

Tests of General Relativity with Einstein Telescope

Andrea Begnoni,^{a,b,*} **Walter Del Pozzo,**^{c,d} **Matteo Pegorin,**^{a,b,e,*}
Joachim Pomper,^{c,d,*} **Angelo Ricciardone**^{c,d}

^aDipartimento di Fisica e Astronomia “Galileo Galilei”, Università degli Studi di Padova, via Marzolo 8, I-35131 Padova, Italy

^bINFN, Sezione di Padova, Via Marzolo 8, I-35131 Padova, Italy

^cDipartimento di Fisica “Enrico Fermi”, Università di Pisa, Largo Bruno Pontecorvo 3, Pisa I-56127, Italy

^dINFN, Sezione di Pisa, Largo Bruno Pontecorvo 3, Pisa I-56127, Italy

^eMax Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, Potsdam 14476, Germany

E-mail: andrea.begnoni@phd.unipd.it, walter.delpozzo@unipi.it,
matteo.pegorin.1@phd.unipd.it, joachim.pomper@phd.unipd.it, angelo.ricciardone@unipi.it

Abstract. Gravitational wave signals from compact binary coalescences offer a powerful and reliable probe of General Relativity. To date, the LIGO-Virgo-KAGRA collaboration has provided stringent consistency tests of General Relativity predictions. In this work, we present forecasts for the accuracy with which General Relativity can be tested using third-generation ground-based interferometers, focusing on Einstein Telescope (ET) and binary black hole mergers. Given the expected high detection rate, performing full Bayesian analyses for each event becomes computationally challenging. To overcome this, we adopt a Fisher matrix approach, simulating parameter estimation in an idealized observation scenario, which allows us to study large populations of compact binary coalescences with feasible computational efforts. Within this framework, we investigate the constraints that ET, in its different configurations, can impose on inspiral post-Newtonian coefficients, by jointly analyzing events using a Bayesian hierarchical methodology. Our results indicate that ET could in principle achieve an accuracy of $\mathcal{O}(10^{-7})$ on the dipole radiation term and $\mathcal{O}(10^{-3})$ on higher-order post-Newtonian coefficients, for both the triangular and the two L-shaped designs, with 10^4 catalog events. We also assess the number of detections required to confidently identify deviations from General Relativity at various post-Newtonian orders and for different detector configurations.

Keywords: Test of General Relativity, Einstein Telescope, Fisher information matrix, post-Newtonian inspiral deformation coefficients, Bayesian hierarchical model

*Corresponding authors.

Contents

1	Introduction	2
2	Methodology	3
2.1	Model independent parametrization of GR deviations	3
2.2	Hierarchical analysis with Fisher approximation	5
3	Validation of the setup with LVK	7
4	Forecast for Einstein Telescope: Tests of General Relativity	9
5	Forecast for Einstein Telescope: Falsifying General Relativity	12
6	Discussion and Conclusion	13
A	Catalog and detector setup	14
B	The impact of the low frequency cutoff f_{\min}	16
C	Fisher approximated hyper-parameter distribution	17
D	On the role of higher-modes and selection effects for LVK	17

1 Introduction

After a decade of observations [1–7], gravitational waves (GWs) from compact binary coalescences (CBCs) have emerged as a powerful tool for testing general relativity (GR) in the strong-field regime, and highly relativistic regime. There are several well motivated reasons to consider that GR is not the full theory of gravity, especially when trying to reconcile classical general relativity with the fundamental principles of quantum mechanics. Therefore, many beyond-GR theories have been proposed and numerous works have attempted to search for GR deviations in the GW data, observed by the LIGO-Virgo-KAGRA (LVK) collaboration [8–14]. These studies face the difficulty posed by the large number of ways GR deviations could manifest [15–18] and the experimental challenges in detecting these deviations [19]. Moreover, this search is also affected by the presence of possible systematics, non-stationary or non-Gaussian noise [20, 21], environmental effects [22–25], waveform biases [26, 27] or parameter degeneracy and sampling issues [11], which could mimic a deviation from GR. Extensive studies have also scrutinized the nature and close environment of compact objects [28–58]. To date, no GR deviation has been detected by the LVK collaboration [10, 11].

In this work we adopt an agnostic approach, first developed in [59] and employed in the LVK’s TIGER pipeline, to perform tests on deviations from GR [60, 61]. The spirit of this approach, as used in the most recent LVK analysis [10], is to add parametrized GR deviations in the post-Newtonian (PN) coefficients [62, 63] and perform inference by varying one coefficient at a time. These modifications are applied to the inspiral phase of the waveform, followed by adjustments to the phase in the merger and ringdown parts to guarantee overall C^1 -regularity in the frequency domain [64]. Another more recent framework to test GR in the inspiral phase is the Flexible Theory Independent method [11, 65, 66], which uses a taper function to isolate the inspiral deviation, reducing leakage to the merger and ringdown phases. Additional methods for testing deviations from GR are also the parameterized post-Newtonian [19, 67–71], the parameterized post-Keplerian [72, 73] and the post-Einsteinian [74–82] frameworks.

Many beyond GR theories motivate an accurate scrutiny of the PN terms, e.g., higher curvature theories [18, 83, 84], Dynamical Chern–Simons [85, 86], Einstein–Maxwell–Dilaton [87–89] or Einstein–Dilaton–Gauss–Bonnet [90–109], for a review [110], and GWs represent a promising tool to probe such GR deviations [8–10, 111–116]. To improve the constraints on potential GR deviations, we can take advantage of the large number of GW detections expected from the third generation of interferometers. For this reason, we forecast the improvements achievable by hierarchically combining information from many individual GW events, utilizing a Gaussian population model [117]. We show that even small deviations from GR can be probed with Einstein Telescope (ET) [118, 119].

ET is a third generation GW interferometer [120], expected to improve the detector sensitivity with respect to second generation interferometers and to extend the observable horizon; it has been included in the 2021 European Strategy Forum on Research Infrastructures (ESFRI) roadmap. The location and the final layout are still under discussion in the community [119, 121]: possible layouts are a triangular configuration (Δ) or a 2L shape detector network (2L) with the arms either aligned (2L_0) or misaligned by 45 degrees (2L_45). In this work, we compare such different ET designs to estimate their forecasting power to measure GR deviations. To perform the forecast, we employ a Fisher matrix approach to analyze a large catalog of sources with complete inspiral–merger–ringdown (IMR) waveforms, in particular, the Phenom-family waveforms [64, 122–125].

We use waveform models that include not only the fundamental mode but also higher harmonics [122, 123], highlighting the importance of higher-mode waveforms for testing deviations from GR, and focusing on models that incorporate spin effects only for aligned spins. The analysis focuses on Binary Black-Holes (BBH) and involves hundreds of thousands of individual GW events. It is carried out using the `GWJulia` code [126], which allows the evaluation of the individual Fisher matrix, for all GR and beyond-GR waveforms, in a very fast and efficient manner.

The paper is organized as follows. In Sec. 2 we describe the theoretical and computational methodology employed in this work; in Sec. 3 we validate our Fisher–matrix forecasts against observational results from the LVK collaboration [10]. Section 4 presents forecasts for ET’s ability to test GR assuming GR as the fiducial theory, whereas Sec. 5 evaluates ET’s constraining capabilities when

assuming deviations from GR. The discussion and conclusions are presented in Sec. 6. App. A details the catalog and detector setup, App. B examines the impact of the low-frequency cutoff, App. C provides further details on the hierarchical hyperparameter distribution within the Fisher approximation, while App. D examines the impact of higher modes in the waveform model on higher PN order constraints when considering the LVK network.

2 Methodology

In this section, we introduce the formalisms and technical definitions used throughout this work, beginning with the analytical modeling of GW, the PN framework and the Fisher information matrix. Then, we give an in-depth explanation of the Bayesian hierarchical methods that will be employed in the following sections.

2.1 Model independent parametrization of GR deviations

A GW signal h in the fundamental mode can be written as

$$h_{+/\times}(t) = \mathcal{A}_{+/\times}(t) e^{i\Phi(t)}, \quad (2.1)$$

where $\mathcal{A}_{+/\times}$ represent the amplitude, Φ is the phase and $+/\times$ indicate the two GW polarizations. This formulation can be extended to include higher order modes [122, 127]. Conventionally, GW waveforms from CBC are divided in inspiral, merger and ringdown phases [128]. In particular, the inspiral part has an analytic description within the PN framework, which accurately models inspirals of comparable-mass compact objects. In this framework, the phase can be written as a Taylor expansion in the characteristic orbital velocity v as $(v/c)^p$, where c is the speed of light and $p/2$ indicates the PN order. Once one moves to the Fourier domain, within the stationary phase and quasi-circular orbit approximations, the inspiral phase can be expressed as [129–131]

$$\Phi(f) = 2\pi f t_c - \Phi_c - \frac{\pi}{4} + \sum_{p=0}^7 [\varphi_p + \varphi_{p,l} \ln f] f^{(p-5)/3}, \quad (2.2)$$

where φ_p indicates the PN term coefficient, f indicates the frequency, t_c and Φ_c the time and phase of coalescence of the binary, respectively. All the PN terms φ_p with $p \in \{0, \dots, 7\}$, as well as φ_{5l} and φ_{6l} have a known analytic expression, e.g., [131, 132]. Each PN term φ_p in GR is divided into spin-dependent and spin-independent components

$$\varphi_p^{\text{GR}} = \varphi_p^{\text{GR, NS}} + \varphi_p^{\text{GR, S}}. \quad (2.3)$$

In this work, we consider modifications to the phase Φ in the inspiral part of the coalescence, due to beyond GR effects. The deviations to the phase are parametrized by adding the deviation coefficients $\delta\varphi_p$ to each PN coefficient φ_p , where the index p now includes also the log terms. Perturbations may be added to the spin-dependent and spin-independent parts of the PN coefficients separately

$$\delta\varphi_p = \delta\varphi_p^{\text{NS}} + \delta\varphi_p^{\text{S}}. \quad (2.4)$$

We follow the procedure described in [9]; thus, we introduce modifications only to the spin-independent part, leaving the spin-dependent part unaltered, i.e., we assume $\delta\varphi_p^{\text{S}} = 0$. The modification of the PN coefficients then amounts to

$$\varphi_p = (1 + \delta\varphi_p) \varphi_p^{\text{GR, NS}} + \varphi_p^{\text{GR, S}}. \quad (2.5)$$

This implementation as a relative deviation is valid when $\varphi_p^{\text{GR, NS}} \neq 0$. In the other cases $\delta\varphi_p$ represent an absolute deviation from zero. No deviation is introduced for the spin-induced quadrupole moment.

We implement these modifications to two waveform models calibrated to numerical relativity: `IMRPhenomD` [64, 125] and `IMRPhenomHM` [122, 123]. The main difference between the two waveforms is that `IMRPhenomHM` contains higher-order modes, with respect to the fundamental (2,2) mode, being the only mode present in `IMRPhenomD`. These higher-order modes are key for breaking the degeneracy

between the inclination angle and the luminosity distance, e.g., [133], and allow for an overall better parameter estimation [134, 135]. The waveform models **IMRPhenomD** and **IMRPhenomHM** assume aligned spins, thus neglecting precession effects. In principle, precession can be taken into account using a twist-up construction [134], but has been neglected for the Fisher analysis presented here. After the modification of the inspiral parameter in (2.2), phase- and timeshift parameters of the waveform template at the boundaries between the different regimes of the model are adjusted, such that \mathcal{C}^1 -continuity of the waveform is maintained in the frequency domain. The modifications to the inspiral part implemented in the **IMRPhenomD** and **IMRPhenomHM** waveform models do not significantly affect the intermediate and merger-ringdown parts¹.

Following the TIGER approach [59], in the analysis we add a single deviation parameters at a time, while fixing the others to zero. This procedure provides a modified inspiral-merger-ringdown (IMR) waveform model:

$$h_{\varphi_p}[\Theta](f) = h_+^{\varphi_p}[\Theta]F_+ + h_\times^{\varphi_p}[\Theta]F_\times , \quad (2.6)$$

where $F_{+/\times}$ denotes the antenna pattern functions of the detector and $h_{+/\times}^{\varphi_p}$ are the individual polarization contributions for which a deviation parameter was added to the PN coefficient φ_p (or φ_{pl}). We combined the deviation parameter $\delta\varphi_p$ with the other standard waveform parameter θ in an extended vector $\Theta = (\theta, \delta\varphi_p)$. For our analysis we consider 11 GR waveform parameters, reducing the information of the black-hole spins to only the component normal to the plane of motion and thus neglecting a description of precession effects. Hence, in this work, Θ is a vector of the total 12 parameters, see App. A for more information. Considering the following set of potential deviations, this provides 10 IMR waveforms, one for each selected deviation

$$\delta\varphi_p \in \{\delta\varphi_{-2}, \delta\varphi_0, \delta\varphi_1, \delta\varphi_2, \delta\varphi_3, \delta\varphi_4, \delta\varphi_{5l}, \delta\varphi_6, \delta\varphi_{6l}, \delta\varphi_7\} . \quad (2.7)$$

We are not considering φ_5 , since this parameter is degenerate with Φ_c for non-precessing models, as used in this work and evident from (2.2). Furthermore, the PN coefficient for $p = -2$ and $p = 1$ vanish and thus $\delta\varphi_{-2}$ and $\delta\varphi_1$ are parametrized as total deviation defined as

$$\Phi_{\ell m}^p = \frac{3}{128\eta} \tilde{f}^{(p-5)/3} \delta\varphi_p , \quad (2.8)$$

where $\tilde{f} = 2\pi M(1+z)f/m$, with $M(1+z)$ the total mass at detector, η the symmetric mass ratio, z the redshift and (ℓ, m) indicate the spherical harmonics multipoles. A non-vanishing $\delta\varphi_{-2}$ coefficient is typically associated with leading-order gravitational dipole radiation, which can arise in scalar-tensor theories of gravity. Moreover, $\delta\varphi_{-2}$ is the PN deviation parameter most tightly constrained by GW observations. Higher PN terms become relevant later in the inspiral, due to their dependence on higher powers of the orbital velocity, as evident from (2.2). This has relevant consequences, e.g., low mass events result in the best constraints for the low PN terms, due to the large number of cycles observed during their early inspiral phase. Another important aspect is that the low PN orders depend significantly on the low frequency limit f_{\min} of the analysis, which will be lower in next generation detectors, hence resulting in significant improvements, as shown in App. B.

Assuming stationarity and Gaussianity, we describe the detector noise using the (one-sided) power-spectral-density (PSD) in Fourier space $S_n(f)$. Using the noise weighted scalar product of two functions a, b in the frequency domain

$$(a | b) := 2 \int_{f_1}^{f_2} \frac{a(f)b^*(f) + a^*(f)b(f)}{S_n(f)} df , \quad (2.9)$$

we introduce the optimal signal-to-noise ratio (SNR) of a signal template in a given detector as

$$\text{SNR}[\Theta] := \sqrt{(h[\Theta] | h[\Theta])} . \quad (2.10)$$

¹We did not employ waveforms of the **PhenomX** family since the implementation of the deformation coefficients is more subtle [134].

We can then introduce the SNR of the network of detectors as

$$\text{SNR}_{\text{Network}} = \sqrt{\sum_d \text{SNR}_d^2}, \quad (2.11)$$

where the sum is performed over the SNR of the individual detector labeled by d . Note that there are three detectors for a triangular design, which in our treatment are considered to be statistically independent, i.e., the joint likelihood is the product of the three detector likelihoods [136]. In the following, we will approximate the single event posteriors by an expansion in the combined detector SNR, obtaining the Fisher Information Matrix.

