Inspiral tests of general relativity and waveform geometry

Brian C. Seymour,[1,2] Jacob Golomb,[1] and Yanbei Chen[1]
[1]*TAPIR, Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, California 91125, USA*
[2]*Center of Gravity, Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark*

(Received 19 February 2026; accepted 8 June 2026; published 22 July 2026)

The phase evolution of gravitational waves encodes critical information about the orbital dynamics of binary systems. In this work, we test the robustness of parametrized tests against unmodeled deviations from general relativity. We demonstrate that these parametrized tests are flexible and sensitive in detecting generic deviations in the waveform using the Cutler-Vallisneri bias formalism. This universality arises from examining the inherent geometry of the waveform signal and understanding how biases manifest. We show how Bayes factors are governed by the intrinsic geometry of the waveform signal manifold when parametrized tests are used to approximate generic violations of GR. We use the singular value decomposition to propose templates that are orthogonal to parametrized tests, identifying degeneracies and enhancing the detection of potential deviations. More broadly, the geometric framework developed here clarifies—at a fundamental level—how subtle waveform effects (including orbital eccentricity, spin precession, waveform systematics, and instrumental glitches) can mimic one another in data, and when they are intrinsically distinguishable.

DOI: 10.1103/nw3l-qld4

I. INTRODUCTION

The detections of gravitational waves (GW) from compact binary coalescences [1–5] have provided new ways to directly test the behavior of strong gravity in general relativity (GR). Despite many searches in LIGO-Virgo-KAGRA (LVK) data, no evidence for deviations has been found [6–13]. The network currently consists of two LIGO interferometers [14], Virgo [15], KAGRA [16], and eventually LIGO India [17,18]. To maximize the scientific reach of these current detectors, there are proposals for A# [19] and LIGO Voyager [20,21]. Third-generation ground-based facilities—Cosmic Explorer [22,23] and the Einstein Telescope [24–26]—are in the planning stages, which are expected to improve the sensitivity compared to LIGO A+ by an order of magnitude. There are a number of space-based detectors planned of which LISA is expected to launch in 2035 [27,28]. There are other concepts including TianQin [29], Taiji [30,31], B-DECIGO/DECIGO [32–34], and TianGO [35,36].

One of the most prominent techniques for searching for deviations to GR is the parametrized post-Einsteinian (ppE) framework [37–42]. This formalism searches for deviations to the phase of the waveform that appear at particular orders in velocity of the waveform. This framework was also extended in a number of ways throughout the years, including additional polarizations [43], precessing waveforms [44], higher-order modes [45,46], eccentric signals [47], approaches for parametrized searches in the plunge-merger phase [48–51], and a neural post-Einstein framework [52]. It is noted that such an expansion in v/c does not work for all types of perturbations, namely logarithmic or screened terms [42,53]. It also needs the perturbations to the binding energy and GW fluxes to be able to be expressed as a factor of $(M/r)^{\alpha}$ which we have found is not possible for, e.g., nonviolent nonlocality that has an essential singularity [54].

These tests of GR are intimately connected to how well these beyond GR deviations are observable in data. It has been noted for a long time that waveform distinguishability [55] and thus correspondingly accuracy requirements [56–60] are tied to the geometry of the signal manifold [61–64]. Systematic bias to the waveform model directly manifests as biases to the measured values of the parameters [55–60,65–71]. Building upon approaches to understand how waveforms are biased due to unmodeled signals [65,66], Vallisneri found that the GR parameters aim to mask the beyond GR signal deviation [72,73] which is called "stealth bias" [73,74]. It is shown that it is only the perpendicular portion of the waveform that contributes to the detectability of GR violations. As one would expect, both systematics of the GR template waveform [75–84], external astrophysical environment [85–94], and even instrumental glitches [95–99] can show up as false positives deviations to GR.

In addition to this statistical hypothesis testing, eventually the community wants to find ways to better test GR than measuring a single parameter. If multiple ppE

 © 2026 American Physical Society

parameters are attempted to be measured for a particular waveform, the covariance between each of them means that constraints are markedly reduced. The singular value decomposition (SVD) approach to searching for deviations to GR was originally devised by Pai and Arun [100,101] which is an approach similar in spirit to surrogate modeling [102–106]. This approach used features of ppE deviations from GR to identify the features that are most common and most precisely measured [100,101]. Additionally, if one attempts to measure multiple ppE parameters in data, a principal component analysis (PCA) can be performed to identify which directions that the covariance matrix is best measured [107–115]. While the SVD and PCA differ due in origin from modeling with least parameters versus the statistical relationship of measuring multiple parameters, they are related to one another.

