deviate significantly from the GR value [90]. This is mostly due to strong degeneracies between the testing parameter and the physical parameters of the binary in certain regions of the parameter space of the parametrized waveform.

V. SUMMARY AND CONCLUSION

We have developed a parametrized testing framework based on the IMRPHENOMX models, incorporating spin precession and higher-order modes. Parametrized deviations are introduced in the phasing coefficients of the dominant quadrupole mode $(\ell, m) = (2, 2)$ for nonprecessing binaries. This new implementation allows deviations to propagate into the higher-order modes during the inspiral phase. When generating parametrized waveforms with precessing spin models like IMRPHENOMXP and IMRPHENOMXPHM, the non-GR correction undergoes a twist up due to the transformation of waveform modes from the coprecessing frame to the inertial frame. However, our beyond-GR correction does not interfere with the determination of spin dynamics.

This TIGER framework with IMRPHENOMX models is integrated into the LIGO Algorithm Library (LAL) [49]; the parametrized waveforms can be generated using the LALSimulation library.² The corresponding Bayesian analysis framework is implemented in the Bilby TGR package [64]. While compiling this manuscript, this TIGER framework had already been applied in various studies, including testing GR with the GW230529 event [90,92] and investigating astrophysical environmental effects in tests of vacuum GR [30].

With this framework, we conducted a series of injection studies on GW200129-like events to investigate the impact of spin precession and higher-order modes if either is omitted in the baseline TIGER waveform, as discussed in Sec. III. Our results indicate that, if these contributions are significant,

neglecting one of them can lead to biased conclusions, suggesting false deviations from GR. We also noticed systematic biases in most deviation parameters, even when these contributions are relatively small. Previous studies on systematic biases also reached to the similar conclusions using the earlier version of TIGER [24], where higher-order modes were not included in the baseline model. Our new implementation resolves these systematic biases.

We then performed the analyses for the events in GWTC-3 and reported a combined statement by including the events in GWTC-1/2. We also performed the analyses for two notable events GW190412 and GW190814, which were associated with highly asymmetric binary and in which strong contribution of higher-order modes were observed. To provide the combined statement, we made use of two techniques: common parameter approach and hierarchical approach. With both approaches, we found the combined posterior is consistent with the prediction of GR.

We note that while the IMRPHENOMX waveform family has undergone several recent improvements [93–96], the TIGER results presented in this paper were obtained using IMRPHENOMXPHM-MSA [the IMRPHENOMXPHM model with the multiscale analysis (MSA) treatment of spin precession] [43], consistent with the waveform choices in the GWTC-3 catalog [68]. A recent upgrade, IMRPHENOMXPHM-SPINTAYLOR [95], is now being employed for parameter estimation of LVK events observed during the fourth observing run. This version uses the SPINTAYLOR framework [97] to numerically evolve the spins, providing a more accurate precession evolution compared to the analytical MSA approximation. The TIGER framework works straightforwardly with IMRPHENOMXPHM-SPINTAYLOR, since this model employ the same underlying coprecessing IMRPHENOMXHM implementation as in IMRPHENOMXPHM-MSA. Further efforts were made to improve the spin-precession sector of the IMRPHENOMX waveform family by calibrating the precession angles to numerical relativity, leading to the waveform models IMRPHENOMXO4a [93,94] and IMRPHENOMXPNR [96] that have more recently become available. However, those introduce additional modifications to the coprecessing frame, requiring a dedicated TIGER implementation and further validation.

We have used NumPy [98] and SciPy [76] for analyses and Matplotlib [99] for preparing the plots in this manuscript.

ACKNOWLEDGMENTS

We are grateful to Michalis Agathos, Justin Janquart, and Elise M. Sanger for useful comments on our draft. S. R. is supported by the Fonds de la Recherche Scientifique—FNRS (Belgium). G. P. gratefully acknowledges support from a Royal Society University Research Fellowship No. URF\R1\221500 and No. RF\ERE\221015, as well as STFC Grants No. ST/V005677/1 and No. ST/Y00423X/1. M. H., P. T. H. P., and C. V. D. B. are supported by the

²Available at [91]. There was an issue in the -1PN implementation.

research program of the Netherlands Organisation for Scientific Research (NWO). The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459. The authors are also grateful for the computational resources provided by Cardiff University supported by STFC Grant No. ST/I006285/1. This research has made use of data or software obtained from the Gravitational Wave Open Science Center [100], a service of the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation, as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN), and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, and Spain. K. A. G. R. A. is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

DATA AVAILABILITY

The data that support the findings of this article are openly available [73].