2.2 Hierarchical analysis with Fisher approximation

In order to estimate the capability of a detector network to constrain the deviations in the PN coefficients $\delta\varphi_p$, we use a Bayesian hierarchical framework to combine the information $D^{N_{\text{obs}}} \equiv \prod_{k=1}^{N_{\text{obs}}} D_k$ of N_{obs} individual gravitational wave observations with datasets D_k , into an overall posterior for the PN coefficients. These individual events originate from a presumably unknown population of black hole binaries. Following [117], we adopt a minimal information approach, modeling the population distribution of $\delta\varphi_p$ as a Gaussian $\mathcal{N}(\delta\varphi_p | \mu, \sigma)$ with two population hyperparameters μ and σ . As discussed earlier, we consider deviations in PN orders one at a time and the population hyperparameters in principle may differ between the PN orders under consideration. The posterior of $\delta\varphi_p$ can be written as

$$P(\delta\varphi_p | D^{N_{\text{obs}}}, I) = \int d\mu d\sigma \mathcal{N}(\delta\varphi_p | \mu, \sigma) P(\mu, \sigma | D^{N_{\text{obs}}}, I). \quad (2.12)$$

Here, I denotes our general information background in the Bayesian sense. The population posterior $P(\mu, \sigma | D^{N_{\text{obs}}}, I)$ of the hyperparameters is the object of interest, combining the information of N_{obs} individual GW observations. Using Bayes' theorem and assuming that the observations are independent, we obtain

$$P(\mu, \sigma | D^{N_{\text{obs}}}, I) = \frac{P(\mu, \sigma | I)}{P(D^{N_{\text{obs}}} | I)} \prod_{k=1}^{N_{\text{obs}}} \int d\delta\varphi_p P(D_k | \delta\varphi_p, I) \mathcal{N}(\delta\varphi_p | \mu, \sigma). \quad (2.13)$$

The posterior (2.13) is computed by marginalizing over the individual events, as

$$P(D_k | \delta\varphi_p, I) = \int d\boldsymbol{\theta} P(D_k | \boldsymbol{\theta}, \delta\varphi_p, I) P(\boldsymbol{\theta} | \delta\varphi_p, I). \quad (2.14)$$

For forecasting purposes, we apply a linear signal approximation (LSA) to the likelihoods [137] of the individual events, labeled by an index $k = 1, \dots, N_{\text{obs}}$. Such an approximation may be justified by the high SNR of signals detected by next-generation detectors, such as ET. To do so, one expands the likelihood in inverse powers of the optimal SNR of the true signal $h[\boldsymbol{\Theta}_k^0]$ (here $\boldsymbol{\Theta}_k^0$ denotes the fiducial point in parameter space, characterizing the k -th observed event) and then self-consistently truncates the perturbative series, keeping only the lowest order terms [137]. For each event, labeled by the index k , the expansion can be written as

$$\ln P(D_k | \boldsymbol{\Theta}, I) = -\frac{1}{2}(\boldsymbol{\Theta} - \bar{\boldsymbol{\Theta}}_k)^\top \Gamma_k (\boldsymbol{\Theta} - \bar{\boldsymbol{\Theta}}_k) + \mathcal{O}\left(\frac{1}{\text{SNR}[\boldsymbol{\Theta}_k^0]}\right), \quad (2.15)$$

where we omit terms independent of $\boldsymbol{\Theta}$, since they can be absorbed in the normalization. After truncation at the lowest order, the resulting likelihood is completely determined by the Fisher information matrix

$$\{\Gamma_k\}_{ij} = \left(\vec{\nabla}_{x_i} h[x_1, x_2, \dots] \Big| \vec{\nabla}_{x_j} h[x_1, x_2, \dots] \right) \Big|_{\mathbf{x}=\boldsymbol{\Theta}_k^0} \quad (2.16)$$

and the maximum of the approximated likelihood

$$\{\bar{\Theta}_k\}_i = \left(n_k \left| \vec{\nabla}_{x_i} h[x_1, x_2, \dots] \right. \right) \Big|_{\mathbf{x}=\Theta_k^0}. \quad (2.17)$$

In the last expression, n_k denotes an explicit noise realization of the detector. Hence, $\bar{\Theta}_k$ is a random variable and depends on the noise realization, while the Fisher matrix is independent of the noise. When combining multiple events within the hierarchical framework, the overlap, and thus the center, of the individual likelihoods is of great importance [138]. In fact, in the limit of large N_{obs} , the results turn out to be heavily biased when the Fisher-approximated posteriors are always centered around the values Θ_k^0 . This highlights the importance of correctly accounting for the noise-dependent shift of the maximum of the likelihoods. For technical implementations, we will only need to calculate the Fisher matrices Γ_k of the individual events, since in the LSA the integrals in (2.13) and (2.14) can be performed analytically. In the following, we will assume

$$P(\boldsymbol{\theta} | \delta\varphi_p, I) = \Pi(\boldsymbol{\theta} | I) = \text{const.} \quad (2.18)$$

While the independence of the distribution of potentially known deviations could be physically reasonable, we stress that the assumption of an entirely flat single event prior $\Pi(\boldsymbol{\theta} | I)$ is purely a technical tool and certainly does not reflect what is done in a full Bayesian analysis. Since we are concerned with combining a large number of informative events, we can also put a flat prior $P(\mu, \sigma | I)$ for μ and σ . In the limit of large N_{obs} we can safely assume the resulting posterior to be likelihood driven. If we define

$$\Delta_k := \{\Gamma_k^{-1}\}_{\delta\varphi_p \delta\varphi_p} \quad \text{and} \quad \delta\bar{\phi}_k := \{\bar{\Theta}_k\}_{\delta\varphi_p}, \quad (2.19)$$

we then obtain the posterior distribution for the two population hyperparameters μ and σ [139]

$$P(\mu, \sigma | D^{N_{\text{obs}}}, I) \propto \prod_{k=1}^{N_{\text{obs}}} \sqrt{\frac{\Delta_k^2}{\Delta_k^2 + \sigma^2}} \exp\left(-\frac{1}{2} \frac{(\mu - \delta\bar{\phi}_k)^2}{\Delta_k^2 + \sigma^2}\right), \quad (2.20)$$

which was the initial object of interest (2.13), encoding all the relevant information about the distribution of the PN deformation coefficients $\delta\varphi$, under the minimal information assumption. The quantities $\delta\bar{\phi}_k$ can be shown to behave as Gaussian random variables

$$\delta\bar{\phi}_k = \{\Theta_k^0\}_{\delta\varphi_p} + \Delta_k \mathcal{X}, \quad (2.21)$$

with $\mathcal{X} \sim \mathcal{N}(0, 1)$ a unit Gaussian random variable [137]. Having computed the Fisher matrix, we can account for different detector noise realizations by sampling from a standard normal and adding the corresponding shifts to the fiducial value.

The hierarchical distribution includes certain limiting scenarios of the population distribution, such as the limit of $\sigma \rightarrow 0$. For this particular subspace in the population parameter space, the PN deformation coefficient $\delta\varphi_p$ takes on the same value μ for all individual events².

From the hierarchical distribution, we may obtain the distribution of the deviations via (2.12). The integral over μ can again be performed analytically, while the integral over σ must be evaluated numerically. We obtain

$$P(\delta\varphi_p | D^{N_{\text{obs}}}, I) \propto \int_0^\infty d\sigma \frac{e^{c(\sigma)}}{\sqrt{1 + a(\sigma)\sigma^2}} \exp\left(-\frac{1}{2} \frac{a(\sigma)\delta\varphi_p^2 - 2b(\sigma)\delta\varphi_p - b^2(\sigma)\sigma^2}{1 + a(\sigma)\sigma^2}\right), \quad (2.22)$$

where we define the auxiliary functions

$$a(\sigma) = \sum_{k=1}^{N_{\text{obs}}} \frac{1}{\Delta_k^2 + \sigma^2}, \quad b(\sigma) = \sum_{k=1}^{N_{\text{obs}}} \frac{\delta\bar{\phi}_k}{\Delta_k^2 + \sigma^2}, \quad (2.23)$$

$$c(\sigma) = \frac{1}{2} \sum_{k=1}^{N_{\text{obs}}} \left(\ln\left(\frac{\Delta_k^2}{\Delta_k^2 + \sigma^2}\right) - \frac{\delta\bar{\phi}_k^2}{\Delta_k^2 + \sigma^2} \right). \quad (2.24)$$

² μ is still a function of PN order. We omit the index p in the notation.

For the Fisher approximation, we need to fix a fiducial point around which the series expansion is performed. For the forecast of consistency tests assuming GR as the null hypothesis, $\delta\varphi_p^0 = 0$ is the fiducial point for every event of the catalog. In this special case, the null-hypothesis is contained within the $\sigma \rightarrow 0$ subspace of the population parameter space. We thus consider also the deviation distribution

$$\mathring{P}(\delta\varphi_p | D^{N_{\text{obs}}}, I) = \lim_{\sigma \rightarrow 0} \int d\mu \mathcal{N}(\delta\varphi_p | \mu, \sigma) P(\mu, \sigma | D^{N_{\text{obs}}}, I) \quad (2.25)$$

$$\propto \prod_{k=1}^{N_{\text{obs}}} \exp\left(-\frac{1}{2} \frac{(\delta\varphi_p - \delta\bar{\phi}_k)^2}{\Delta_k^2}\right) \quad (2.26)$$

conditioned to the hyperparameter subspace, for which the deviation manifests the same among all individual events, by replacing the flat prior on σ with a δ -function conditioned on $\sigma = 0$. In this case, the combined deviation distribution reduces to the product of the individual deviation distributions. Clearly, the conditioned distribution $\mathring{P}(\delta\varphi | D^{N_{\text{obs}}}, I)$ is always narrower than the full hierarchical deviation distribution (2.22). We refer to Appendix C for a further comparison. Thus, combined with the omission of the information in the tails of the posterior due to the Fisher-approximation, we use the conditioned distribution $\mathring{P}(\delta\varphi | D^{N_{\text{obs}}}, I)$ to estimate lower bounds of our capability to constrain deviations in the PN coefficients with future detectors. To do so, we calculate the 90% credible upper bound of the deviation posterior $\mathring{P}(\delta\varphi_p | D^{N_{\text{obs}}}, I)$, defined as the absolute value of the PN deviation coefficient

$$|\delta\varphi_p| \quad \text{such that} \quad \int_{-|\delta\varphi_p|}^{+|\delta\varphi_p|} \mathring{P}(\delta\varphi_p | D^{N_{\text{obs}}}, I) = 0.9 \quad (2.27)$$

holds.

3 Validation of the setup with LVK

To validate our full analysis, before applying it to the case of Einstein Telescope, we perform an estimation of the constraining power of a GW observation run, such as the O3b observing run of the LVK collaboration [10]. This is intended to provide verification of our software implementation, as well as for the methods presented in Section 2.

To do so, we set up a detector network composed of three detectors: LIGO Hanford (LHO), LIGO Livingston (LLO) [140], and Virgo (V) [141]. The network composed of these three detectors will be referred to as LHV. We compute the conditioned posterior $\mathring{P}(\delta\varphi_p | D^{N_{\text{obs}}}, I)$ in (2.26) from a total of 9 observed events, drawn from a catalog of BBHs and fiducial expansion point $\varphi_p^0 = 0$. From the combined deviation distribution, we estimate the upper bounds on $|\delta\varphi_p|$. For this, an event is defined as observable if it satisfies the following three criteria: i) The event has a network SNR larger than 12, ii) the network SNR contribution from the inspiral is larger than 6, iii) the Fisher information matrix of the event is invertible. The second criterion has a two-fold purpose. On the one hand, it has been adopted to mimic the selection of events in the analysis by the LVK collaboration [10], allowing for a consistent comparison. On the other hand, it guarantees that the signals taken into account in the analysis have a significant amount of SNR in the inspiral portion of the signal. This seems important to avoid potential systematics, given that the parametrization of GR deviation was introduced via phase modifications during the inspiral. The inspiral portion of the signal, used for the second criterion, is defined via the inspiral cutoff frequency $f_{\text{insp,cutoff}}$, defined within the **Phenom** waveform template family [125]. The third criterion is purely technical, ensuring the applicability of the Fisher-approximation. The number of events that do not pass the third criterion depends on the waveform and network used; however, it reached at most 1% of events. Contrary to what will be done for ET later, we adopt a frequency cutoff of $f_{\text{min}} = 20$, in order to be able to compare our result with the measurements performed by the LVK. The minimum analysis frequency has a large impact on the final constraints of the PN deviations, in particular at lower PN orders. Further details are discussed in Appendix B. Details about the generation of the catalog and the used PSDs for the setup of the detectors H, L and V can be found in Appendix A.

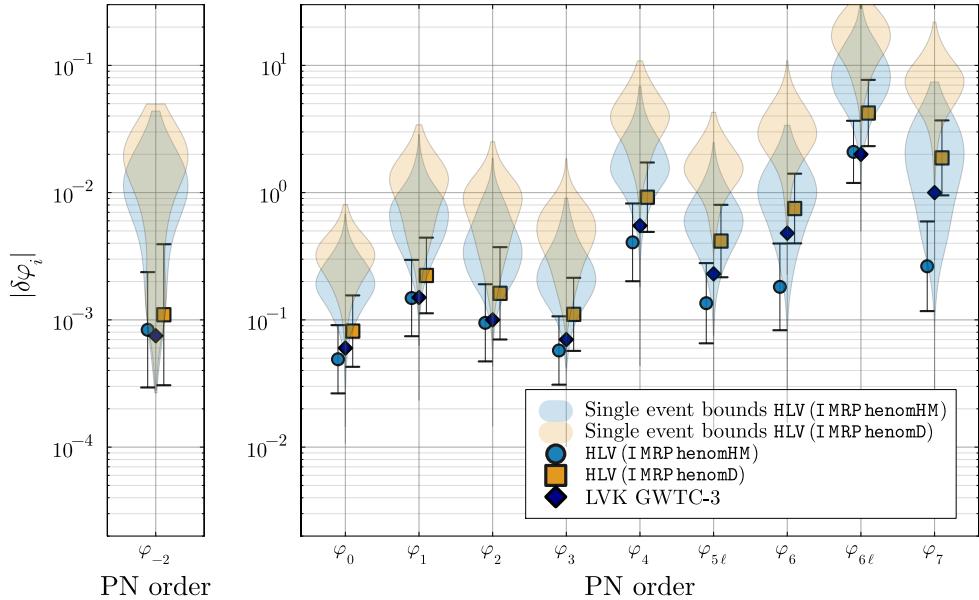


Figure 1: Forecast of the 90% upper bounds on the magnitude of the restricted PN deformation coefficients $|\delta\varphi_p|$ for a LVK O3b-like observing run, for the different PN orders considered. The bounds are obtained from the distribution $\hat{P}(\delta\varphi_p | D^{N_{\text{obs}}}, I)$ in (2.26), assuming that deviations take on the same value for all observed events. The label φ_p refers to the $\frac{p}{2}$ -PN order coefficient, $\varphi_{p\ell}$ refers to the $\frac{n}{2}$ -PN order coefficient with the log term. Each marker and its corresponding error bar represent the mean and 90% confidence interval, for $|\delta\varphi_p|$ calculated for $\mathcal{O}(50)$ independent realizations of the observing run and different noise realizations. The color-coded violin plots show the distribution of the single event 90% upper bounds on $|\delta\varphi_i|$. We show the results of the forecast employing two different waveform models: IMRPhenomHM and IMRPhenomD. For comparison, we report the results obtained by the LVK collaboration, performing a full Bayesian analysis with the SEOBNRv4_ROM waveform template on the data of 9 observed real events, see Figure 6 in [10]. The present FIM analysis provides forecast in good agreements with the GWTC-3 results.