In this paper, we will build upon the existing literature and relate the tests of GR to the geometry of the signal manifold from the noise-weighted inner product. We will begin by identifying how GR parameters are biased if a beyond GR signal [72] is introduced and show how the ppE formalism can capture generic deviations remarkably well due to the behavior of the geometrical picture. We will show how multiparameter tests of GR are difficult because the ppE deviations to GR have similar features when orthogonalized away from GR. Finally, we will build upon the work of Pai *et al.* [100,101] and introduce a new form of the SVD that identifies the common features of the ppE tests. Throughout, we will use geometric units where $c = G = 1$. The Python notebooks and environment used to generate the main figures of this paper is given in the Github repository WAVEFORM-GEOMETRY-TESTING-GR [116].

II. BACKGROUND AND METHODS

This analysis is motivated by work done on calibration errors in a waveform using the Cutler-Vallisneri bias formalism [65,66]. In Fig. 1, we show a visualization of how the true GR parameters $\boldsymbol{\theta}_{\rm t}$ are biased due to a beyond GR signal $h_{\rm s}(\boldsymbol{\theta}_{\rm t}, \boldsymbol{\lambda}_{\rm t})$ (blue) where $\boldsymbol{\lambda}_{\rm t}$ are the beyond GR parameters. The possible values of the model waveform $h_{\rm m}(\boldsymbol{\theta}_{\rm t})$ are shown in the gray manifold while the best fitting parameter is $\boldsymbol{\theta}_{\rm ML}$. In this section, we will derive how the GR parameters are biased replicating [65,66] and show how the Bayes factor depends strongly on the residual deviation from GR $\Delta h^{\perp}$ as found in [72,73].

A. Review of biased parameter estimation

Suppose we have a true waveform that is $s_t \equiv h_{\rm s}(\boldsymbol{\theta}_t, \boldsymbol{\lambda}_t)$ and we are attempting to recover it with a model $h_{\rm m}(\boldsymbol{\theta})$. The likelihood in GR is just the true waveform plus Gaussian noise

$$d = s_t + n, \tag{1}$$

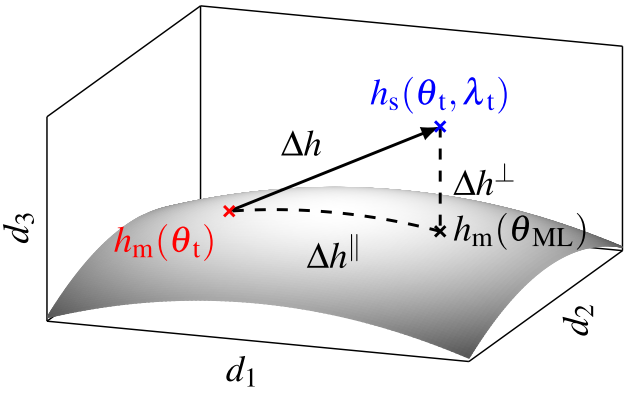

FIG. 1. Illustration of degeneracy when testing GR. We show the injected signal (blue) which depends on the true GR parameters $\boldsymbol{\theta}_{\rm t}$ and the beyond GR parameters $\boldsymbol{\lambda}_{\rm t}$. The model signal at the true GR parameters $\boldsymbol{\theta}_{\rm t}$ (red) is shown and the best fit signal is at the maximum likelihood point $\boldsymbol{\theta}_{\rm ML}$ (black). The GR waveform is modified by Δh which causes biases to the GR waveform, thus residual signal to measure beyond GR deviations is given by the perpendicular signal $\Delta h^{\perp}$. Note that this is a high dimensional manifold where (d_1, d_2, d_3) are the values of the signal at particular frequency bins.

while our template is $h_{\rm m}(\boldsymbol{\theta})$. For stationary and Gaussian noise the likelihood for a waveform is equal to

$$\log L(\boldsymbol{\theta}) \propto -\frac{1}{2}(d - h_{\rm m}(\boldsymbol{\theta})|d - h_{\rm m}(\boldsymbol{\theta})), \tag{2}$$

where the noise-weighted inner product is defined as

$$(a|b) \equiv 4{\rm Re}\int_0^{\infty} \frac{a^*(f)b(f)}{S_n(f)}df, \tag{3}$$

where $S_n(f)$ is the power spectral density (PSD) of the detector.