APPENDIX A: IMPACT OF HIGHER HARMONICS AND PRECESSION

In this section, we highlight systematic biases that can arise from neglecting higher-order modes and precession in the baseline waveform of the parametrized test, which can lead to a false indication of deviation from GR. Here, we show the posterior of all the deviation parameters implemented in the new TIGER framework. Figure 6 shows the results for GW200129-like injection described in Sec. III. Figure 7 shows the full results obtained for the GW190412 and GW190814 events, as demonstrated in Sec. IV D.

APPENDIX B: COMPARISON BETWEEN TIGER AND FTI

Another common approach for parametrized tests of GR with GW observations is known as FTI [39]. Similar to the TIGER method, FTI introduces deviations in the phasing coefficients to generate a non-GR waveform. However, instead of directly incorporating deviation parameters into the frequency-domain phase coefficients, it computes the non-GR correction for a given PN term and then adds it to the frequency-domain phase of the GR waveform. To ensure that the postinspiral part of the waveform remains unaffected by the non-GR correction, a windowing function is applied to smoothly taper it off to zero. The current FTI framework is applicable to any aligned-spin waveform model. The FTI results in the LVK analysis were obtained using the SEOBNRv4_ROM model [101], a frequency-domain waveform model for aligned-spin binaries that neglects higher-order modes and precession effects.

We present a comparison between our TIGER results and FTI results reported in Ref. [37]. The data for the FTI analysis have been obtained from the LVK data release [102]. Figure 8 highlights the comparison for two O3b events. For the GW200225_060421 event, we found that the TIGER and FTI posteriors are largely consistent across all deviation parameters, except for the 3PN logarithmic term. We also highlight results for the GW191204_171526 event, which exhibited the largest inconsistency among the GWTC-3 events, with discrepancies visible in multiple PN terms. Since the baseline models of these two approaches differ, full agreement between the results is not expected. Additionally, the absence of higher-order modes and precession effects in the FTI model could contribute to the inconsistencies if the observed signal contains these components.

We note that the FTI analyses of the GW190814 event were carried out using the SEOBNRv4HM_ROM model [103] to mitigate biases arising from the omission of higher-order harmonics [36]. A comparison with the TIGER results is presented in the bottom row in Fig. 8. The posterior distributions for both methods are consistent at lower PN orders (from -1PN to 1PN). However, significant differences appear at higher PN orders, starting from 2.5PN . Notably, for the 3PN and 3.5PN terms, the GR value lies outside the 90% credible interval in the TIGER analysis, while the FTI posteriors remain consistent with the GR predictions. For all three events, the differences are mostly noticeable for the higher PN terms. Since the baseline model of the FTI framework does not account for precession effects, their omission may potentially lead to inconsistencies.