Using our setup, specified above, allows for a reasonable comparison with LVK’s O3b analysis, which employed a full Bayesian analysis applied to the data of 9 real observed events. Figure 1 shows a comparison of upper limits on the deviation parameters: our estimates based on IMRPhenomD and IMRPhenomHM compared to the LVK constraints from real data using SEOBNRv4_ROM [10]. The two are in good agreement, despite the known limitations of the Fisher–matrix approximation. The colored violin plots in the Figure show the distribution of the 90% upper bounds of $|\delta\varphi_p|$, as obtained from individual observable events drawn from the catalog. The LVK results quoted represent the upper bounds on PN deviation parameter, obtained with the SEOBNRv4_ROM waveform template [10]. For these specific results, also the LVK analysis adopted a δ -prior for σ , enforcing any GR deviation to manifest with the same value in all individual observations. In order to quantify the fluctuations of our estimate of the upper bounds on $|\delta\varphi_p|$, obtained from the distribution $\hat{P}(\delta\varphi_p | D^{N_{\text{obs}}}, I)$ in (2.22), we produce $\mathcal{O}(50)$ realizations.³ Repeating the calculation this way multiple times allows us to plot the fluctuation of the lower bound [138], visualized as error bars in Figure 1, representing the 90% confidence intervals around the estimated mean value.

A comparison of the forecasts employing IMRPhenomD vs. IMRPhenomHM suggests that the inclusion of higher modes in the analysis results in a $\mathcal{O}(2)$ improvement in the bounds, consistent with results

³We generate a large catalog of sources, then we assemble sub-catalogs of 9 detected events from the detected population, leading to $\mathcal{O}(50)$ realizations of O3b-like observing runs.

from full Bayesian parameter estimation, e.g., [65, 134]. The improvements on the constraints for $\delta\varphi_{5l}$, $\delta\varphi_6$, and $\delta\varphi_7$ exceed the expectations inferred from individual Bayesian analyses showing pronounced tails, which apparently indicate tighter bounds. Nevertheless, it is well known that these higher-order deviation coefficients can exhibit pathological or unphysical behavior even within full Bayesian frameworks [9]. Therefore, the potential overestimation of the constraints observed in our simplified Fisher analysis is not entirely unexpected. The pronounced tails observed in our analysis originate from the contribution of high-mass binary systems, whose signals enter the detector's sensitivity band only during the late inspiral phase. Such events involve additional systematic uncertainties and should be interpreted with some caution, as the deviations from general relativity were introduced only in the inspiral portion of the signal. Moreover, the population mass distribution with a peak at $M_{\text{peak}} \approx 34M_\odot$ further enhances the prevalence of such high-mass sources. A more detailed discussion is provided in Appendix D. Overall, it appears that the analysis benefits from the inclusion of higher modes [65, 134].

4 Forecast for Einstein Telescope: Tests of General Relativity

After validating our method, we proceed to forecast how precisely ET can constrain deviations in the PN coefficients using BBH observations. For this analysis, we set the fiducial expansion point of the Fisher approximation to the PN coefficient ϕ_p predicted by GR, i.e., $\delta\phi_p^0 = 0$, for all individual observations.

Given that the final ET design is not decided yet, we consider and compare three different detector configurations: a single 10km triangular configuration Δ , a network of two L-shaped 15km misaligned detectors 2L_45, and a network of two L-shaped 15km aligned detectors 2L_0. Further details on the detector setup, including the employed PSD, are provided in Appendix A.

Given our observations in the previous section, we restrict to the `IMRPhenomHM` waveform template. Thanks to its increased sensitivity, ET should more effectively exploit the physical information encoded in higher-order modes. For the ET analysis, we use a lower frequency cutoff $f_{\min} = 2\text{Hz}$. For a discussion on the impact of the choice for f_{\min} see Appendix B. Given ET's improved sensitivity, the number of events observed is expected to increase significantly. As a benchmark, during an observation run of about 3 to 4 months, we expect to have $\mathcal{O}(10^4)$ BBH sources [126, 133]. Hence, for our forecast, we generated a catalog of 10^4 BBH events, and by calculating the SNR and Fisher matrix, we selected the number of observable events among these. In this analysis, we drop the criterion based on the inspiral portion of the SNR, employed in Section 4. We verified that its inclusion does not affect the obtained results, suggesting that signals detected by ET generally show a sufficiently large inspiral SNR to robustly support this class of inspiral-based tests. Hence we define an event as observable if it satisfies the following two criteria: i) The event has a network SNR larger than 12, ii) the Fisher information matrix of the event is invertible. Note that the subset and number of observed events, selected from these 10,000 events in the catalog, depend on the detector configuration. For the details on the catalog, see Appendix A. In particular, 69.6% of events in the catalog for the Δ -configuration, 81.9% for the 2L_45-configuration and 76.8% for the 2L_0-configuration were used to compute the hierarchical posterior. We find that only very few events, of order $\sim 1\%$, have been neglected based solely on a non-invertible Fisher matrix. We emphasize that this procedure differs from the one in Section 3 and corresponds to calculating the constraints based on a fixed window of observation time, rather than a fixed number of observed events. However, this allows us to draw meaningful conclusions when comparing different detector designs, as done in this section.

Since we are performing the forecast for a GR consistency test, hence assuming GR as the null-hypothesis, we compute the combined 90% credible upper bound $|\varphi_p|$ from the conditioned deviation distribution $\dot{P}(\delta\varphi_p | D^{N_{\text{obs}}}, I)$ in (2.26). In appendix C, we further provide plots of the hierarchical hyperparameter distribution $P(\mu, \sigma | D^{N_{\text{obs}}}, I)$ in (2.20) and estimates from the hierarchical deviation distribution $P(\delta\varphi_p | D^{N_{\text{obs}}}, I)$ in (2.22). The discussion of these results is deferred to the appendix since the main conclusions and most optimistic constraints on the improvements with ET can be inferred by using $\dot{P}(\delta\varphi_p | D^{N_{\text{obs}}}, I)$. To quantify the statistical uncertainty of our forecast, due to specific catalog and detector noise realizations, we repeat the calculation for 20 different catalog

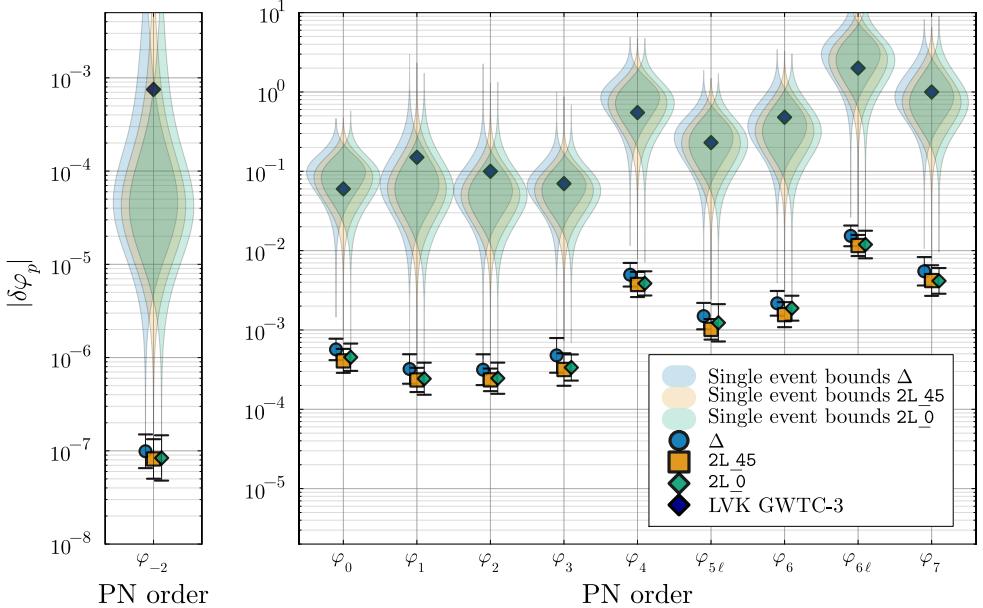


Figure 2: Forecast for the 90% upper bounds on the magnitude of the post-Newtonian inspiral deviation coefficients $\delta\varphi_i$ for the three different ET configurations considered: triangular Δ , two L-shaped interferometers with 45° misalignment 2L_45, two aligned L-shaped interferometers 2L_0. The bounds are obtained from the distribution $\hat{P}(\delta\varphi_p \mid D^{N_{\text{obs}}}, I)$ in (2.26), assuming that deviations take on the same value for all observed events, and are represented by three different markers, respectively. The estimate used $\mathcal{O}(8000)$ events observed by each detector configuration in a fixed period of approximately 4 months. Each marker represents the average upper bound, obtained by averaging over 20 independent realizations of the simulated observations, while the error bars represent the 90% confidence intervals of the mean. The color-coded violin plots show the distribution of the single event 90% upper bounds on $|\delta\varphi_i|$. For comparison, as in the previous Figure, the blue diamonds represent the current GWTC-3 constraints [10].

realizations with 10^4 events and different noise realizations. We depict the results of our forecast of the 90% credible upper bound $|\varphi_p|$ of GR deviations for ET in Figure 2. The marker in Figure 2 represents the mean value over the 20 realizations, while the error bars mark the 90% confidence interval, quantifying the statistical fluctuation of our forecast.

The combined 90% credible upper bounds from ET for BBH events, after an observation time of about four months, present an improvement between 2 to 4 orders of magnitude with respect to the GWTC-3 results [10], with larger improvements for lower PN orders. These improvements can be primarily attributed to three factors: ET's extended frequency range, down to $f_{\min} = 2$ Hz, its enhanced sensitivity across the entire frequency range, especially the region already accessible to the HLV network, and the increased statistical power from a larger number of detected events. To estimate the improvement due to increased detector sensitivity, we plot the distributions of the estimated individual single event bounds, depicted as violin plots in Figure 2. The largest improvement for single events occurs at lower PN orders, primarily due to the enhanced low-frequency sensitivity of ET. For $\delta\varphi_{-2}$, this results in an improvement of two orders of magnitude, while for $\delta\varphi_0$ through $\delta\varphi_4$ the bound is refined by about one order of magnitude. The contributions of the higher-order PN coefficients are entering only at higher frequencies and are less affected, see also Figure 6 in Appendix B.

To assess the improvements of the upper bounds due to the increased statistics, given the large number of expected detections, in Figure 3 we plot the bounds as a function of the number of detected

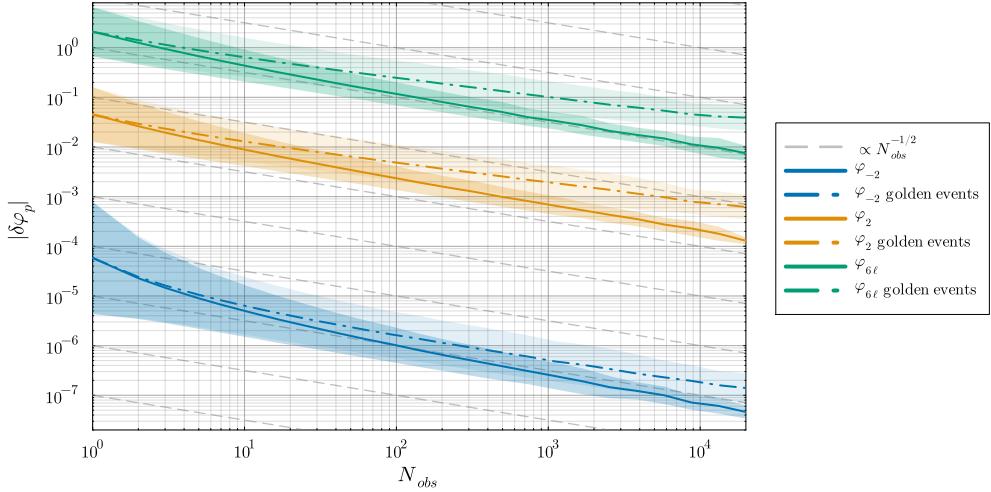


Figure 3: Plot showing the 90% credible upper bounds on the magnitude of the post-Newtonian inspiral deformation coefficients N_{obs} . The solid lines show the average combined upper bounds on $|\delta\varphi_p|$ obtained from the distribution $\hat{P}(\delta\varphi_p | D^{N_{\text{obs}}}, I)$ in (2.26). The average is taken over several catalogs and noise realizations, with the faint color band representing the resulting 90% confidence intervals. The dash-dotted lines show the average value of the tightest single event upper bounds on $|\delta\varphi_p|$, among N_{obs} events drawn from the catalog. The fainter colored bands quantify the corresponding 90% confidence intervals of the best single event constraints, obtained by drawing multiple catalog subsets of size N_{obs} and identifying the best bound. The detector was fixed to ET in the 2L_45 configuration.

events, comparing the upper bounds obtainen from the distribution $\hat{P}(\delta\varphi_p | D^{N_{\text{obs}}}, I)$ to the tightest 90% credible upper bounds obtained from a single golden event, as a function of N_{obs} observed events within a catalog realizations. For clarity, we fix the 2L_45 ET detector configuration and show only the -1PN, 1PN, and 3PN log terms. We observe that for all PN orders the combined constraint yields tighter bounds with respect to the single event ones, with the relative gain increasing with the number of observed events N_{obs} , as expected. We find that for $N_{\text{obs}} = 10^4$, the average relative improvement between the two bounds is of $\mathcal{O}(3 - 5)$. The combined bounds, obtained from $\hat{P}(\delta\varphi_p | D^{N_{\text{obs}}}, I)$, will further improve with more observed events. The asymptotic scaling can be asymptotically approximated using the central limit theorem, i.e., $|\delta\varphi_p| \propto N_{\text{obs}}^{-1/2}$. The scaling of the single events bounds, on the other hand, can be approximated asymptotically using the probability of observing a close-by high SNR golden event [142], which results in $|\delta\varphi_p|_{\text{golden}} \propto N_{\text{obs}}^{-1/3}$. It appears this asymptotic scaling becomes accurate for most PN orders already after a few hundred observed events, apart from the $\delta\varphi_{-2}$ order, which exhibits a steeper power scaling. The faint color bands, quantifying the variance of the upper bound, due to different catalog and noise realizations, are narrower for the combined bounds than for the ones obtained from the best single event. Considering the scaling of the bounds on the PN deviation coefficients with observation time, $|\delta\varphi_i|(t_{\text{obs}}) \propto t_{\text{obs}}^{1/2}$, we expect that after some years of observations, the bound on the deviation coefficient at 0PN order ($\delta\varphi_0$) from ET should be comparable to those from double pulsars [112, 143–154]. To accurately compare the ET and the double pulsars bounds on the -1PN order ($\delta\varphi_{-2}$), the present analysis should be extended to also include binary neutron star systems [112, 114, 155].