We wish to find the Fisher information about the point to characterize detectability [55,65,117–119]. The maximum likelihood estimator is found where the derivative of the likelihood is zero

$$\partial_i \log L|_{\boldsymbol{\theta}_{\rm ML}} = (\partial_i h_{\rm m}(\boldsymbol{\theta})|d - h_{\rm m}(\boldsymbol{\theta}))|_{\boldsymbol{\theta}_{\rm ML}} = 0. \tag{4}$$

In general, this is a nonlinear equation to solve for $\boldsymbol{\theta}_{\rm ML}$, but we make the standard assumption that it is sharply peaked in $\boldsymbol{\theta}$ so that $\boldsymbol{\theta}_{\rm ML} - \boldsymbol{\theta}_{\rm t} \ll 1$. Let us define Δh as

$$\Delta h = h_{\rm s}(\boldsymbol{\theta}_{\rm t}, \boldsymbol{\lambda}_{\rm t}) - h_{\rm m}(\boldsymbol{\theta}_{\rm t}). \tag{5}$$

We also assume that Δh is small.[1] Therefore, we can use $d - h_{\rm m}(\boldsymbol{\theta}) = n + \Delta h + \Delta\theta^i \partial_i h_{\rm m}$. The maximum

[1]Note that stability of this equation is poor for actual numerical implementation, especially when $\Delta h_{\parallel} \gg \Delta h_{\perp}$. Appendix A discusses how this is rectified by modifying the form of Δh used in (8).

likelihood estimate is thus equal to

$$\theta^i_{\rm ML} = \theta^i_{\rm t} + \Delta\theta^i_{\rm bias} + \Delta\theta^i_{\rm stat}, \tag{6}$$

where the bias and statistical errors are equal to

$$\Delta\theta^i_{\rm stat} = \Sigma^{ij}(\partial_j h_{\rm m}|n), \tag{7}$$

$$\Delta\theta^i_{\rm bias} = \Sigma^{ij}(\partial_j h_{\rm m}|\Delta h), \tag{8}$$

where Σ^{ij} is the matrix inverse of the Fisher matrix $\Gamma_{ij} \equiv (\partial_i h_{\rm m}|\partial_j h_{\rm m})$. One can show that the covariance of the $\Delta\theta^i_{\rm stat}$ random variable is equal to the inverse of the Fisher matrix $E[\Delta\theta^i_{\rm stat}\Delta\theta^j_{\rm stat}] \equiv \Sigma^{ij} = (\Gamma^{-1})^{ij}$ [117,120]. We emphasize that the $\boldsymbol{\Delta\theta}_{\rm bias}$ comes from the mismodeling of the waveform and is independent of SNR.

Suppose that we are not in the nested case as described above, where our modeled waveform actually depends on some other parameters which are not in the signal. In this case, we model it as $h_{\rm m}(\boldsymbol{\theta},\boldsymbol{\mu})$ where $\boldsymbol{\mu}$ are some small parameters which we want to measure in the waveform. In this case, we can perform the same analysis as above. If we define $\Theta^I = (\boldsymbol{\theta},\boldsymbol{\mu})$ as the full parameter set, the maximum likelihood estimate is

$$\Theta^I_{\rm ML} = \Theta^I_{\rm t} + \Delta\Theta^I_{\rm stat} + \Delta\Theta^I_{\rm bias}, \tag{9}$$

where $\Theta^I_{\rm t} = (\boldsymbol{\theta}_{\rm t}, \mathbf{0})$ so we are assuming the true value of $\boldsymbol{\mu}_{\rm t} \sim 0$. The statistical error is the same $\Delta\Theta^I_{\rm stat} = \Sigma^{IJ}(\partial_J h_{\rm m}|n)$ while the bias is

$$\Delta\Theta^I_{\rm bias} = \Sigma^{IJ}(\partial_J h_{\rm m}|\Delta h). \tag{10}$$

We emphasize that at high SNR, our small model parameters will scale like

$$\boldsymbol{\mu}_{\rm ML} = \Delta\boldsymbol{\mu}_{\rm bias}(\boldsymbol{\theta}_{\rm t},\boldsymbol{\lambda}_{\rm t}) + \Delta\boldsymbol{\mu}_{\rm stat}, \tag{11}$$

where the bias term is linearly proportional to the $\boldsymbol{\lambda}_{\rm t}$ parameter $\Delta\boldsymbol{\mu}_{\rm bias} \propto \boldsymbol{\lambda}_{\rm t}$. This key fact will be used to explain the unreasonable effectiveness of ppE templates for testing GR. If you imagine that there is a generic phase deviation in the true waveform $\Delta\Psi_{\rm t}$, this will be accessible if we measure any $\delta\varphi_k$ since it is responsive to any modification to the waveform.