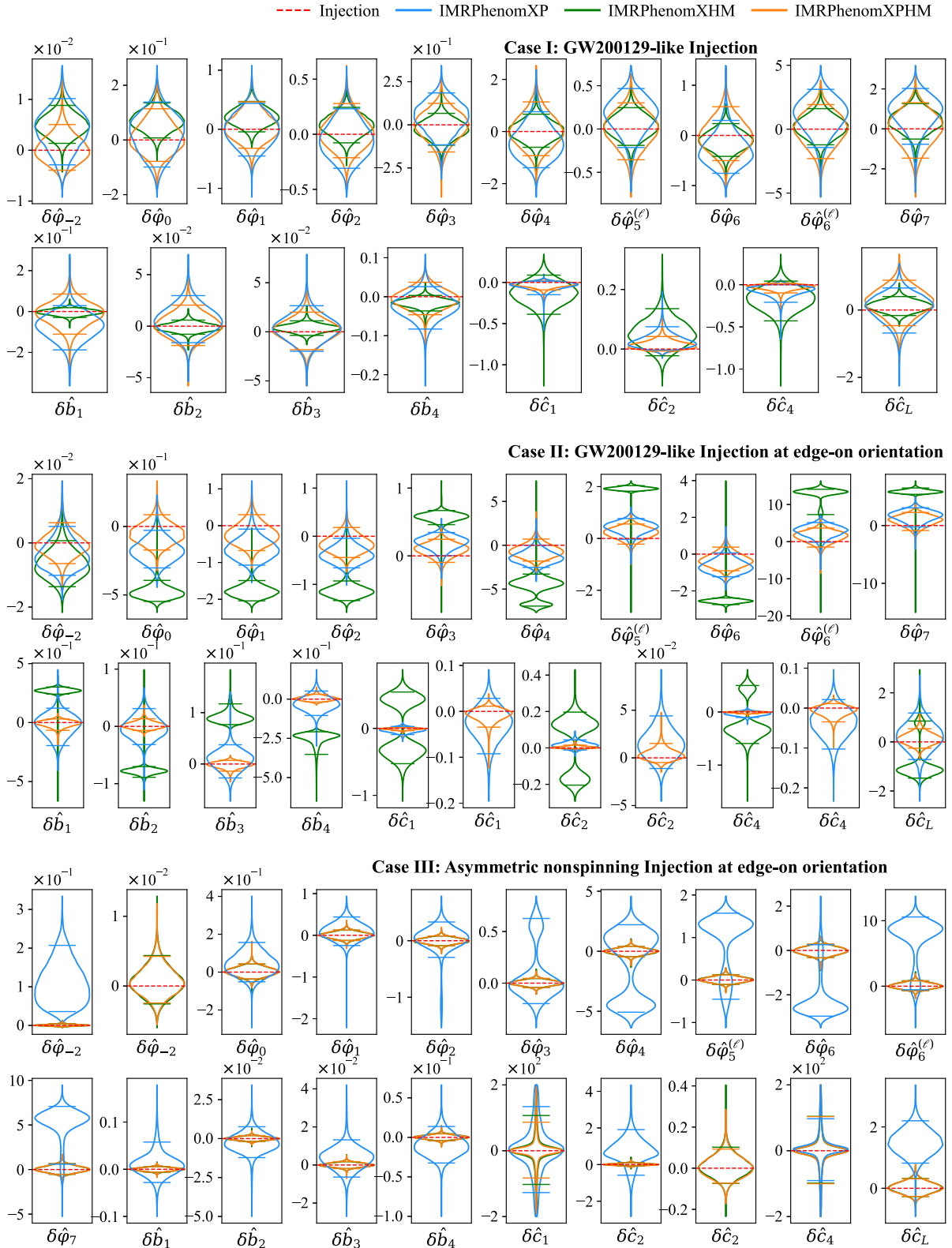


FIG. 6. The same as Fig. 1, but displaying the posterior distributions for all deviation parameters. For some deviation parameters, the posteriors obtained using IMRPHENOMXP are significantly broader than those from the other two models. These cases are shown in two consecutive subplots to highlight the difference and to clearly display the narrower posteriors. These TIGER results for the GW200129-like injections are obtained using different waveform models to highlight the impact of higher-order modes and precession effects in parametrized tests, as described in Sec. III and Appendix A.