Regarding the constraints from ET, the three configurations considered yield quite similar and consistent results, with the 2L_45 setup showing some slightly tighter bounds. The observed differences among them remain well within the expected statistical scatter due to catalog realizations and noise fluctuations. This overall similarity could be an indication that the estimated BGR deviations, being related to intrinsic parameters, are quite insensitive to the specific detector geometry, as also noted in

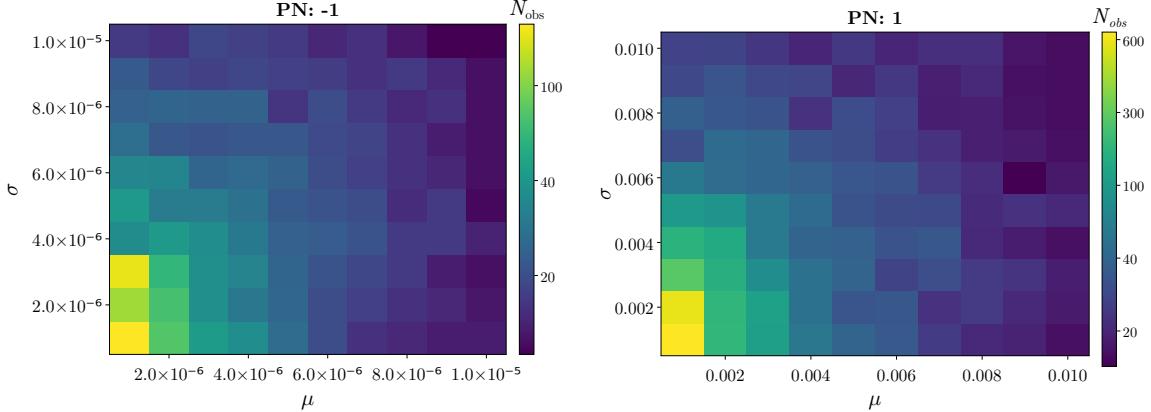


Figure 4: Number of detections needed to claim a deviation from GR with 90% confidence level, given fiducial deviation parameter $\delta\varphi_p^0$, drawn from a Gaussian with mean $\mu_{\delta\varphi_p}^0$ and standard deviation $\sigma_{\delta\varphi_p}^0$. We plot the results for $\delta\varphi_{-2}$ on the left and for $\delta\varphi_2$ on the right, for the 2L_45 configuration. The numbers plotted are the median of 150 repetitions of the forecast, with different catalogs and noise realizations.

other papers (e.g., [126, 133]). As a side note, the presence of correlated noise between the detectors in the triangular configuration can have a non-negligible impact that will be important to quantify [156–158].

5 Forecast for Einstein Telescope: Falsifying General Relativity

In the previous two sections, our analysis focused on forecasting consistency tests of GR. Thus, we set the fiducial point of the Fisher expansion to $\delta\varphi_p^0 = 0$ for all individual events and mostly used the conditional deviation distribution (2.26). In the following, however, we assume General Relativity to not represent the true theory of gravity, and we are interested in forecasting the expected number of observed events, showing a small deviation away from GR, required to detect a significant deviation from GR.

In general such deviations may manifest in several ways, possibly also depending on any of the compact binaries’ parameters in non-trivial ways [17, 18]. Due to the multitude of available theoretical models and the lack of experimental preference, we chose a minimal and self-consistent approach for our study. To this end, we model the deviation parameter following a Gaussian population distribution, specified by two fiducial population hyperparameters $\mu_{\delta\varphi_p}^0$ and $\sigma_{\delta\varphi_p}^0$. The fiducial points for the individual Fisher approximated single event posteriors will then be drawn from a Normal distribution, i.e.

$$\delta\varphi_p^0 \sim \mathcal{N}\left(\mu_{\delta\varphi_p}^0, \sigma_{\delta\varphi_p}^0\right). \quad (5.1)$$

Hence, our analysis amounts to estimating the number of observed events required to recover the injected hyperparameter. We consider a deviation as significant if the fiducial deviation hyperparameters are recovered within the 90% region of the hierarchical hyperparameter posterior, while the volume enclosed by the iso-surface containing GR is larger than 90% [159]. This guarantees an exclusion of GR at the 90% level. For this analysis, we utilize the full hierarchical hyperparameter distribution discussed in section 2.2.

In Figure 4, we show the number of detected events required to claim such a 90% CL deviation of GR injecting fiducial hyperparameters $(\mu_{\delta\varphi_p}^0, \sigma_{\delta\varphi_p}^0)$ for a fixed PN order and 2L_45 as detector configuration. To estimate the minimum number of events required to claim a deviation from GR, we employ a bisection algorithm, repeating the hierarchical evaluation for different numbers of events.

Notably, such an estimate is very sensitive to the catalog and noise realization. Furthermore, individual golden events providing particularly tight constraints on deviations from GR may dominate the overall estimate. Thus, to mitigate these statistical fluctuations, we produce 150 catalog realizations and plot the median number of events required to claim the 90% deviation from GR. From Figure 4 we see the significant improvements that ET will bring, being able to constrain deviations of $\mathcal{O}(10^{-6})$ for $\delta\varphi_{-2}$ and $\mathcal{O}(10^{-3})$ for $\delta\varphi_2$ with only a few hundreds of observed events, i.e., within days of observations. This aligns with the results of Section 4: given the same number of events, the better sensitivity of ET will result in tighter bounds with respect to LVK. This effect is further compounded with the higher statistic due to the increased detection rate, such that few days of data taking will suffice to improve the LVK bounds. In particular, considering the -1PN order, only $\mathcal{O}(20)$ events are required to exclude $(\mu_{\delta\varphi_p}^0, \sigma_{\delta\varphi_p}^0) = (0.01, 0.01)$ while $\mathcal{O}(600)$ events are required to exclude $(\mu_{\delta\varphi_p}^0, \sigma_{\delta\varphi_p}^0) = (0.001, 0.001)$, which corresponds to one week of data taking.

6 Discussion and Conclusion

Testing General Relativity remains a primary objective of current and next-generation gravitational-wave detectors. Since the first detection of GW150914, and especially with GW170817, GW observations [1–4, 6] have enabled a wide range of tests of GR, including measurements of the speed of GWs, constraints on spin-induced quadrupole moments, and bounds on the number of polarizations [8–10, 111–113].

In this work, we address, for the case of ET, a central objective of the tests of GR: evaluating the consistency of the inspiral phase of compact binary coalescences with GR predictions. Modeling the inspiral via a PN expansion, we employ the TIGER framework to forecast the ET’s sensitivity to deviations in the PN coefficients.

We employed a Fisher-matrix approach, implemented with the `GWJulia` code. First, to validate our framework, we compared our predictions with the GWTC-3 results reported by the LVK Collaboration, finding very good agreement with the published data (see Figure 1). Then, we performed forecast analyses for ET, estimating the value of future bounds on potential deviations from GR, with the fiducial point of the Fisher approximation given by the PN coefficients, assuming GR as the fiducial model.

We analyzed a catalog of ten thousand binary black hole mergers, corresponding to roughly three to four months of observations, for each of the ten PN terms, considering three different ET configurations. The analysis employed `IMRPhenomHM` waveform model, which includes higher modes. Alternatively, it would have been possible to use the effective-one-body (EOB) waveforms [127, 160] within the FTI framework [10, 11, 65]; however, such a choice is expected to have only a limited impact on our results [10, 134]. We find that, after a few months of observation, each ET configuration results in tighter constraints on all PN coefficients by two to three orders of magnitude relative to current bounds, with the notable exception of the -1PN term, for which the improvement reaches four orders of magnitude. This significant improvement stems from the higher sensitivity of ET at lower frequencies, which enables more precise measurements of the early inspiral phase. This, in turn, underscores the importance of mitigating seismic noise and advancing cryogenic technology [119, 121]. Furthermore, several years of observations should result in ET upper bounds for the 0PN order that could be competitive with double pulsars constraints.

Our analysis indicates that the several detector configuration considered all provide comparable bounds. The `2L_45` design provides slightly better bounds across all PN orders, yet the difference is smaller than the variance due to catalog and detector noise realization.

Furthermore, we carried out an agnostic test to assess the constraining capabilities of ET in presence of real deviations from GR. By injecting GR deviations with magnitudes drawn from a Gaussian distribution, we quantified ET’s ability to recover such deviations. Our results indicate that even a few days of observations would already exclude large regions of the beyond-GR parameter space.

The present analysis highlights the scientific potential of ET, demonstrating its capability to provide precise tests of GR, and to significantly tighten current constraints when implementing the

hierarchical Bayesian method. The efficiency and flexibility of our implementation, enabled by the `GWJulia` code, were key to achieving the forecast results presented here. Our analysis also presents some approximations and limitations. While the Fisher matrix proved reasonably accurate in our validation, a full Bayesian study with more complete waveform models [25, 134, 160–171] would provide more faithful estimates at the expense of high computational costs, which makes an analysis like this infeasible for now with a full Bayesian study. The source catalog could be extended by including binary neutron stars and neutron star-black hole binaries. Finally, future work should assess the impact of non-stationary or non-Gaussian detector noise, glitches, overlapping signals, waveform systematics, possible correlations between the GR deviations and astrophysical parameters [23, 27, 172, 173], as well as correlated noise in the triangular configuration [156, 158].

Promising directions for future work include developing a more refined parameterization of beyond-GR effects. For instance, one could introduce explicit dependencies of the post-Newtonian deviations on binary parameters such as the mass ratio or total mass, or apply principal component analysis to capture correlated deviations more effectively [17, 18, 174]. Further extensions may involve testing broader classes of modified gravity scenarios, such as theories with massive gravitons [175–177]. A natural next step is to apply this framework to binary neutron star or black hole–neutron star systems, where the higher event rates [126, 178] and longer inspiral durations can provide tighter constraints, particularly at the low PN orders [112]. Overall, the framework introduced here is highly flexible and can be straightforwardly extended to probe more complex and theory-specific deviations from GR. It thus provides a robust and computationally efficient foundation for forecasting tests of general relativity across diverse detector networks and astrophysical populations.

Acknowledgments

M.P. wishes to thank Nicola Bartolo, Alessandra Buonanno, Elisa Maggio, Pierpaolo Mastrolia, Lorenzo Pompili, Elise M. Sänger, Jan Steinhoff and Sebastian Völkel for useful comments and insightful discussions. We thank Soumen Roy and Jan Steinhoff for valuable comments and feedback on the manuscript. We thank Yanbei Chen for orchestrating the LVK internal review. AB is supported by ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union – NextGenerationEU. The work of M.P. is supported by the European Union under the Next Generation EU programme. M.P. gratefully acknowledges financial support from Fondazione Ing. Aldo Gini, from the INFN initiatives *Amplitudes* and *InDark*, and the hospitality and support of the Albert Einstein Institute. The PhD fellowship of J.P. is funded by the Italian Ministry of University through the project ‘Nano-Meta-Materials and Devices: New Frontier Concepts for Particle and Radiation Detection’ (Grant ‘Dipartimento di Eccellenza’ 2023-2027, CUP I57G22000720004) at the Department of Physics of the University of Pisa. CloudVeneto is acknowledged for the use of computing and storage facilities. This material is based upon work supported by NSF’s LIGO Laboratory which is a major facility fully funded by the National Science Foundation.

Data/code availability statement

The code/data used to conduct/produced during the presented study is available from the corresponding authors on reasonable request.

A Catalog and detector setup

This appendix provides more details on some important quantities for our analysis. In particular, we introduce and characterize the properties of the detectors and we comment on the catalog properties. Each detector is determined by its shape, position on Earth, orientation, and the PSD. The shape, i.e., the geometry, is defined through the opening angle between the arms ζ , which is 90° for L-shaped and 60° for triangular detectors. The latitude λ and longitude φ define the position of the center of the detector, i.e., the location of the beam-splitter for L-shaped and the center of the triangle for triangular detectors. Then, the configuration is completely described with the orientation angle, representing

a rotation in the detector plane, which is particularly important for a network of detectors. Each detector has an orientation γ w.r.t. a given direction, which in this work (and in **GWJulia**) we choose w.r.t. the local East. The current network proposals for ET contemplate [119, 121] a triangular design or 2L networks, and, in the latter case, there is also the freedom of the orientation between the two detectors, β . The optimal configuration to measure CBC corresponds to two detectors misaligned with an angle of $\beta_{\text{CBC}} = \pi/4 + n\pi/2$, where n is an integer⁴. Once the curvature of the Earth is considered, the detectors are not coplanar, so the orientation between the two detectors is not β [179, 180] and we define it as $\alpha = \gamma_2 - \gamma_1$. This way, we obtain $\alpha = \beta + 2.51^\circ$ for two detectors, one in Sardinia, one in the MR region. All the following considerations apply to β , which is the physical parameter in the FIM evaluation. Finally, the PSDs used for the comparison with GTWC-3 [6, 10] are the ones corresponding to the O3b run⁵ [182–185] and in fig. 5 we show the amplitude spectral densities, i.e., the square root of the PSD.

Since ET’s final design and configuration have not been decided yet [119, 121], we consider different setups to compare their capabilities in detecting and estimating the parameters of CBCs. In particular, we consider:

- Δ : triangular ET with 10 km arms in Sardinia
- $2L_0$: two 15 km L-shape interferometers, one in Sardinia, one in the MR Euroregion. The orientations of the two detectors are chosen such that $\beta = 0^\circ$.
- $2L_45$: same as $2L_0$ with the exception that the orientations lead to $\beta = 45^\circ$. This is expected to be the best network for measuring BGR parameters.

All the ET networks have cryogenic technology, improving the sensitivity to lower frequencies. We then consider two PSDs [121, 186], one for a detector arm length of 10km for the triangular design and one of 15km arm length for the L-shape. In this work, and in **GWJulia**, we also consider that the beam splitters presented in the triangular configuration have slightly different positions. We also include the effects caused by the Earth’s rotation, which are supposed to help marginally the FIM analysis, in the case of BBH [126].

A CBC is identified by numerous parameters, both intrinsic, i.e., characterizing the source properties, and extrinsic, i.e., illustrating the relation between the detector and the source, which we collectively denote with θ . Among the intrinsics, in this work, we consider the masses $m_{1/2}$ and the longitudinal spins $\chi_{1/2z}$. Within the extrinsics, we instead consider: the inclination angle ι , i.e., the angle between the binary angular momentum and Earth position, the sky position, indicated with (θ, ϕ) , the polarization angle ψ and the time and phase to coalescence, indicated respectively with t_{coal} and Φ_{coal} .

When possible, we rely on the recent populations provided by the LVK collaboration [178]. In particular, the masses are drawn from a **PowerLaw + Peak** distribution [178], which consists of a power law plus a gaussian peak at $M_{\text{peak}} \approx 34M_\odot$ and the second mass is drawn from a power-law proportional to $\propto q^{-1.1}$ [178]. Regarding the spins, in the case of BBH, we employ the Default model from [178]. For the redshift distribution, we use the Madau-Dickinson distribution [187], which is convoluted with a time delay distribution. The time delay represents the time between the star formation and the merger of the BBH, and we consider a power-law distribution. A key and still uncertain parameter in this distribution is the minimum time delay allowed $t_{d,\min}$, which we consider 20 Myrs [188]. The sky positions are drawn from a uniform distribution over the sphere, i.e., φ , the azimuthal angle is drawn from a uniform distribution, and θ , the polar angle, is uniform in the cosine. Similarly to θ , the inclination angle ι is drawn from a uniform distribution in the cosine. Then, the polarization angle ψ represents the rotation needed to align the source frame, where the $h_{+/\times}$ polarizations are defined, with the detector frame [189]. This angle is sampled uniformly in $[0, \pi]$. The time to coalescence, t_{coal} , represents the time of the merger and is uniform in a day. The

⁴Notice that this statement is only valid statistically, i.e., when dealing with a large number of sources. Indeed, some CBCs may be better measured with a different orientation angle between the two detectors.