B. Geometric interpretation of bias equation

In the previous section, we reviewed how an injection can bias the parameter estimation in a mismodeled waveform. Now, we will look at Eq. (10) and discover how it can be understood more deeply by thinking about the geometry of the waveform. We define the parallel and perpendicular components as

$$\Delta h_{\parallel} = \Delta\Theta^J_{\rm bias}\partial_J h_{\rm m},$$
$$\Delta h_{\perp} = \Delta h - \Delta\Theta^J_{\rm bias}\partial_J h_{\rm m}. \tag{12}$$

Using these identities, one can see that Eq. (10) becomes

$$\Delta\Theta^I_{\rm bias} = \Sigma^{IJ}(\partial_J h_{\rm m}|\Delta h_{\parallel}). \tag{13}$$

This demonstrates that only deviations parallel to waveform manifold bias the parameter estimation.

Let us now investigate the case that we have a waveform with a bias that is due to a small parameter $\boldsymbol{\lambda}$ and our model has a small parameter $\boldsymbol{\mu}$. For simplicity assume that $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are both one dimensional parameters. Then, one can show that the estimate for μ is

$$\Delta\mu_{\rm bias} = \frac{((\partial_\mu h_{\rm m})^{\perp\boldsymbol{\theta}}|(\Delta h)^{\perp\boldsymbol{\theta}})}{\|(\partial_\mu h_{\rm m})^{\perp\boldsymbol{\theta}}\|^2}, \tag{14}$$

where the $\perp\boldsymbol{\theta}$ denotes removing the part of the signal which is parallel to the main $\boldsymbol{\theta}$ parameters. Thus, we see evidence for $\boldsymbol{\mu}$ when both the waveform $h_{\rm m}(\boldsymbol{\theta},\boldsymbol{\mu})$ and the residual $\Delta h(\boldsymbol{\theta}_t,\boldsymbol{\lambda})$ cannot eliminate the effect by a redefinition of $\boldsymbol{\theta}$.

In Eq. (14), we performed this calculation for the simplest case that you are trying to measure a single parameter μ. If $\boldsymbol{\mu} = \mu^a$ is an array of beyond GR parameters, the equation above can be generalized using Schur decomposition which yields

$$\Delta\mu^a_{\rm bias} = (\Gamma^{-1}_{\rm red})_{ab}((\partial_{\mu^b} h_{\rm m})^{\perp\boldsymbol{\theta}}|(\Delta h)^{\perp\boldsymbol{\theta}}), \tag{15}$$

where the reduced Fisher matrix is equal to

$$\Gamma^{\rm red}_{ab} = ((\partial_{\mu^a} h_{\rm m})^{\perp\boldsymbol{\theta}}|(\partial_{\mu^b} h_{\rm m})^{\perp\boldsymbol{\theta}}). \tag{16}$$

One can see in the single μ case the equation reduces to Eq. (14) since $\Gamma^{\rm red}_{\mu\mu} = \|(\partial_\mu h_{\rm m})^{\perp\boldsymbol{\theta}}\|^2$. In Appendix B, we derive Eq. (15) using the Schur decomposition identities for a matrix consisting of blocks.

C. Multidetector geometry

The manifold language introduced in this chapter straightforwardly generalizes to the case of multiple detectors observing one GW event. The likelihood for parameters is

$$p(d|\boldsymbol{\theta},\boldsymbol{\mu}) \propto \exp\left[-\frac{1}{2}\sum_A \|d_A - h^A_{\rm m}(\boldsymbol{\theta},\boldsymbol{\mu})\|^2\right]. \tag{17}$$

It is useful to introduce the summed network inner product and use square brackets to represent this

$$[a|b] = \sum_A (a_A|b_A)_A, \tag{18}$$

then the maximum likelihood point is the solution to

$$\Gamma_{IJ}\Delta\Theta^J = [\partial_I h_{\rm m}|\Delta h] + [\partial_I h_{\rm m}|n], \tag{19}$$

where the network Fisher information matrix is $\Gamma_{IJ} \equiv [\partial_I h_{\rm m}|\partial_J h_{\rm m}]$. The equations for the statistical error and bias straightforwardly generalize from Eq. (7) and are equal to

$$\Delta\Theta^I_{\rm stat} = (\Gamma^{-1})^{IJ}[\partial_I h_{\rm m}|n], \tag{20}$$

$$\Delta\Theta^I_{\rm bias} = (\Gamma^{-1})^{IJ}[\partial_I h_{\rm m}|\Delta h]. \tag{21}$$

Henceforth, we will continue to do calculations as if they were a single detector and use the normal inner product $(\cdot|\cdot)$ notation, but the geometric description of testing GR in multiple detectors follows the same paradigm. The explicit antenna-response and polarization structure is given in Appendix D.