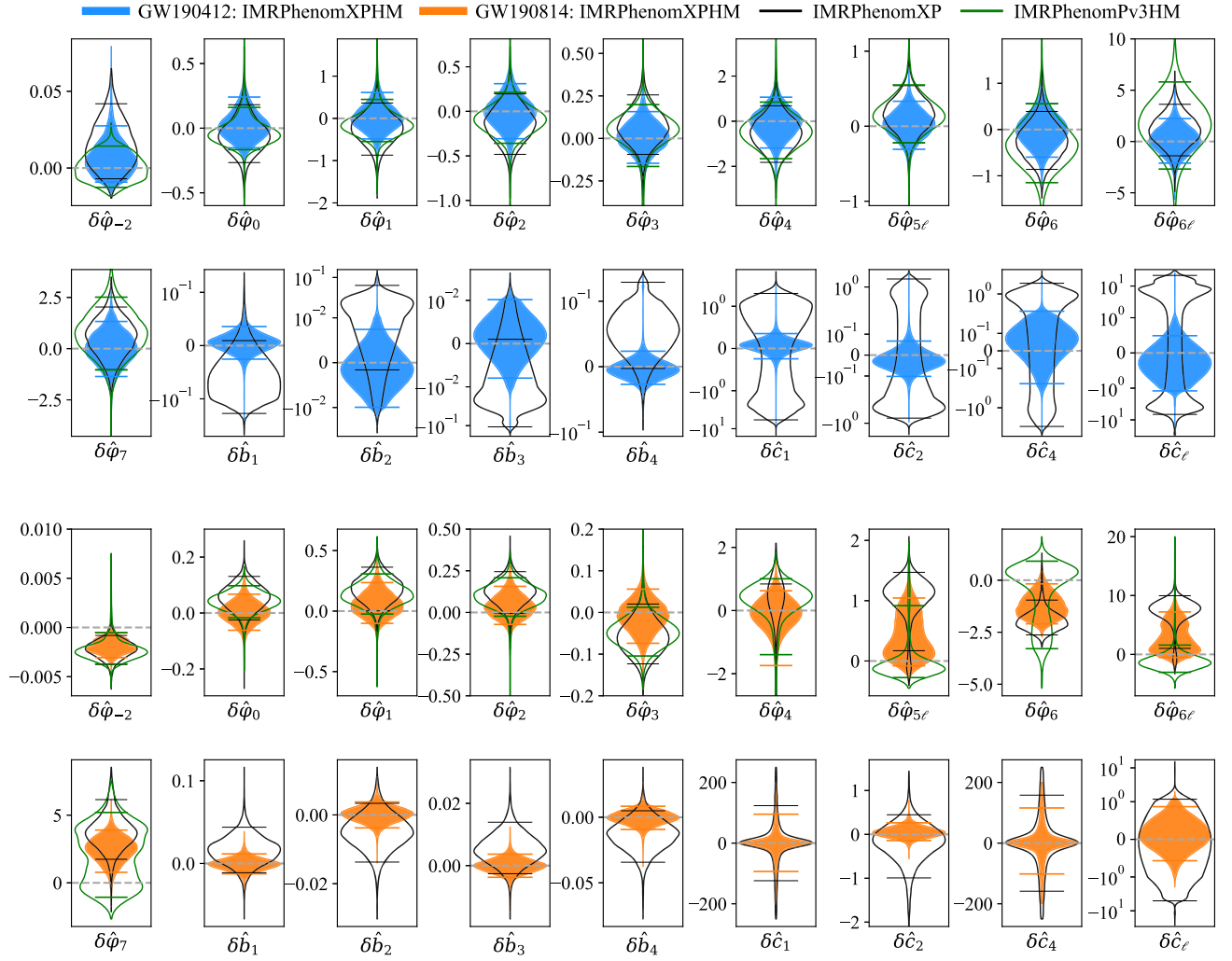


FIG. 7. The same as Fig. 5, but showing the posterior distributions for all deviation parameters. These TIGER results for the GW190412 and GW190814 events are obtained using different waveform models to highlight the impact of higher-order modes and precession effects in parametrized tests, as discussed in Sec. IV D and Appendix A. For the inspiral parameters, we also show the results obtained with the IMRPHENOMPv3HM model, as reported in the GWTC-2 tests of GR paper by LVK [36].