⁵from [181], the data that was used to create Figure 2 in the GWTC-3 analysis.

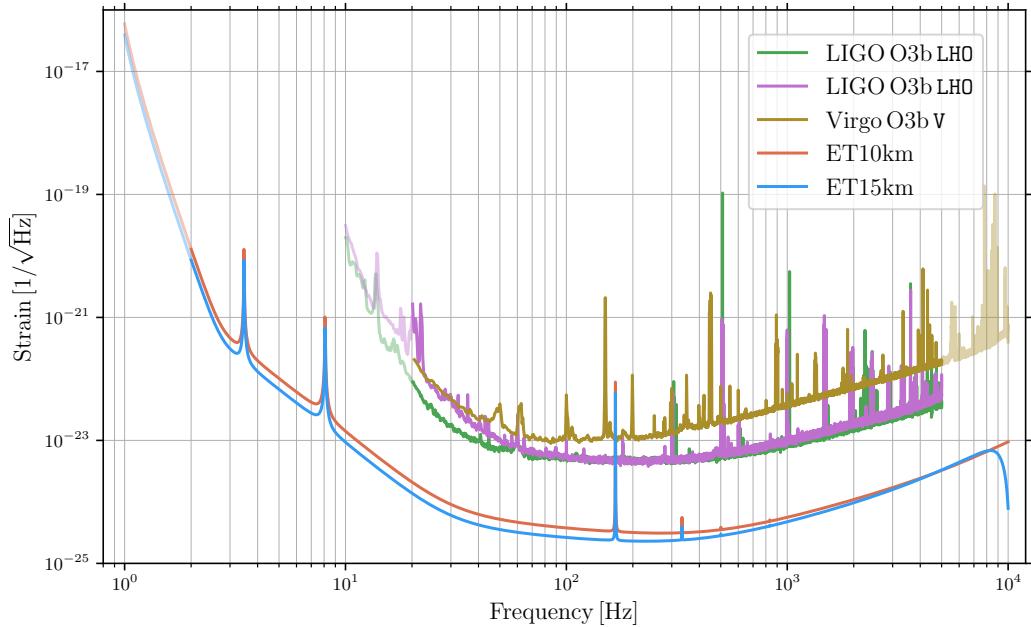


Figure 5: Plot of the detector amplitude spectral densities (ASD) for detectors used in this work: LHO, LLO, and V during the O3b observation run and ET with 10km and 15km arm length. The faint parts of the ASDs mark the regions below f_{\min} and above f_{\max} , omitted in the Fisher analysis.

phase to coalescence Φ_{coal} represents the phase at which the merger happens and, similar to t_{coal} , it is drawn from a uniform distribution, this time in the interval $[0, 2\pi]$.

Finally, for the values of local merger rate, i.e., the number of sources in a volume at redshift zero, we consider $R_0 = 17 \text{ [Gpc}^{-3}\text{yr}^{-1}]$ [178]. Moreover, the total number of sources in one year, which is obtained by integrating the redshift distribution of the sources, after the convolution with the time delay distribution, times the local rate, leads to $n_{\text{sources}} = 3.35 \times 10^4 \text{ yr}^{-1}$. In this work, we do not consider the overlap of different waveforms in the detector, which should be considered in a more refined analysis.

B The impact of the low frequency cutoff f_{\min}

This appendix discusses the choice of the minimum frequency used in the analysis. We explore the effect of changing f_{\min} in the PSD in fig. 5, in particular, f_{\min} acts as a hard cut-off, setting the PSD infinite at lower frequencies. In fig. 6 we consider a GW150914-like event, for different values of f_{\min} to assess the impact of the low frequency cutoff on the obtained GR deviation constraints at different PN orders in the case of ET. The error bars in the plot stem from sampling multiple GW150914-like events, with the binary component masses and the luminosity distance fixed, but varying the other binary parameters such as spins, inclination etc., as well as using different noise realizations for each event. The influence of f_{\min} is particularly important for the -1PN term, though other PN terms are also affected. As expected, lower-order PN terms are generally more sensitive to changes in f_{\min} , with the 0PN and 3PN terms being notable exceptions. For a direct comparison, fig. 6 also includes data points from Figure 1.1 of [190], where the authors conducted a full Bayesian parameter estimation of a comparable event. This allows for a clear comparison of our Fisher analysis with their full injection-recovery study.

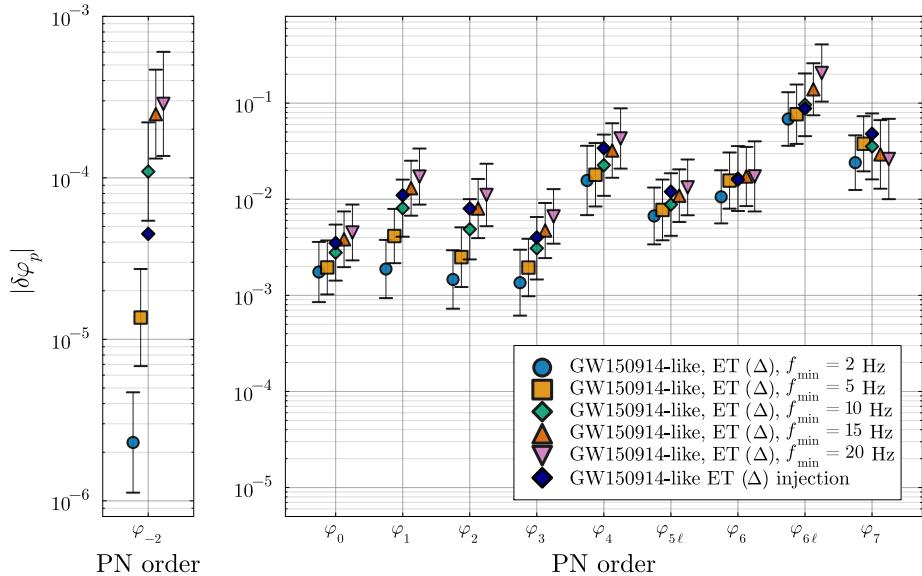


Figure 6: Forecast of the 90% upper bounds on the magnitude of the post-Newtonian deformation coefficients $\delta\varphi_p$ as a function of the minimum frequency f_{\min} used in the analysis, for a GW150914-like event. The forecast was performed using the `IMRPhenomHM` waveform model for the Einstein Telescope Δ configuration by varying the minimum frequency f_{\min} of the signal, as reported in the legend. The markers and the error bars for several f_{\min} configurations quantify, respectively, the average value and the 90% confidence interval of the bounds over 100 realizations of a GW150914-like event. This allows for taking into consideration detector noise realizations and different parameters of the event, keeping masses and luminosity distance fixed according to GW150914. We also report with dark blue diamonds the bounds of figure 1.1 of [190], for comparison of our result with their Bayesian injection-recovery analysis of a comparable event in the ET Δ configuration.

C Fisher approximated hyper-parameter distribution

In this appendix, we show explicit results of the Fisher approximated hyper-parameter distributions, comparing the full two-dimensional distributions to the conditioned ones. In Figure 7 we depict the resulting distribution in the deviations $\delta\varphi_p$ for all PN orders, both for the generic result (2.22), and for the conditioned result (2.26), which assumes all the GR deviations take on the same value for all of the single events. As can be seen from Figure 7, the conditioned posterior is always narrower, reflecting the additional physical prior information.

In Figure 8 we depict the 90% contours, as obtained from the Fisher approximated two-dimensional hyper-parameter distribution. For the analysis in Section 5, these contours were used to decide whether GR is contained in the 90% credibility region.

Comparing these plots depicted in Figures 7 and 8 with the corresponding posterior distributions of the GWTC-3 analysis [10], we find a similar improvement as in Figure 2, with distributions narrower between a factor of $\mathcal{O}(10^2)$ to $\mathcal{O}(10^4)$, especially for lower PN orders. The apparent preference towards positive deviations is an artifact of the particular realization depicted in Figure 8. This feature does not persist, for example, for different noise realizations while using the same catalog of events.

D On the role of higher-modes and selection effects for LVK

To clarify the origin of the tails in the distributions of single-event constraints for `IMRPhenomHM`, observed in the HLV network analysis, we examine how the constraints improve when transitioning

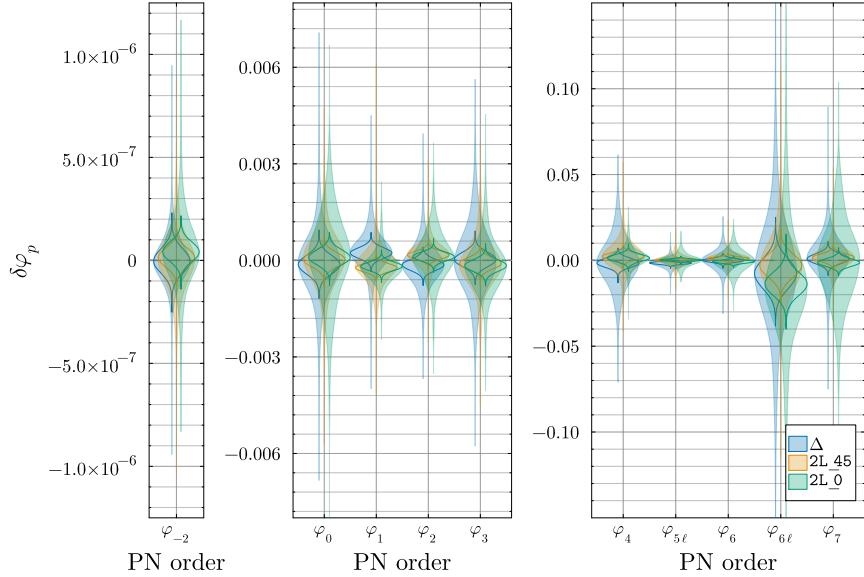


Figure 7: Visualization of the hierarchically combined posterior distribution (2.22) for the PN deviation coefficients $\delta\varphi_p$, as obtained from the Fisher approximation, for all three ET configurations considered in this work. The distributions have been obtained by hierarchically using a catalog of 10^4 events corresponding to an observation period of approximately four months, as described in Section 5. The filled violin plots represent the posterior distribution $P(\delta\varphi_p | D^{N_{\text{obs}}}, I)$ for the $\delta\varphi_p$ deformation coefficients. The unfilled violin plots, with thick outline, represent the conditioned posterior distribution $\hat{P}(\delta\varphi_p | D^{N_{\text{obs}}}, I)$, describing the scenario in which all GR deviations manifest with the same value in all single events.

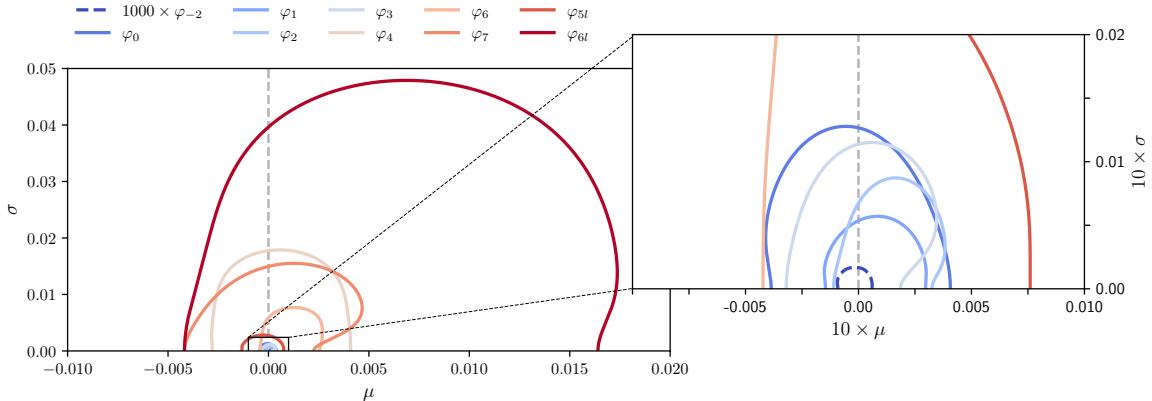


Figure 8: Visualization of the hierarchical hyper-parameter distribution $P(\mu, \sigma | D^{N_{\text{obs}}}, I)$ for the 2L_45 detector configuration of ET. The contours show the 90% credible region. The inset is scaled by a factor of 10. The contour plot corresponding to the -1PN order has been scaled by a factor of 1000.

from IMRPhenomD to IMRPhenomHM as a function of the binary parameters. Since the improvements are most pronounced for φ_7 , we will focus on the 3.5PN order. However, a similar discussion also holds for φ_{5l} and φ_6 . From the entire generated catalog, we restrict to the distribution of events with redshift $z < 0.5$, since the largest portion of events lies within this redshift range. Further, we restrict to events with detector frame chirp mass $(1+z)\mathcal{M} < 80M_\odot$, covering almost all events defined as

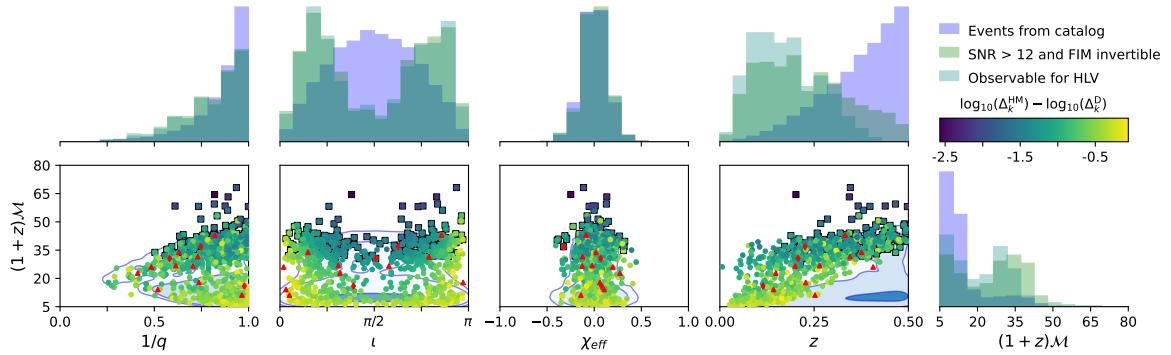


Figure 9: Visualization of the distribution of events for the single event constraints on $\delta\varphi_7$ for the HLV network. The histograms in blue represent the distribution of event parameters as generated by the catalog. The green histograms represent the distribution of events with an invertible Fisher information matrix (FIM) and overall SNR > 12 with both waveforms. The teal histogram represents the distribution of events defined as observable by HLV with both waveforms. To be defined as observable, these events must also have an inspiral SNR > 6 . The round marker in the scatter plots represents the parameters of the same observable events, with the color encoding the logarithmic improvement for the single event deviation constraint when passing from `IMRPhenomD` to `IMRPhenomHM`, quantified by the logarithmic ratio of the quantities Δ_k , defined in equation (2.19). Square markers denote events with SNR > 12 and an invertible FIM, but which are not considered among the observable ones, due to lacking SNR in the inspiral, with the color coding being the same as for the round ones. Red triangle markers denote events only observable when employing `IMRPhenomHM`, and diamond markers denote those only observable with `IMRPhenomD`.

observable in section 3.