D. Bayes factors for tests of GR

We wish to compare whether there is support for beyond GR (bGR) signals in the data. Vallisneri [72,73] showed how the Bayes factor provides support for the bGR signal and identified the geometric meaning. If we have two hypotheses (a) GR ($\boldsymbol{\theta}$) and (b) bGR ($\boldsymbol{\theta}, \boldsymbol{\lambda}$), the Bayes factor is equal to

$$\mathcal{B}^{\rm bGR}_{\rm GR} \equiv \frac{p(d|{\rm bGR})}{p(d|{\rm GR})}, \tag{22}$$

where the evidence of data d in model $\mathcal{M}$ with parameters $\boldsymbol{\Theta}$ is

$$p(d|\mathcal{M}) = \int d\boldsymbol{\Theta} p(d|\boldsymbol{\Theta}, \mathcal{M}) p(\boldsymbol{\Theta}), \tag{23}$$

where $p(\boldsymbol{\Theta})$ is the prior and $p(d|\boldsymbol{\Theta}, \mathcal{M})$ is the likelihood. Vallisneri computed that the Bayes factor is

$$\mathcal{B}^{\rm bGR}_{\rm GR}|_{s_{\rm bGR}} = \frac{(2\pi)^{1/2}\Delta\lambda_{\rm stat}}{\Delta\lambda_{\rm prior}} e^{\rho_\perp^2/2 + x\rho_\perp + x^2/2}, \tag{24}$$

where $x \equiv (\Delta h_\perp|n)/\|\Delta h_\perp\| \sim \mathcal{N}(0,1)$ is a standard normal random variable, $\Delta\lambda_{\rm stat}$ is the statistical error on $\boldsymbol{\lambda}$, and we assumed a flat prior with width $\Delta\lambda_{\rm prior}$. The residual SNR is

$$\rho_\perp \equiv \|\Delta h_\perp\| \tag{25}$$

and $\Delta h_\perp$ is the *true* perpendicular signal to the GR waveform as defined in Eqs. (5) and (12) due to $\boldsymbol{\lambda}_t$. One can see that the residual SNR is what provides evidence for a GR violation. If the true signal is a GR signal, the Bayes factor is just

$$\mathcal{B}^{\rm bGR}_{\rm GR}|_{s_{\rm GR}} = \frac{(2\pi)^{1/2}\Delta\lambda_{\rm stat}}{\Delta\lambda_{\rm prior}} e^{x^2/2}, \tag{26}$$

which has a randomly distributed exponent $x^2/2$ and an *Occam factor* prefactor $\propto \Delta\lambda_{\rm stat}/\Delta\lambda_{\rm prior}$ that will cause it to favor GR on average.

Suppose we searched for deviations from GR with post-Newtonian (PN) deviations from GR in the formalism of the parametrized post-Einsteinian/TIGER/FTI tests of GR [37–41,46] which mismodeled the bGR waveform. We will now generalize the Bayes factor formula in this case. We assume that the injected signal is $s^t_{\rm bGR} = h(\boldsymbol{\theta}_{\rm t}, \boldsymbol{\lambda}_{\rm t})$. We will compare two hypotheses (a) GR ($\boldsymbol{\theta}$) and (b) ppE ($\boldsymbol{\theta}, \boldsymbol{\mu}$). Careful computation reveals that the Bayes factor is

$$\mathcal{B}^{\rm ppE}_{\rm GR}\Big|_{s_{\rm bGR}} = \frac{(2\pi)^{1/2}\Delta\mu_{\rm stat}}{\Delta\mu_{\rm prior}} e^{(\rho^{\rm ppE}_\perp)^2/2 + y\rho^{\rm ppE}_\perp + y^2/2}, \tag{27}$$

where $y \equiv (\Delta h^\perp_{\rm ppE}|n)/\|\Delta h^\perp_{\rm ppE}\|$ is a unit normal Gaussian random variable and the captured residual SNR $\rho^{\rm ppE}_\perp$ is defined via

$$\rho^{\rm ppE}_\perp = \mathcal{O}\rho_\perp, \tag{28}$$

which depends on the overlap of the bGR and ppE waveform deviations. The overlap is equal to

$$\mathcal{O}(\Delta h^\perp_{\rm ppE}, \Delta h^\perp_{\rm bGR}) = \frac{(\Delta h^\perp_{\rm ppE}|\Delta h^\perp_{\rm bGR})}{\|\Delta h^\perp_{\rm ppE}\|\|\Delta h^\perp_{\rm bGR}\|}. \tag{29}$$

Note that this overlap is between 0 and 1, and quantifies how similar two signals are. We use notation $\Delta h^\perp_{\rm ppE} \equiv (\Delta h^{\rm ML}_{\rm ppE})^\perp = \Delta\mu_{\rm bias}(\partial_\mu h_{\rm m})^\perp$ where $\Delta\mu_{\rm bias}$ is the best fit for the ppE test as a function of $\Delta h^\perp_{\rm bGR}$. The lost residual SNR is then equal to

$$\begin{aligned}\rho^{\rm diff}_\perp &\equiv \|\Delta h^\perp_{\rm bGR} - \Delta h^\perp_{\rm ppE}\|, \\ &= \rho_\perp\sqrt{1-\mathcal{O}^2}.\end{aligned} \tag{30}$$

In Appendix C, we rederive the optimal Bayes factor in Eq. (24) using methods from [72] and generalize it to the parametrized test Bayes factor in Eq. (27). One can see that the geometric structure comes from features related to how Fisher matrices invert when you have blocks using the Schur decomposition.