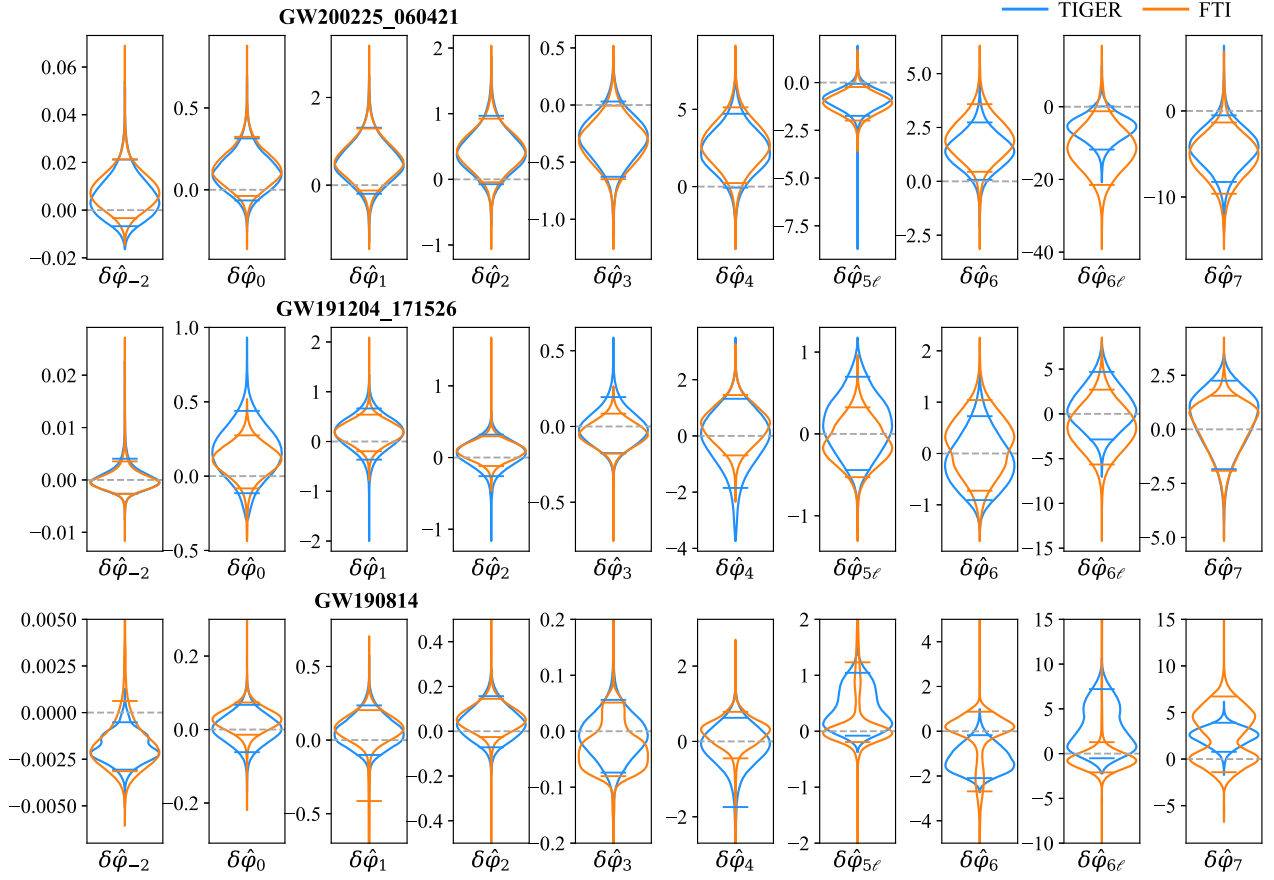


FIG. 8. Comparison of TIGER and FTI results. The top and middle rows correspond to the O3b events GW200225_060421 and GW191204_171526, respectively. The bottom row shows the comparison for the GW190814 event. The FTI results are obtained from the LVK data release [102]. These results are further discussed in Appendix B. The small horizontal lines indicate the 90% credible intervals of the posteriors, and the gray dashed line at $\delta\hat{\phi}_i = 0$ represents the GR prediction.

-
- [1] J. Aasi *et al.* (LIGO Scientific Collaboration), *Classical Quantum Gravity* **32**, 074001 (2015).
 - [2] F. Acernese *et al.* (Virgo Collaboration), *Classical Quantum Gravity* **32**, 024001 (2015).
 - [3] W. E. East and J. L. Ripley, *Phys. Rev. D* **103**, 044040 (2021).
 - [4] M. Bezares, R. Aguilera-Miret, L. ter Haar, M. Crisostomi, C. Palenzuela, and E. Barausse, *Phys. Rev. Lett.* **128**, 091103 (2022).
 - [5] P. Figueras and T. França, *Phys. Rev. D* **105**, 124004 (2022).
 - [6] L. Aresté Saló, K. Clough, and P. Figueras, *Phys. Rev. Lett.* **129**, 261104 (2022).
 - [7] L. Aresté Saló, K. Clough, and P. Figueras, *Phys. Rev. D* **108**, 084018 (2023).
 - [8] L. Aresté Saló, S. E. Brady, K. Clough, D. Doneva, T. Evstafyeva, P. Figueras, T. França, L. Rossi, and S. Yao, *J. Open Source Softw.* **9**, 6369 (2024).
 - [9] M. Corman, J. L. Ripley, and W. E. East, *Phys. Rev. D* **107**, 024014 (2023).
 - [10] M. Corman and W. E. East, *Phys. Rev. D* **110**, 084065 (2024).
 - [11] M. Okounkova, L. C. Stein, M. A. Scheel, and S. A. Teukolsky, *Phys. Rev. D* **100**, 104026 (2019).
 - [12] M. Okounkova, *Phys. Rev. D* **102**, 084046 (2020).
 - [13] R. Cayuso, P. Figueras, T. França, and L. Lehner, *Phys. Rev. Lett.* **131**, 111403 (2023).
 - [14] L. Blanchet and B. S. Sathyaprakash, *Phys. Rev. Lett.* **74**, 1067 (1995).
 - [15] K. G. Arun, B. R. Iyer, M. S. S. Qusailah, and B. S. Sathyaprakash, *Classical Quantum Gravity* **23**, L37 (2006).
 - [16] K. G. Arun, B. R. Iyer, M. S. S. Qusailah, and B. S. Sathyaprakash, *Phys. Rev. D* **74**, 024006 (2006).
 - [17] N. Yunes and F. Pretorius, *Phys. Rev. D* **80**, 122003 (2009).