The histograms in Figure 9 show a comparison between the distribution of single event parameters, before and after imposing the selection criteria that define an event as observable, and omitting the selection criteria on the inspiral SNR. We see that after the selection, a bimodal structure in the chirp mass distribution emerges. The peak in the mass reflects the Gaussian peak at $M_{\text{Peak}} \approx 34 M_\odot$ in the primary source mass distribution of the population model [178], used to generate our catalogs, see Appendix A. Further, we notice that the improvements between the two waveform models under consideration arise for the high-mass events. The improvement observed when transitioning from `IMRPhenomD` to `IMRPhenomHM` can be attributed to the enhanced contribution of the typically subdominant higher multipoles relative to the fundamental mode in these systems [191].

High mass detected events generate most of their signal in the late-inspiral and merger part of the waveform, making them not optimal candidates for BGR inspiral tests. Nonetheless, these high mass events contribute significantly to the fraction of observed LVK events due to their large detection horizon and the peak in the mass distribution of the population model. As evident from Figure 9, the most pronounced differences between the two waveform models under consideration arise for the high-mass events. The improvement observed when transitioning from `IMRPhenomD` to `IMRPhenomHM` can be attributed to the enhanced contribution of the typically subdominant higher multipoles relative to the fundamental mode in these systems [191].

The GR deviations have been parameterized by allowing modifications of the PN coefficients in the inspiral phase, which have been extended to the rest of the waveform by matching conditions, adjusting the onset-phase and time of the merger and ringdown. Moreover, we would like to stress that the magnitude of improvement has to be interpreted with caution, due to the short signal duration in the inspiral phase for these events. We thus consider the unexpectedly large improvements seen when using `IMRPhenomHM` to be a potential concern that merits additional scrutiny. This is further corroborated by looking at Figure 9, where events with inspiral SNR < 6 are marked with squares instead of round markers, providing the largest improvements when including higher modes. The extreme cases are excluded from the analysis resulting in Figure 1, by the definition of observable

events for the LHV network, see Section 3. Due to the frequency dependence of the various PN orders, explicit in (2.2), the higher PN coefficients are more sensitive to this effect. The combination of the bimodal structure in the chirp mass and the trend of improvements with increasing mass allows to explain the long tail in the single event distribution of $\delta\varphi_7$ observed in Figure 1. It is noteworthy that for ET differences between the two waveforms are less pronounced across the whole parameter space, due to ET’s lower minimum frequency $f_{\min} = 2$ Hz and overall improved sensitivity, allowing to have access to more information from the earlier inspiral portion of the signal, especially for the higher mass events present in the population. For this reason, we could omit the inspiral criterion for the results presented in Section 4.

References

- [1] LIGO SCIENTIFIC, VIRGO collaboration, *Observation of Gravitational Waves from a Binary Black Hole Merger*, *Phys. Rev. Lett.* **116** (2016) 061102 [[1602.03837](#)].
- [2] LIGO SCIENTIFIC, VIRGO, FERMI GBM, INTEGRAL, ICECUBE, ASTROSAT CADMIUM ZINC TELLURIDE IMAGER TEAM, IPN, INSIGHT-HXMT, ANTARES, SWIFT, AGILE TEAM, 1M2H TEAM, DARK ENERGY CAMERA GW-EM, DES, DLT40, GRAWITA, FERMI-LAT, ATCA, ASKAP, LAS CUMBRES OBSERVATORY GROUP, OZGRAV, DWF (DEEPER WIDER FASTER PROGRAM), AST3, CAASTRO, VINROUGE, MASTER, J-GEM, GROWTH, JAGWAR, CALTECHNRAO, TTU-NRAO, NuSTAR, PAN-STARRS, MAXI TEAM, TZAC CONSORTIUM, KU, NORDIC OPTICAL TELESCOPE, EPESSTO, GROND, TEXAS TECH UNIVERSITY, SALT GROUP, TOROS, BOOTES, MWA, CALET, IKI-GW FOLLOW-UP, H.E.S.S., LOFAR, LWA, HAWC, PIERRE AUGER, ALMA, EURO VLBI TEAM, PI OF SKY, CHANDRA TEAM AT MCGILL UNIVERSITY, DFN, ATLAS TELESCOPES, HIGH TIME RESOLUTION UNIVERSE SURVEY, RIMAS, RATIR, SKA SOUTH AFRICA/MEERKAT collaboration, *Multi-messenger Observations of a Binary Neutron Star Merger*, *Astrophys. J. Lett.* **848** (2017) L12 [[1710.05833](#)].
- [3] LIGO SCIENTIFIC, VIRGO collaboration, *GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs*, *Phys. Rev. X* **9** (2019) 031040 [[1811.12907](#)].
- [4] LIGO SCIENTIFIC, VIRGO collaboration, *GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run*, *Phys. Rev. X* **11** (2021) 021053 [[2010.14527](#)].
- [5] LIGO SCIENTIFIC, VIRGO collaboration, *GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run*, *Phys. Rev. D* **109** (2024) 022001 [[2108.01045](#)].
- [6] KAGRA, VIRGO, LIGO SCIENTIFIC collaboration, *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run*, *Phys. Rev. X* **13** (2023) 041039 [[2111.03606](#)].
- [7] LIGO SCIENTIFIC, VIRGO, KAGRA collaboration, *GWTC-4.0: Updating the Gravitational-Wave Transient Catalog with Observations from the First Part of the Fourth LIGO-Virgo-KAGRA Observing Run*, [2508.18082](#).
- [8] LIGO SCIENTIFIC, VIRGO collaboration, *Tests of General Relativity with the Binary Black Hole Signals from the LIGO-Virgo Catalog GWTC-1*, *Phys. Rev. D* **100** (2019) 104036 [[1903.04467](#)].
- [9] LIGO SCIENTIFIC, VIRGO collaboration, *Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog*, *Phys. Rev. D* **103** (2021) 122002 [[2010.14529](#)].
- [10] LIGO SCIENTIFIC, VIRGO, KAGRA collaboration, *Tests of General Relativity with GWTC-3*, [2112.06861](#).
- [11] E.M. Sänger et al., *Tests of General Relativity with GW230529: a neutron star merging with a lower mass-gap compact object*, [2406.03568](#).
- [12] LIGO SCIENTIFIC, VIRGO, KAGRA collaboration, *GW250114: Testing Hawking's Area Law and the Kerr Nature of Black Holes*, *Phys. Rev. Lett.* **135** (2025) 111403 [[2509.08054](#)].
- [13] LIGO SCIENTIFIC, VIRGO, KAGRA collaboration, *Directed searches for gravitational waves from ultralight vector boson clouds around merger remnant and galactic black holes during the first part of the fourth LIGO-Virgo-KAGRA observing run*, [2509.07352](#).
- [14] LIGO SCIENTIFIC, VIRGO, KAGRA collaboration, *GW241011 and GW241110: Exploring Binary Formation and Fundamental Physics with Asymmetric, High-spin Black Hole Coalescences*, *Astrophys. J. Lett.* **993** (2025) L21 [[2510.26931](#)].
- [15] E. Berti et al., *Testing General Relativity with Present and Future Astrophysical Observations*, *Class. Quant. Grav.* **32** (2015) 243001 [[1501.07274](#)].
- [16] M. Ishak, *Testing General Relativity in Cosmology*, *Living Rev. Rel.* **22** (2019) 1 [[1806.10122](#)].

- [17] E. Payne, M. Isi, K. Chatzioannou, L. Lehner, Y. Chen and W.M. Farr, *Curvature Dependence of Gravitational-Wave Tests of General Relativity*, *Phys. Rev. Lett.* **133** (2024) 251401 [[2407.07043](#)].
- [18] L. Bernard, S. Giri, L. Lehner and R. Sturani, *Generic EFT-motivated beyond General Relativity gravitational wave tests and their curvature dependence: from observation to interpretation*, [2507.17143](#).
- [19] C.M. Will, *The Confrontation between General Relativity and Experiment*, *Living Rev. Rel.* **17** (2014) 4 [[1403.7377](#)].
- [20] J.Y.L. Kwok, R.K.L. Lo, A.J. Weinstein and T.G.F. Li, *Investigation of the effects of non-Gaussian noise transients and their mitigation in parameterized gravitational-wave tests of general relativity*, *Phys. Rev. D* **105** (2022) 024066 [[2109.07642](#)].
- [21] L. Reali, A. Maselli and E. Berti, *The impact of compact binary confusion noise on tests of fundamental physics with next-generation gravitational-wave detectors*, [2307.01264](#).
- [22] S. Roy and R. Vicente, *Compact binary coalescences in dense gaseous environments can pose as ones in vacuum*, *Phys. Rev. D* **111** (2025) 084037 [[2410.16388](#)].
- [23] A. Gupta et al., *Possible causes of false general relativity violations in gravitational wave observations*, [2405.02197](#).
- [24] E. Maggio, H.O. Silva, A. Buonanno and A. Ghosh, *Tests of general relativity in the nonlinear regime: A parametrized plunge-merger-ringdown gravitational waveform model*, *Phys. Rev. D* **108** (2023) 024043 [[2212.09655](#)].
- [25] L. Pompili, E. Maggio, H.O. Silva and A. Buonanno, *A parametrized spin-precessing inspiral-merger-ringdown waveform model for tests of general relativity*, [2504.10130](#).
- [26] S.M. Brown, B. Krishnan, R. Somasundaram and I. Tews, *Can Neutron Star Tidal Effects Obscure Deviations from General Relativity?*, *Astrophys. J.* **982** (2025) 133 [[2411.19129](#)].
- [27] A. Dhani, S.H. Völkel, A. Buonanno, H. Estelles, J. Gair, H.P. Pfeiffer et al., *Systematic Biases in Estimating the Properties of Black Holes Due to Inaccurate Gravitational-Wave Models*, *Phys. Rev. X* **15** (2025) 031036 [[2404.05811](#)].
- [28] V. Cardoso and P. Pani, *Testing the nature of dark compact objects: a status report*, *Living Rev. Rel.* **22** (2019) 4 [[1904.05363](#)].
- [29] O. Lunin and S.D. Mathur, *AdS / CFT duality and the black hole information paradox*, *Nucl. Phys. B* **623** (2002) 342 [[hep-th/0109154](#)].
- [30] P.O. Mazur and E. Mottola, *Gravitational vacuum condensate stars*, *Proc. Nat. Acad. Sci.* **101** (2004) 9545 [[gr-qc/0407075](#)].
- [31] S.L. Liebling and C. Palenzuela, *Dynamical boson stars*, *Living Rev. Rel.* **26** (2023) 1 [[1202.5809](#)].
- [32] G.F. Giudice, M. McCullough and A. Urbano, *Hunting for Dark Particles with Gravitational Waves*, *JCAP* **10** (2016) 001 [[1605.01209](#)].
- [33] U. Danielsson, L. Lehner and F. Pretorius, *Dynamics and observational signatures of shell-like black hole mimickers*, *Phys. Rev. D* **104** (2021) 124011 [[2109.09814](#)].
- [34] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *Black Holes: Complementarity or Firewalls?*, *JHEP* **02** (2013) 062 [[1207.3123](#)].
- [35] R. Brito, V. Cardoso and P. Pani, *Superradiance: New Frontiers in Black Hole Physics*, *Lect. Notes Phys.* **906** (2015) pp.1 [[1501.06570](#)].
- [36] V. Cardoso, S. Hopper, C.F.B. Macedo, C. Palenzuela and P. Pani, *Gravitational-wave signatures of exotic compact objects and of quantum corrections at the horizon scale*, *Phys. Rev. D* **94** (2016) 084031 [[1608.08637](#)].
- [37] V. Cardoso, E. Franzin and P. Pani, *Is the gravitational-wave ringdown a probe of the event horizon?*, *Phys. Rev. Lett.* **116** (2016) 171101 [[1602.07309](#)].
- [38] V. Cardoso and P. Pani, *Tests for the existence of black holes through gravitational wave echoes*, *Nature Astron.* **1** (2017) 586 [[1709.01525](#)].

- [39] E. Maggio, P. Pani and G. Raposo, *Testing the nature of dark compact objects with gravitational waves*, [2105.06410](#).
- [40] S. Chakraborty, E. Maggio, A. Mazumdar and P. Pani, *Implications of the quantum nature of the black hole horizon on the gravitational-wave ringdown*, *Phys. Rev. D* **106** (2022) 024041 [[2202.09111](#)].
- [41] M. Silvestrini, E. Maggio, S. Chakraborty and P. Pani, *One Membrane to Love them all: Tidal deformations of compact objects from the membrane paradigm*, [2506.16516](#).
- [42] S. Mastrogiovanni, E. Maggio and A.F. Mascioli, *Limits on the existence of totally reflective exotic compact objects with current and future gravitational-wave detectors*, *Class. Quant. Grav.* **42** (2025) 135016 [[2502.07675](#)].
- [43] E. Berti et al., *Black hole spectroscopy: from theory to experiment*, [2505.23895](#).
- [44] A. Iannicciari, A.J. Iovino, A. Kehagias, P. Pani, G. Perna, D. Perrone et al., *Deciphering the Instability of the Black Hole Ringdown Quasinormal Spectrum*, *Phys. Rev. Lett.* **133** (2024) 211401 [[2407.20144](#)].
- [45] X.J. Forteza, S. Bhagwat, S. Kumar and P. Pani, *Novel Ringdown Amplitude-Phase Consistency Test*, *Phys. Rev. Lett.* **130** (2023) 021001 [[2205.14910](#)].
- [46] D. Baumann, H.S. Chia and R.A. Porto, *Probing Ultralight Bosons with Binary Black Holes*, *Phys. Rev. D* **99** (2019) 044001 [[1804.03208](#)].
- [47] D. Baumann, G. Bertone, J. Stout and G.M. Tomaselli, *Sharp Signals of Boson Clouds in Black Hole Binary Inspirals*, *Phys. Rev. Lett.* **128** (2022) 221102 [[2206.01212](#)].
- [48] D. Baumann, H.S. Chia, J. Stout and L. ter Haar, *The Spectra of Gravitational Atoms*, *JCAP* **12** (2019) 006 [[1908.10370](#)].
- [49] T. Ikeda, L. Bernard, V. Cardoso and M. Zilhão, *Black hole binaries and light fields: Gravitational molecules*, *Phys. Rev. D* **103** (2021) 024020 [[2010.00008](#)].
- [50] D. Baumann, G. Bertone, J. Stout and G.M. Tomaselli, *Ionization of gravitational atoms*, *Phys. Rev. D* **105** (2022) 115036 [[2112.14777](#)].
- [51] R. Della Monica and R. Brito, *Detectability of gravitational atoms in black hole binaries with the Einstein Telescope*, *Phys. Rev. D* **112** (2025) 024074 [[2503.23419](#)].
- [52] S. Barsanti, A. Maselli, T.P. Sotiriou and L. Gualtieri, *Detecting Massive Scalar Fields with Extreme Mass-Ratio Inspirals*, *Phys. Rev. Lett.* **131** (2023) 051401 [[2212.03888](#)].
- [53] A. Maselli, S. Yi, L. Pierini, V. Vellucci, L. Reali, L. Gualtieri et al., *Black hole spectroscopy beyond Kerr: Agnostic and theory-based tests with next-generation interferometers*, *Phys. Rev. D* **109** (2024) 064060 [[2311.14803](#)].
- [54] M. Della Rocca, G. Antoniou, L. Gualtieri and A. Maselli, *Probing time-dependent scalar wigs with extreme-mass-ratio inspirals*, *Phys. Rev. D* **112** (2025) 024042 [[2412.15131](#)].
- [55] C. Bambi et al., *Black hole mimickers: from theory to observation*, 5, 2025 [[2505.09014](#)].
- [56] A. Begnoni and S. Profumo, *Primordial black holes versus their impersonators at gravitational wave observatories*, [2509.04028](#).
- [57] G. Nicolini, A. Maselli and M. Zilhão, *Dancing in the dark: probing Dark Matter through the dynamics of eccentric binary pulsars*, [2507.16904](#).
- [58] S. Kejriwal, E. Barausse and A.J.K. Chua, *Hierarchical modeling of gravitational-wave populations for disentangling environmental and modified-gravity effects*, [2510.17398](#).
- [59] M. Agathos, W. Del Pozzo, T.G.F. Li, C. Van Den Broeck, J. Veitch and S. Vitale, *TIGER: A data analysis pipeline for testing the strong-field dynamics of general relativity with gravitational wave signals from coalescing compact binaries*, *Phys. Rev. D* **89** (2014) 082001 [[1311.0420](#)].
- [60] J. Meidam et al., *Parametrized tests of the strong-field dynamics of general relativity using gravitational wave signals from coalescing binary black holes: Fast likelihood calculations and sensitivity of the method*, *Phys. Rev. D* **97** (2018) 044033 [[1712.08772](#)].