Finally, if we were to detect deviations from GR using, e.g., one of the parametrized tests, it is important to see how well we could distinguish a bGR model from a ppE model. This would be important in the hypothetical case that we need to identify which particular bGR models are consistent with the data and favored. We show that the Bayes factor between bGR ($\boldsymbol{\theta}, \boldsymbol{\lambda}$) and ppE ($\boldsymbol{\theta}, \boldsymbol{\mu}$) is equal to

$$\mathcal{B}^{\rm bGR}_{\rm ppE}|_{s_{\rm bGR}} = \frac{\Delta\lambda_{\rm stat}\Delta\mu_{\rm prior}}{\Delta\lambda_{\rm prior}\Delta\mu_{\rm stat}} e^{(x+\rho_\perp)^2/2-(y+\mathcal{O}\rho_\perp)^2/2}, \tag{31}$$

where the x, y are unit normal Gaussian random variables as defined below Eqs. (24) and (27). Note that $E[xy] = \sqrt{1-\mathcal{O}^2}$ so these variables are correlated with one another. Additionally, we note that the fidelity at which we can distinguish between bGR and ppE is determined by the quantity $\rho^{\rm diff}_\perp$ that we derived in Eq. (30). This expression is fully derived in Appendix C, but it involves using the fact that Bayes factor between bGR and ppE in Eq. (31) can be expressed as a ratio of the Bayes factors in Eqs. (24) and (27). We also note that there is no simple expression for this Bayes factor, but it can be written as a sum of two uncorrelated random variables as we detail in Eq. (C27) of Appendix C 3.

In the case of high residual SNR, we can express Eq. (31) more simply and provide a z-score for how much the theory deviates. If we take the limit that $\rho_\perp \gg 1$, the difference between the chi-squared random variables can be expressed in terms of unit normal Gaussian random variable $z \sim \mathcal{N}(0,1)$

$$\log \mathcal{B}^{\rm bGR}_{\rm ppE}|_{s_{\rm bGR}} \approx \mathcal{C} + \frac{(\rho^{\rm diff}_\perp)^2}{2} + \sqrt{1+\mathcal{O}^2-2\mathcal{O}\sqrt{1-\mathcal{O}^2}}\rho_\perp z, \tag{32}$$

where the normal approximation was used for the chi-squared variables and $\mathcal{C}$ is the constant coming from prior terms of Eq. (31). One can see that the z-score for the terms in the exponential is

$$Z = \frac{1-\mathcal{O}^2}{2\sqrt{1+\mathcal{O}^2-2\mathcal{O}\sqrt{1-\mathcal{O}^2}}}\rho_\perp, \tag{33}$$

which characterizes the degree of preference for bGR vs ppE in the presence of noise. This example demonstrates that the smaller the overlap is, the easier it is to discern two signals while the significance scales linearly with the residual SNR.

In Fig. 2, we illustrate the geometric properties of the Bayes factor derivations that we have done. The GR waveform manifold is represented by the black line between the GR + ppE and GR + bGR manifolds (blue and red, respectively). The bGR signal (red x) is injected at the true parameter values $(\boldsymbol{\theta}_{\rm t}, \boldsymbol{\lambda}_{\rm t})$. We also show the value of the GR waveform at $h_{\rm GR}(\boldsymbol{\theta}_{\rm t})$ at true parameters and also its value which happens at the maximum likelihood point $h_{\rm GR}(\boldsymbol{\theta}_{\rm ML})$ as computed in Eq. (6) where we ignore the statistical noise. We can see that the maximum likelihood point in the GR manifold minimizes the value of $\rho_\perp$.