- [18] C. K. Mishra, K. G. Arun, B. R. Iyer, and B. S. Sathyaprakash, *Phys. Rev. D* **82**, 064010 (2010).
- [19] M. Agathos, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, J. Veitch, and S. Vitale, *Phys. Rev. D* **89**, 082001 (2014).
- [20] 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).
- [21] J. Meidam *et al.*, *Phys. Rev. D* **97**, 044033 (2018).
- [22] H. Narola, S. Roy, and A. S. Sengupta, *Phys. Rev. D* **107**, 024017 (2023).
- [23] A. Sharma, S. Roy, and A. S. Sengupta, *Phys. Rev. D* **109**, 124049 (2024).
- [24] P. T. H. Pang, J. Calderón Bustillo, Y. Wang, and T. G. F. Li, *Phys. Rev. D* **98**, 024019 (2018).
- [25] E. Maggio, H. O. Silva, A. Buonanno, and A. Ghosh, *Phys. Rev. D* **108**, 024043 (2023).
- [26] R. S. Chandramouli, K. Prokup, E. Berti, and N. Yunes, *Phys. Rev. D* **111**, 044026 (2025).
- [27] P. Narayan, N. K. Johnson-McDaniel, and A. Gupta, *Phys. Rev. D* **108**, 064003 (2023).
- [28] M. A. Shaikh, S. A. Bhat, and S. J. Kapadia, *Phys. Rev. D* **110**, 024030 (2024).
- [29] G. Caneva Santoro, S. Roy, R. Vicente, M. Haney, O. J. Piccinni, W. Del Pozzo, and M. Martinez, *Phys. Rev. Lett.* **132**, 251401 (2024).
- [30] S. Roy and R. Vicente, *Phys. Rev. D* **111**, 084037 (2025).
- [31] A. Mishra, N. V. Krishnendu, and A. Ganguly, *Phys. Rev. D* **110**, 084009 (2024).
- [32] M. Wright, J. Janquart, and N. Johnson-McDaniel, *Astrophys. J.* **981**, 133 (2025).
- [33] A. Liu, R. S. Chandramouli, O. A. Hannuksela, N. Yunes, and T. G. F. Li, *Phys. Rev. D* **112**, 064043 (2025).
- [34] A. Gupta *et al.*, *SciPost Phys. Commun. Rep.* **5** (2025).
- [35] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **100**, 104036 (2019).
- [36] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **103**, 122002 (2021).
- [37] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Phys. Rev. D* **112**, 084080 (2025).
- [38] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **123**, 011102 (2019).
- [39] A. K. Mehta, A. Buonanno, R. Cotesta, A. Ghosh, N. Sennett, and J. Steinhoff, *Phys. Rev. D* **107**, 044020 (2023).
- [40] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. Jiménez Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044006 (2016).
- [41] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Jiménez Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044007 (2016).
- [42] S. Khan, F. Ohme, K. Chatziioannou, and M. Hannam, *Phys. Rev. D* **101**, 024056 (2020).
- [43] G. Pratten *et al.*, *Phys. Rev. D* **103**, 104056 (2021).
- [44] G. Pratten, S. Husa, C. Garcia-Quiros, M. Colleoni, A. Ramos-Buades, H. Estelles, and R. Jaume, *Phys. Rev. D* **102**, 064001 (2020).
- [45] 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).