- [61] T.G.F. Li, W. Del Pozzo, S. Vitale, C. Van Den Broeck, M. Agathos, J. Veitch et al., *Towards a generic test of the strong field dynamics of general relativity using compact binary coalescence*, *Phys. Rev. D* **85** (2012) 082003 [[1110.0530](#)].
- [62] L. Blanchet, *Post-Newtonian Theory for Gravitational Waves*, *Living Rev. Rel.* **17** (2014) 2 [[1310.1528](#)].
- [63] K.S. Thorne, *Multipole Expansions of Gravitational Radiation*, *Rev. Mod. Phys.* **52** (1980) 299.
- [64] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Jiménez Forteza et al., *Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era*, *Phys. Rev. D* **93** (2016) 044007 [[1508.07253](#)].
- [65] A.K. Mehta, A. Buonanno, R. Cotesta, A. Ghosh, N. Sennett and J. Steinhoff, *Tests of general relativity with gravitational-wave observations using a flexible theory-independent method*, *Phys. Rev. D* **107** (2023) 044020 [[2203.13937](#)].
- [66] M. Piarulli, S. Marsat, E.M. Sänger, A. Buonanno, J. Steinhoff and N. Tamanini, *A Parametrized Test of General Relativity for LISA Massive Black Hole Binary Inspirals*, [2510.06330](#).
- [67] K. Nordtvedt, *Equivalence Principle for Massive Bodies. 2. Theory*, *Phys. Rev.* **169** (1968) 1017.
- [68] C.M. Will and K. Nordtvedt, Jr., *Conservation Laws and Preferred Frames in Relativistic Gravity. I. Preferred-Frame Theories and an Extended PPN Formalism*, *Astrophys. J.* **177** (1972) 757.
- [69] C.M. Will, *Theoretical Frameworks for Testing Relativistic Gravity. 2. Parametrized Post-Newtonian Hydrodynamics, and the Nordtvedt Effect*, *Astrophys. J.* **163** (1971) 611.
- [70] K.J. Nordtvedt and C.M. Will, *Conservation Laws and Preferred Frames in Relativistic Gravity. II. Experimental Evidence to Rule Out Preferred-Frame Theories of Gravity*, *Astrophys. J.* **177** (1972) 775.
- [71] C.M. Will, *Relativistic Gravity in the Solar System. 111. Experimental Disproof of a Class of Linear Theories of Gravitation*, *Astrophys. J.* **185** (1973) 31.
- [72] T. Damour and J.H. Taylor, *On the orbital period change of the binary pulsar PSR-1913+16*, *Astrophys. J.* **366** (1991) 501.
- [73] T. Damour and J.H. Taylor, *Strong field tests of relativistic gravity and binary pulsars*, *Phys. Rev. D* **45** (1992) 1840.
- [74] N. Yunes and F. Pretorius, *Fundamental Theoretical Bias in Gravitational Wave Astrophysics and the Parameterized Post-Einsteinian Framework*, *Phys. Rev. D* **80** (2009) 122003 [[0909.3328](#)].
- [75] N. Cornish, L. Sampson, N. Yunes and F. Pretorius, *Gravitational Wave Tests of General Relativity with the Parameterized Post-Einsteinian Framework*, *Phys. Rev. D* **84** (2011) 062003 [[1105.2088](#)].
- [76] K. Chatzioannou, N. Yunes and N. Cornish, *Model-Independent Test of General Relativity: An Extended post-Einsteinian Framework with Complete Polarization Content*, *Phys. Rev. D* **86** (2012) 022004 [[1204.2585](#)].
- [77] N. Yunes, K. Yagi and F. Pretorius, *Theoretical Physics Implications of the Binary Black-Hole Mergers GW150914 and GW151226*, *Phys. Rev. D* **94** (2016) 084002 [[1603.08955](#)].
- [78] S. Tahura and K. Yagi, *Parameterized Post-Einsteinian Gravitational Waveforms in Various Modified Theories of Gravity*, *Phys. Rev. D* **98** (2018) 084042 [[1809.00259](#)].
- [79] G.S. Bonilla, P. Kumar and S.A. Teukolsky, *Modeling compact binary merger waveforms beyond general relativity*, *Phys. Rev. D* **107** (2023) 024015 [[2203.14026](#)].
- [80] N. Loutrel, P. Pani and N. Yunes, *Parametrized post-Einsteinian framework for precessing binaries*, *Phys. Rev. D* **107** (2023) 044046 [[2210.10571](#)].
- [81] Y. Xie, D. Chatterjee, G. Narayan and N. Yunes, *Neural post-Einsteinian framework for efficient theory-agnostic tests of general relativity with gravitational waves*, *Phys. Rev. D* **110** (2024) 024036 [[2403.18936](#)].
- [82] Y. Xie, G. Narayan and N. Yunes, *Neural Post-Einsteinian Test of General Relativity with the Third Gravitational-Wave Transient Catalog*, [2510.02515](#).

- [83] S. Endlich, V. Gorbenko, J. Huang and L. Senatore, *An effective formalism for testing extensions to General Relativity with gravitational waves*, *JHEP* **09** (2017) 122 [[1704.01590](#)].
- [84] N. Sennett, R. Brito, A. Buonanno, V. Gorbenko and L. Senatore, *Gravitational-Wave Constraints on an Effective Field-Theory Extension of General Relativity*, *Phys. Rev. D* **102** (2020) 044056 [[1912.09917](#)].
- [85] N. Yunes and F. Pretorius, *Dynamical Chern-Simons Modified Gravity. I. Spinning Black Holes in the Slow-Rotation Approximation*, *Phys. Rev. D* **79** (2009) 084043 [[0902.4669](#)].
- [86] C. Molina, P. Pani, V. Cardoso and L. Gualtieri, *Gravitational signature of Schwarzschild black holes in dynamical Chern-Simons gravity*, *Phys. Rev. D* **81** (2010) 124021 [[1004.4007](#)].
- [87] F.-L. Julié, *On the motion of hairy black holes in Einstein-Maxwell-dilaton theories*, *JCAP* **01** (2018) 026 [[1711.10769](#)].
- [88] M. Khalil, N. Sennett, J. Steinhoff, J. Vines and A. Buonanno, *Hairy binary black holes in Einstein-Maxwell-dilaton theory and their effective-one-body description*, *Phys. Rev. D* **98** (2018) 104010 [[1809.03109](#)].
- [89] F.-L. Julié, *Gravitational radiation from compact binary systems in Einstein-Maxwell-dilaton theories*, *JCAP* **10** (2018) 033 [[1809.05041](#)].
- [90] P. Kanti, N.E. Mavromatos, J. Rizos, K. Tamvakis and E. Winstanley, *Dilatonic black holes in higher curvature string gravity*, *Phys. Rev. D* **54** (1996) 5049 [[hep-th/9511071](#)].
- [91] A. Maselli, P. Pani, L. Gualtieri and V. Ferrari, *Rotating black holes in Einstein-Dilaton-Gauss-Bonnet gravity with finite coupling*, *Phys. Rev. D* **92** (2015) 083014 [[1507.00680](#)].
- [92] J.L. Blázquez-Salcedo, C.F.B. Macedo, V. Cardoso, V. Ferrari, L. Gualtieri, F.S. Khoo et al., *Perturbed black holes in Einstein-dilaton-Gauss-Bonnet gravity: Stability, ringdown, and gravitational-wave emission*, *Phys. Rev. D* **94** (2016) 104024 [[1609.01286](#)].
- [93] N. Sennett, L. Shao and J. Steinhoff, *Effective action model of dynamically scalarizing binary neutron stars*, *Phys. Rev. D* **96** (2017) 084019 [[1708.08285](#)].
- [94] H.O. Silva, J. Sakstein, L. Gualtieri, T.P. Sotiriou and E. Berti, *Spontaneous scalarization of black holes and compact stars from a Gauss-Bonnet coupling*, *Phys. Rev. Lett.* **120** (2018) 131104 [[1711.02080](#)].
- [95] H. Witek, L. Gualtieri, P. Pani and T.P. Sotiriou, *Black holes and binary mergers in scalar Gauss-Bonnet gravity: scalar field dynamics*, *Phys. Rev. D* **99** (2019) 064035 [[1810.05177](#)].
- [96] A. Kuntz, F. Piazza and F. Vernizzi, *Effective field theory for gravitational radiation in scalar-tensor gravity*, *JCAP* **05** (2019) 052 [[1902.04941](#)].
- [97] M. Khalil, N. Sennett, J. Steinhoff and A. Buonanno, *Theory-agnostic framework for dynamical scalarization of compact binaries*, *Phys. Rev. D* **100** (2019) 124013 [[1906.08161](#)].
- [98] F.-L. Julié and E. Berti, *Post-Newtonian dynamics and black hole thermodynamics in Einstein-scalar-Gauss-Bonnet gravity*, *Phys. Rev. D* **100** (2019) 104061 [[1909.05258](#)].
- [99] S.E. Perkins, R. Nair, H.O. Silva and N. Yunes, *Improved gravitational-wave constraints on higher-order curvature theories of gravity*, *Phys. Rev. D* **104** (2021) 024060 [[2104.11189](#)].
- [100] F.-L. Julié, H.O. Silva, E. Berti and N. Yunes, *Black hole sensitivities in Einstein-scalar-Gauss-Bonnet gravity*, *Phys. Rev. D* **105** (2022) 124031 [[2202.01329](#)].
- [101] L. Bernard, L. Blanchet and D. Trestini, *Gravitational waves in scalar-tensor theory to one-and-a-half post-Newtonian order*, *JCAP* **08** (2022) 008 [[2201.10924](#)].
- [102] F.-L. Julié, V. Baibhav, E. Berti and A. Buonanno, *Third post-Newtonian effective-one-body Hamiltonian in scalar-tensor and Einstein-scalar-Gauss-Bonnet gravity*, *Phys. Rev. D* **107** (2023) 104044 [[2212.13802](#)].
- [103] G. Creci, T. Hinderer and J. Steinhoff, *Tidal properties of neutron stars in scalar-tensor theories of gravity*, *Phys. Rev. D* **108** (2023) 124073 [[2308.11323](#)].
- [104] L. Bernard, E. Dones and S. Mousiakakos, *Tidal effects up to next-to-next-to-leading post-Newtonian order in massless scalar-tensor theories*, *Phys. Rev. D* **109** (2024) 044006 [[2310.19679](#)].

- [105] R.F. Diedrichs, D. Schmitt and L. Sagunski, *Binary systems in massive scalar-tensor theories: Next-to-leading order gravitational wave phase from effective field theory*, *Phys. Rev. D* **110** (2024) 104073 [[2311.04274](#)].
- [106] G.L. Almeida, *Binary dynamics to second post-Newtonian order in scalar-tensor and Einstein-scalar-Gauss-Bonnet gravity from effective field theory*, *Phys. Rev. D* **109** (2024) 084060 [[2402.13996](#)].
- [107] F.-L. Julié, L. Pompili and A. Buonanno, *Inspiral-merger-ringdown waveforms in Einstein-scalar-Gauss-Bonnet gravity within the effective-one-body formalism*, *Phys. Rev. D* **111** (2025) 024016 [[2406.13654](#)].
- [108] G.L. Almeida and S.-Y. Zhou, *Post-Newtonian dynamics of spinning black hole binaries in Einstein-scalar-Gauss-Bonnet gravity*, *Phys. Rev. D* **110** (2024) 124016 [[2408.14196](#)].
- [109] E. Dones and L. Bernard, *Tidal effects in gravitational and scalar waveforms and fluxes to one post-Newtonian order in massless scalar-tensor theories*, [2507.07676](#).
- [110] N. Yunes and X. Siemens, *Gravitational-Wave Tests of General Relativity with Ground-Based Detectors and Pulsar Timing-Arrays*, *Living Rev. Rel.* **16** (2013) 9 [[1304.3473](#)].
- [111] LIGO SCIENTIFIC, VIRGO collaboration, *Tests of general relativity with GW150914*, *Phys. Rev. Lett.* **116** (2016) 221101 [[1602.03841](#)].
- [112] LIGO SCIENTIFIC, VIRGO collaboration, *Tests of General Relativity with GW170817*, *Phys. Rev. Lett.* **123** (2019) 011102 [[1811.00364](#)].
- [113] LIGO SCIENTIFIC, VIRGO collaboration, *GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence*, *Phys. Rev. Lett.* **119** (2017) 141101 [[1709.09660](#)].
- [114] K. Chamberlain and N. Yunes, *Theoretical Physics Implications of Gravitational Wave Observation with Future Detectors*, *Phys. Rev. D* **96** (2017) 084039 [[1704.08268](#)].
- [115] S.E. Perkins, N. Yunes and E. Berti, *Probing Fundamental Physics with Gravitational Waves: The Next Generation*, *Phys. Rev. D* **103** (2021) 044024 [[2010.09010](#)].
- [116] A. Gupta, S. Datta, S. Kastha, S. Borhanian, K.G. Arun and B.S. Sathyaprakash, *Multiparameter tests of general relativity using multiband gravitational-wave observations*, *Phys. Rev. Lett.* **125** (2020) 201101 [[2005.09607](#)].
- [117] M. Isi, K. Chatzioannou and W.M. Farr, *Hierarchical test of general relativity with gravitational waves*, *Physical Review Letters* **123** (2019) .
- [118] ET collaboration, *Science Case for the Einstein Telescope*, *JCAP* **03** (2020) 050 [[1912.02622](#)].
- [119] A. Abac et al., *The Science of the Einstein Telescope*, [2503.12263](#).
- [120] M. Punturo et al., *The third generation of gravitational wave observatories and their science reach*, *Class. Quant. Grav.* **27** (2010) 084007.
- [121] M. Branchesi et al., *Science with the Einstein Telescope: a comparison of different designs*, *JCAP* **07** (2023) 068 [[2303.15923](#)].
- [122] L. London, S. Khan, E. Fauchon-Jones, C. García, M. Hannam, S. Husa et al., *First higher-multipole model of gravitational waves from spinning and coalescing black-hole binaries*, *Phys. Rev. Lett.* **120** (2018) 161102 [[1708.00404](#)].
- [123] C. Kalaghatgi, M. Hannam and V. Raymond, *Parameter estimation with a spinning multimode waveform model*, *Phys. Rev. D* **101** (2020) 103004 [[1909.10010](#)].
- [124] G. Pratten, S. Husa, C. Garcia-Quiros, M. Colleoni, A. Ramos-Buades, H. Estelles et al., *Setting the cornerstone for a family of models for gravitational waves from compact binaries: The dominant harmonic for nonprecessing quasicircular black holes*, *Phys. Rev. D* **102** (2020) 064001 [[2001.11412](#)].
- [125] S. Husa, S. Khan, M. Hannam, M. Pürller, F. Ohme, X. Jiménez Forteza et al., *Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal*, *Phys. Rev. D* **93** (2016) 044006 [[1508.07250](#)].
- [126] A. Begnoni, S. Ansaldi, M. Pieroni, A. Renzi and A. Ricciardone, *Detectability and Parameter Estimation for Einstein Telescope Configurations with GWJulia*, [2506.21530](#).