We also show the best fit value of the waveform if our recovery model is GR + ppE. This will peak at the location $h^{\rm ML}_{\rm ppE} = h^{\rm ML}_{\rm GR} + \Delta h^\perp_{\rm ppE}$ which is the point which minimizes

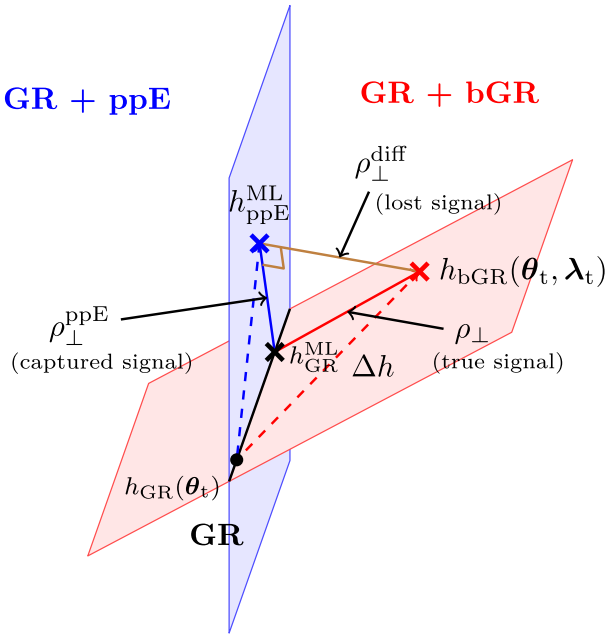

FIG. 2. In this plot, we visually show how the ppE tests of GR can capture generic bGR deviations. The GR manifold (black) is a line that is at the intersection between the true bGR manifold (red) and the ppE manifold (blue). One can see that the perpendicular part of the signal from GR is $\rho_\perp = \|\Delta h_\perp\|$ which is the residual SNR after allowing the GR parameters to be biased $h^{\rm ML}_{\rm GR} = h_{\rm GR}(\theta_{\rm t} + \Delta\theta_{\rm bias})$. One can see that the best fit ppE parameter is located at the blue mark and has residual SNR $\rho^{\rm ppE}_\perp = \mathcal{O}\rho_\perp$ as given in Eq. (28). Finally, the brown line is the missed signal from our ppE model. One can see that the best fit for the ppE model corresponds to a value such that the residual (brown line) is perpendicular to the ppE manifold. With this picture in mind, we can explore how well tests of GR capture generic deviations.

the distance from injected bGR signal $h_{\rm bGR}(\boldsymbol{\theta}_{\rm t}, \boldsymbol{\lambda}_{\rm t})$. The distance between the best fit value of $h_{\rm ppE}(\boldsymbol{\theta}_{\rm ML}, \boldsymbol{\mu}_{\rm ML})$ and $h_{\rm bGR}(\boldsymbol{\theta}_{\rm t}, \boldsymbol{\lambda}_{\rm t})$ is equal to the signal which was lost by the ppE model. The size of this is given by the length of the brown line and size $\rho^{\rm diff}_\perp$ in Fig. 2. In fact, as we showed in Eqs. (31) and (32), the size of $\rho^{\rm diff}_\perp$ characterizes how well one can distinguish between our true bGR model and the ppE model. Also note that the brown line is perpendicular to the ppE manifold[2] as shown in Fig. 2 which physically makes sense because the best fit point optimizes the location where the blue manifold is closest to the red signal injection. This geometrical picture shows that the systematic error of using parametrized tests is primarily characterized by the overlap $\mathcal{O}$ between the ppE and bGR manifolds.

[2]Note that $(\Delta h^\perp_{\rm bGR} - \Delta h^\perp_{\rm ppE}|\Delta h^\perp_{\rm ppE}) = 0$ since it is a ML point. This can also be shown explicitly via $\|\Delta h^\perp_{\rm ppE}\| = \mathcal{O}\rho_\perp$ and $(\Delta h^\perp_{\rm bGR}|\Delta h^\perp_{\rm ppE}) = \mathcal{O}^2\rho^2_\perp$.

The fitting factor is a number that characterizes how good of a fit the waveform is, and is related to the residual SNR. The fitting factor is defined as

$$\mathrm{FF}_{\mathcal{M}} = \max_{\Delta\theta^i} \frac{(h_{\mathcal{M}} + \Delta h, h_{\mathcal{M}} + \Delta\theta^i \partial_i h)}{|h_{\mathcal{M}} + \Delta h| \cdot |h_{\mathcal{M}} + \Delta\theta^i \partial_i h|}, \tag{34}$$

where $\mathcal{M}$ is the model. If $\mathcal{M} = \mathrm{GR}$ and we want to see how similar a bGR signal is, the fitting factor is given by [72]

$$1 - \mathrm{FF}_{\mathrm{GR}} = \frac{1}{2} \frac{\rho_\perp^2}{\|h_{\mathrm{GR}}\|^2}. \tag{35}$$

For the case of a mismodeled bGR search with $\mathcal{M} = \mathrm{ppE}$, the fitting factor is now

$$1 - \mathrm{FF}_{\mathrm{ppE}} = \frac{1}{2} \frac{(\rho_\perp^{\mathrm{diff}})^2}{\|h_{\mathrm{GR}}\|^2}, \tag{36}$$

so the mismatch of mismodeling a bGR signal is small and depends directly proportionally to the overlap in Eq. (29).