- [46] M. Colleoni, F. A. Ramis Vidal, N. K. Johnson-McDaniel, T. Dietrich, M. Haney, and G. Pratten, *Phys. Rev. D* **111**, 064025 (2025).
- [47] A. Abac, T. Dietrich, A. Buonanno, J. Steinhoff, and M. Ujevic, *Phys. Rev. D* **109**, 024062 (2024).
- [48] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **116**, 221101 (2016); **121**, 129902 (E) (2018).
- [49] LIGO Scientific Collaboration, LIGO Algorithm Library—LALSuite, free software (GPL) (2023).
- [50] M. Cabero, A. B. Nielsen, A. P. Lundgren, and C. D. Capano, *Phys. Rev. D* **95**, 064016 (2017).
- [51] <https://git.ligo.org/lscsoft/lalsuite/-/commits/653065c0f77c76f413aae3b09c41b3b878c11840>.
- [52] C. M. Will, *Phys. Rev. D* **50**, 6058 (1994).
- [53] N. Yunes, P. Pani, and V. Cardoso, *Phys. Rev. D* **85**, 102003 (2012).
- [54] K. G. Arun, *Classical Quantum Gravity* **29**, 075011 (2012).
- [55] E. Barausse, N. Yunes, and K. Chamberlain, *Phys. Rev. Lett.* **116**, 241104 (2016).
- [56] K. Yagi, L. C. Stein, and N. Yunes, *Phys. Rev. D* **93**, 024010 (2016).
- [57] P. Schmidt, M. Hannam, and S. Husa, *Phys. Rev. D* **86**, 104063 (2012).
- [58] T. A. Apostolatos, C. Cutler, G. J. Sussman, and K. S. Thorne, *Phys. Rev. D* **49**, 6274 (1994).
- [59] L. E. Kidder, *Phys. Rev. D* **52**, 821 (1995).
- [60] P. Schmidt, M. Hannam, S. Husa, and P. Ajith, *Phys. Rev. D* **84**, 024046 (2011).
- [61] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, *Phys. Rev. Lett.* **113**, 151101 (2014).
- [62] A. Bohé, M. Hannam, S. Husa, F. Ohme, M. Pürrer, and P. Schmidt, PhenomPv2—technical notes for the LAL implementation, Technical Report No. LIGO-T1500602-v4, 2019, <https://dcc.ligo.org/LIGO-T1500602/public>.
- [63] J. Veitch and A. Vecchio, *Phys. Rev. D* **81**, 062003 (2010).
- [64] G. Ashton, C. Talbot, S. Roy, G. Pratten, T.-H. Pang, M. Agathos, T. Baka, E. Sängner, A. Mehta, J. Steinhoff, E. Maggio, A. Ghosh, and A. Vijaykumar, Bilby TGR (2024).
- [65] G. Ashton *et al.*, *Astrophys. J. Suppl. Ser.* **241**, 27 (2019).
- [66] J. S. Speagle, *Mon. Not. R. Astron. Soc.* **493**, 3132 (2020).
- [67] M. Hannam *et al.*, *Nature (London)* **610**, 652 (2022).
- [68] R. Abbott *et al.* (KAGRA, Virgo, and LIGO Scientific Collaborations), *Phys. Rev. X* **13**, 041039 (2023).
- [69] E. Payne, S. Hourihane, J. Golomb, R. Udall, R. Udall, D. Davis, and K. Chatziioannou, *Phys. Rev. D* **106**, 104017 (2022).
- [70] L. Barsotti, S. Gras, M. Evans, and P. Fritschel, The updated Advanced LIGO design curve, LIGO Technical Note T1800044-v5, T0900288-v3, 2018.
- [71] A. Manzotti and A. Dietz, *arXiv:1202.4031*.
- [72] LIGO Scientific Collaboration, LALSimulation-LALInference Parameter Transformations, https://lscsoft.docs.ligo.org/lalsuite/lalsimulation/group_lalsimulation_inference.html (accessed: October 06, 2024).