- [127] R. Cottesta, A. Buonanno, A. Bohé, A. Taracchini, I. Hinder and S. Ossokine, *Enriching the Symphony of Gravitational Waves from Binary Black Holes by Tuning Higher Harmonics*, *Phys. Rev. D* **98** (2018) 084028 [[1803.10701](#)].
- [128] M. Maggiore, *Gravitational waves*, vol. 2, Oxford university press (2008).
- [129] C. Cutler and E.E. Flanagan, *Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral wave form?*, *Phys. Rev. D* **49** (1994) 2658 [[gr-qc/9402014](#)].
- [130] T. Damour, B.R. Iyer and B.S. Sathyaprakash, *A Comparison of search templates for gravitational waves from binary inspiral*, *Phys. Rev. D* **63** (2001) 044023 [[gr-qc/0010009](#)].
- [131] A. Buonanno, B. Iyer, E. Ochsner, Y. Pan and B.S. Sathyaprakash, *Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors*, *Phys. Rev. D* **80** (2009) 084043 [[0907.0700](#)].
- [132] L. Blanchet, G. Faye, Q. Henry, F. Larrouture and D. Trestini, *Gravitational-Wave Phasing of Quasicircular Compact Binary Systems to the Fourth-and-a-Half Post-Newtonian Order*, *Phys. Rev. Lett.* **131** (2023) 121402 [[2304.11185](#)].
- [133] F. Iacovelli, M. Mancarella, S. Foffa and M. Maggiore, *Forecasting the Detection Capabilities of Third-generation Gravitational-wave Detectors Using GWFAST*, *Astrophys. J.* **941** (2022) 208 [[2207.02771](#)].
- [134] S. Roy, M. Haney, G. Pratten, P.T.H. Pang and C. Van Den Broeck, *An improved parametrized test of general relativity using the IMRPhenomX waveform family: Including higher harmonics and precession*, [2504.21147](#).
- [135] N. Christensen and R. Meyer, *Parameter estimation with gravitational waves*, *Rev. Mod. Phys.* **94** (2022) 025001 [[2204.04449](#)].
- [136] A. Freise, S. Chelkowski, S. Hild, W. Del Pozzo, A. Perreca and A. Vecchio, *Triple Michelson Interferometer for a Third-Generation Gravitational Wave Detector*, *Class. Quant. Grav.* **26** (2009) 085012 [[0804.1036](#)].
- [137] M. Vallisneri, *Use and abuse of the fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects*, *Physical Review D* **77** (2008) .
- [138] C. Pacilio, D. Gerosa and S. Bhagwat, *Catalog variance of testing general relativity with gravitational-wave data*, *Physical Review D* **109** (2024) L081302.
- [139] H. Zhong, M. Isi, K. Chatzioannou and W.M. Farr, *Multidimensional hierarchical tests of general relativity with gravitational waves*, *Physical Review D* **110** (2024) 044053.
- [140] LIGO SCIENTIFIC collaboration, *Advanced LIGO*, *Class. Quant. Grav.* **32** (2015) 074001 [[1411.4547](#)].
- [141] VIRGO collaboration, *Advanced Virgo: a second-generation interferometric gravitational wave detector*, *Class. Quant. Grav.* **32** (2015) 024001 [[1408.3978](#)].
- [142] H.-Y. Chen and D.E. Holz, *The Loudest Gravitational Wave Events*, [1409.0522](#).
- [143] K.G. Arun, *Generic bounds on dipolar gravitational radiation from inspiralling compact binaries*, *Class. Quant. Grav.* **29** (2012) 075011 [[1202.5911](#)].
- [144] E. Barausse, N. Yunes and K. Chamberlain, *Theory-Agnostic Constraints on Black-Hole Dipole Radiation with Multiband Gravitational-Wave Astrophysics*, *Phys. Rev. Lett.* **116** (2016) 241104 [[1603.04075](#)].
- [145] A.G. Lyne et al., *A Double - pulsar system - A Rare laboratory for relativistic gravity and plasma physics*, *Science* **303** (2004) 1153 [[astro-ph/0401086](#)].
- [146] M. Kramer et al., *Tests of general relativity from timing the double pulsar*, *Science* **314** (2006) 97 [[astro-ph/0609417](#)].
- [147] N. Yunes and S.A. Hughes, *Binary Pulsar Constraints on the Parameterized post-Einsteinian Framework*, *Phys. Rev. D* **82** (2010) 082002 [[1007.1995](#)].
- [148] N. Wex, *Testing Relativistic Gravity with Radio Pulsars*, [1402.5594](#).

- [149] M. Kramer, *Pulsars as probes of gravity and fundamental physics*, *Int. J. Mod. Phys. D* **25** (2016) 1630029 [[1606.03843](#)].
- [150] L. Shao, N. Sennett, A. Buonanno, M. Kramer and N. Wex, *Constraining nonperturbative strong-field effects in scalar-tensor gravity by combining pulsar timing and laser-interferometer gravitational-wave detectors*, *Phys. Rev. X* **7** (2017) 041025 [[1704.07561](#)].
- [151] P.C.C. Freire, N. Wex, G. Esposito-Farese, J.P.W. Verbiest, M. Bailes, B.A. Jacoby et al., *The relativistic pulsar-white dwarf binary PSR J1738+0333 II. The most stringent test of scalar-tensor gravity*, *Mon. Not. Roy. Astron. Soc.* **423** (2012) 3328 [[1205.1450](#)].
- [152] D. Anderson, P. Freire and N. Yunes, *Binary pulsar constraints on massless scalar-tensor theories using Bayesian statistics*, *Class. Quant. Grav.* **36** (2019) 225009 [[1901.00938](#)].
- [153] R. Nair and N. Yunes, *Improved binary pulsar constraints on the parametrized post-Einsteinian framework*, *Phys. Rev. D* **101** (2020) 104011 [[2002.02030](#)].
- [154] M. Kramer et al., *Strong-Field Gravity Tests with the Double Pulsar*, *Phys. Rev. X* **11** (2021) 041050 [[2112.06795](#)].
- [155] J. Zhao, L. Shao, Y. Gao, C. Liu, Z. Cao and B.-Q. Ma, *Probing dipole radiation from binary neutron stars with ground-based laser-interferometer and atom-interferometer gravitational-wave observatories*, *Phys. Rev. D* **104** (2021) 084008 [[2106.04883](#)].
- [156] F. Cireddu, M. Wils, I.C.F. Wong, P.T.H. Pang, T.G.F. Li and W. Del Pozzo, *Likelihood for a network of gravitational-wave detectors with correlated noise*, *Phys. Rev. D* **110** (2024) 104060 [[2312.14614](#)].
- [157] I.C.F. Wong, P.T.H. Pang, M. Wils, F. Cireddu, W. Del Pozzo and T.G.F. Li, *Potential impact of noise correlation in next-generation gravitational wave detectors*, *Phys. Rev. D* **111** (2025) 044046 [[2407.08728](#)].
- [158] I. Caporali, G. Capurri, W. Del Pozzo, A. Ricciardone and L. Valbusa Dall’Armi, *Impact of correlated noise on the reconstruction of the stochastic gravitational wave background with the Einstein Telescope*, *Phys. Rev. D* **112** (2025) 022005 [[2501.09057](#)].
- [159] A. Ghosh, N.K. Johnson-McDaniel, A. Ghosh, C.K. Mishra, P. Ajith, W.D. Pozzo et al., *Testing general relativity using gravitational wave signals from the inspiral, merger and ringdown of binary black holes*, *Classical and Quantum Gravity* **35** (2017) 014002.
- [160] L. Pompili et al., *Laying the foundation of the effective-one-body waveform models SEOBNRv5: Improved accuracy and efficiency for spinning nonprecessing binary black holes*, *Phys. Rev. D* **108** (2023) 124035 [[2303.18039](#)].
- [161] A. Ramos-Buades, A. Buonanno, H. Estellés, M. Khalil, D.P. Mihaylov, S. Ossokine et al., *Next generation of accurate and efficient multipolar precessing-spin effective-one-body waveforms for binary black holes*, *Phys. Rev. D* **108** (2023) 124037 [[2303.18046](#)].
- [162] M. van de Meent, A. Buonanno, D.P. Mihaylov, S. Ossokine, L. Pompili, N. Warburton et al., *Enhancing the SEOBNRv5 effective-one-body waveform model with second-order gravitational self-force fluxes*, *Phys. Rev. D* **108** (2023) 124038 [[2303.18026](#)].
- [163] A. Gamboa, M. Khalil and A. Buonanno, *Third post-Newtonian dynamics for eccentric orbits and aligned spins in the effective-one-body waveform model seobnrv5ehm*, *Phys. Rev. D* **112** (2025) 044037 [[2412.12831](#)].
- [164] S. Albanesi, R. Gamba, S. Bernuzzi, J. Fontbuté, A. Gonzalez and A. Nagar, *Effective-one-body modeling for generic compact binaries with arbitrary orbits*, [2503.14580](#).
- [165] H. Estellés, A. Buonanno, R. Enficiaud, C. Foo and L. Pompili, *Adding equatorial-asymmetric effects for spin-precessing binaries into the SEOBNRv5PHM waveform model*, [2506.19911](#).
- [166] G. Pratten, S. Husa, C. Garcia-Quiros, M. Colleoni, A. Ramos-Buades, H. Estelles et al., *Setting the cornerstone for a family of models for gravitational waves from compact binaries: The dominant harmonic for nonprecessing quasicircular black holes*, *Phys. Rev. D* **102** (2020) 064001 [[2001.11412](#)].
- [167] G. Pratten et al., *Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes*, *Phys. Rev. D* **103** (2021) 104056 [[2004.06503](#)].

- [168] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades et al., *Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries*, *Phys. Rev. D* **102** (2020) 064002 [[2001.10914](#)].
- [169] H. Estellés, M. Colleoni, C. García-Quirós, S. Husa, D. Keitel, M. Mateu-Lucena et al., *New twists in compact binary waveform modeling: A fast time-domain model for precession*, *Phys. Rev. D* **105** (2022) 084040 [[2105.05872](#)].
- [170] M.d.L. Planas, A. Ramos-Buades, C. García-Quirós, H. Estellés, S. Husa and M. Haney, *Time-domain phenomenological multipolar waveforms for aligned-spin binary black holes in elliptical orbits*, [2503.13062](#).
- [171] E. Hamilton et al., *PhenomXPNR: An improved gravitational wave model linking precessing inspirals and NR-calibrated merger-ringdown*, [2507.02604](#).
- [172] M. Garg, L. Sberna, L. Speri, F. Duque and J. Gair, *Systematics in tests of general relativity using LISA massive black hole binaries*, *Mon. Not. Roy. Astron. Soc.* **535** (2024) 3283 [[2410.02910](#)].
- [173] E. Payne, M. Isi, K. Chatzioannou and W.M. Farr, *Fortifying gravitational-wave tests of general relativity against astrophysical assumptions*, *Phys. Rev. D* **108** (2023) 124060 [[2309.04528](#)].
- [174] P. Mahapatra et al., *Confronting General Relativity with Principal Component Analysis: Simulations and Results from GWTC-3 Events*, [2508.06862](#).
- [175] C.M. Will, *Bounding the mass of the graviton using gravitational wave observations of inspiralling compact binaries*, *Phys. Rev. D* **57** (1998) 2061 [[gr-qc/9709011](#)].
- [176] M. Johnston, M. Kalomenopoulos and C.-J. Haster, *An analysis and cosmological study of the propagation of gravitational radiation under a graviton of nonzero mass*, [2509.18358](#).
- [177] C. de Rham, *Massive Gravity*, *Living Rev. Rel.* **17** (2014) 7 [[1401.4173](#)].
- [178] KAGRA, VIRGO, LIGO SCIENTIFIC collaboration, *Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3*, *Phys. Rev. X* **13** (2023) 011048 [[2111.03634](#)].
- [179] E.E. Flanagan, *The Sensitivity of the laser interferometer gravitational wave observatory (LIGO) to a stochastic background, and its dependence on the detector orientations*, *Phys. Rev. D* **48** (1993) 2389 [[astro-ph/9305029](#)].
- [180] N. Christensen, *Optimal detection strategies for measuring the stochastic gravitational radiation background with laser interferometric antennas*, *Phys. Rev. D* **55** (1997) 448.
- [181] L.S. Collaboration, V. Collaboration and K. Collaboration, *Gwtc-3: Compact binary coalescences observed by ligo and virgo during the second part of the third observing run — data behind the figures*, Oct., 2023. 10.5281/zenodo.7997424.
- [182] ALIGO collaboration, *Sensitivity and performance of the Advanced LIGO detectors in the third observing run*, *Phys. Rev. D* **102** (2020) 062003 [[2008.01301](#)].
- [183] M. Tse et al., *Quantum-Enhanced Advanced LIGO Detectors in the Era of Gravitational-Wave Astronomy*, *Phys. Rev. Lett.* **123** (2019) 231107.
- [184] VIRGO collaboration, *Increasing the Astrophysical Reach of the Advanced Virgo Detector via the Application of Squeezed Vacuum States of Light*, *Phys. Rev. Lett.* **123** (2019) 231108.
- [185] VIRGO collaboration, *Virgo detector characterization and data quality: results from the O3 run*, *Class. Quant. Grav.* **40** (2023) 185006 [[2210.15633](#)].
- [186] S. Hild et al., *Sensitivity Studies for Third-Generation Gravitational Wave Observatories*, *Class. Quant. Grav.* **28** (2011) 094013 [[1012.0908](#)].
- [187] P. Madau and M. Dickinson, *Cosmic Star Formation History*, *Ann. Rev. Astron. Astrophys.* **52** (2014) 415 [[1403.0007](#)].
- [188] M. Mapelli, N. Giacobbo, E. Ripamonti and M. Spera, *The cosmic merger rate of stellar black hole binaries from the Illustris simulation*, *Mon. Not. Roy. Astron. Soc.* **472** (2017) 2422 [[1708.05722](#)].
- [189] J.T. Whelan, *The geometry of gravitational wave detection*, *en. In:()* (2013) 15.
- [190] A. Abac et al., *The Science of the Einstein Telescope*, [2503.12263](#).

- [191] A. Puecher, C. Kalaghatgi, S. Roy, Y. Setyawati, I. Gupta, B.S. Sathyaprakash et al., *Testing general relativity using higher-order modes of gravitational waves from binary black holes*, *Phys. Rev. D* **106** (2022) 082003 [[2205.09062](#)].