- [73] LIGO Scientific, Virgo, and KAGRA Collaborations, [10.5281/zenodo.8177023](https://zenodo.org/record/8177023) (2023).
- [74] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **109**, 022001 (2024).
- [75] A. Zimmerman, C.-J. Haster, and K. Chatziioannou, *Phys. Rev. D* **99**, 124044 (2019).
- [76] P. Virtanen *et al.*, *Nat. Methods* **17**, 261 (2020).
- [77] M. Isi, K. Chatziioannou, and W. M. Farr, *Phys. Rev. Lett.* **123**, 121101 (2019).
- [78] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, *J. Stat. Softw.* **76**, 1–32 (2017).
- [79] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **102**, 043015 (2020).
- [80] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **896**, L44 (2020).
- [81] S. Roy, A. S. Sengupta, and K. G. Arun, *Phys. Rev. D* **103**, 064012 (2021).
- [82] S. Kastha, A. Gupta, K. G. Arun, B. S. Sathyaprakash, and C. Van Den Broeck, *Phys. Rev. D* **98**, 124033 (2018).
- [83] P. Mahapatra and S. Kastha, *Phys. Rev. D* **109**, 084069 (2024).
- [84] S. Dhanpal, A. Ghosh, A. K. Mehta, P. Ajith, and B. S. Sathyaprakash, *Phys. Rev. D* **99**, 104056 (2019).
- [85] A. Puecher, C. Kalaghatgi, S. Roy, Y. Setyawati, I. Gupta, B. S. Sathyaprakash, and C. Van Den Broeck, *Phys. Rev. D* **106**, 082003 (2022).
- [86] T. Islam, A. K. Mehta, A. Ghosh, V. Varma, P. Ajith, and B. S. Sathyaprakash, *Phys. Rev. D* **101**, 024032 (2020).
- [87] C. D. Capano and A. H. Nitz, *Phys. Rev. D* **102**, 124070 (2020).
- [88] P. Mahapatra, *Phys. Rev. D* **109**, 024050 (2024).
- [89] P. Mahapatra, S. Kastha, A. Gupta, B. S. Sathyaprakash, and K. G. Arun, *Phys. Rev. D* **109**, 064036 (2024).
- [90] E. M. Sanger, S. Roy, M. Agathos, O. Birnholtz, A. Buonanno, T. Dietrich, M. Haney, F.-L. Julie, G. Pratten, J. Steinhoff *et al.*, [arXiv:2406.03568](https://arxiv.org/abs/2406.03568).
- [91] <https://git.ligo.org/lscsoft/lalsuite>.
- [92] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Astrophys. J. Lett.* **970**, L34 (2024).
- [93] E. Hamilton, L. London, J. E. Thompson, E. Fauchon-Jones, M. Hannam, C. Kalaghatgi, S. Khan, F. Pannarale, and A. Vano-Vinuales, *Phys. Rev. D* **104**, 124027 (2021).
- [94] J. E. Thompson, E. Hamilton, L. London, S. Ghosh, P. Kolitsidou, C. Hoy, and M. Hannam, *Phys. Rev. D* **109**, 063012 (2024).
- [95] M. Colleoni, F. A. R. Vidal, C. Garca-Quiros, S. Akay, and S. Bera, *Phys. Rev. D* **111**, 104019 (2025).
- [96] E. Hamilton *et al.*, [arXiv:2507.02604](https://arxiv.org/abs/2507.02604).
- [97] R. Sturani, Note on the derivation of the angular momentum and spin precessing equations in SpinTaylor codes, Technical Report No. LIGO-T1500554-v23, LIGO Scientific Collaboration, 2019.
- [98] C. R. Harris *et al.*, *Nature (London)* **585**, 357 (2020).
- [99] J. D. Hunter, *Comput. Sci. Eng.* **9**, 90 (2007).
- [100] <https://www.gwosc.org/>.
- [101] A. Bohe *et al.*, *Phys. Rev. D* **95**, 044028 (2017).
- [102] LIGO Scientific, Virgo, KAGRA Collaborations, Data release for tests of general relativity with GWTC-3, Technical Report No. LIGO-P2100456, 2021, <https://dcc.ligo.org/P2100456>.
- [103] R. Cotesta, S. Marsat, and M. Purrer, *Phys. Rev. D* **101**, 124040 (2020).