III. PPE PARAMETERS CAPTURE GENERIC PHASE DEVIATIONS

In the previous section, we showed that the perpendicular signal is what dominates the test of GR. So far, we have laid the groundwork where we stressed the relationship between the statistical nature of the problem and how it relates to the intrinsic behavior of the frequency domain waveforms. We will now show what the residual of the ppE tests of GR look like in the frequency domain.

We will use the ppE framework to add PN deviations to the GR waveform phase. We will use the normalizations that the LVK Collaboration uses [10]. During the inspiral, the GR phase is

$$\Psi_{\mathrm{PN}}(f) = 2\pi f t_c - \varphi_c - \frac{\pi}{4} \tag{37}$$

$$+ \frac{3}{128\eta} (\pi\tilde{f})^{-5/3} \sum_{k=0}^{7} [\varphi_k + \varphi_{kl} \log(\pi\tilde{f})] (\pi\tilde{f})^{k/3}, \tag{38}$$

where t_c, φ_c are the time and phase of coalescence at the geocenter, η is the symmetric mass ratio, and $\tilde{f} = M(1+z)f$ is the dimensionless GW frequency. M is the source frame mass while the detector frame total mass is $(1+z)M$ which accounts for how the detector frame mass changes with redshift z. The GR phasing coefficients are φ_k and φ_{kl} which depend on the GR parameters $\boldsymbol{\theta}$ (see Ref. [121] for explicit expressions). In this work, we will focus on the nonlog terms and look for dephasing that occurs as

$$\Delta\Psi_k(f) = \frac{3}{128\eta} \delta\varphi_k \varphi_k (\pi\tilde{f})^{(k-5)/3}, \tag{39}$$

where $\delta\varphi_k$ is the kth order fractional deviation to the GR phase. The goal is to see how well this parametrized template can capture the effects of a generic bGR deviation that is parametrized like

$$\Delta\Psi_{\mathrm{bGR}}(f) = \lambda\psi_\lambda(f). \tag{40}$$

Both the parametrized test and true bGR waveform are assumed to be phase deviations in the frequency domain so the waveforms are

$$h(f) = h_{\mathrm{GR}}(f) e^{i\Delta\Psi(f)}, \tag{41}$$

so the nominal waveform discrepancy is

$$\Delta h(f) \approx i\Delta\Psi(f) h_{\mathrm{GR}}(f; \boldsymbol{\theta}_{\mathrm{t}}), \tag{42}$$

because the phase deviation to the waveform is small. In fact, using Eq. (42) as the definition for Δh is more numerically stable than Eq. (41) as is discussed in Sec. II.D of Ref. [65] and in Appendix A of this paper.

We can visualize how the parametrized test coefficients affect the waveform by plotting the nominal and residual deviations. On the left side of Fig. 3, we plot nominal Δh as shown from Eq. (42) using GR parameters given by GW150914. We normalize the value of $\delta\varphi_k$ so that the residual SNR has unit size $\|\Delta h_k^\perp\| = 1$. As we have shown throughout this paper, it is the perpendicular component of the waveform that really influences observation. On the right side of the plot we show the residual component of $\Delta h_\perp^k$. One can see that the amount of residual signal in the right panel is significantly reduced compared to the original signal in the left panel—i.e., $\Delta h_k^{\perp\mathrm{GR}} \ll \Delta h_k$. Additionally, the signals $\Delta h_k^{\perp\mathrm{GR}}$ are overlapping and look very similar to one another as can be seen on the right. This illustrates that there is a similarity between each perpendicular ppE test. This occurs because the GR parameters are biased—notably the masses (M, η) and spins $(\chi_{\mathrm{eff}}, \chi_{\mathrm{p}})$ compensate for this deviation. Eliminating the biases in the GR parameters leads to a marked suppression of the residuals compared to the nominal deviation Δh_k for a GW150914-like injection.

Figure 3 is difficult to interpret because we are used to seeing the phase plotted instead of the amplitude. If we instead look for the perpendicular phase, we can see how biased GR parameters change the effective power law behavior of the ppE tests. We can read off the phase in the following fashion

$$\Delta h_\perp = i\Delta\Psi_\perp(f) h_{\mathrm{GR}}(f), \tag{43}$$

where $\Delta\Psi_\perp(f)$ is the phase deviation after removing the GR biases. Note that in general, there could be a second

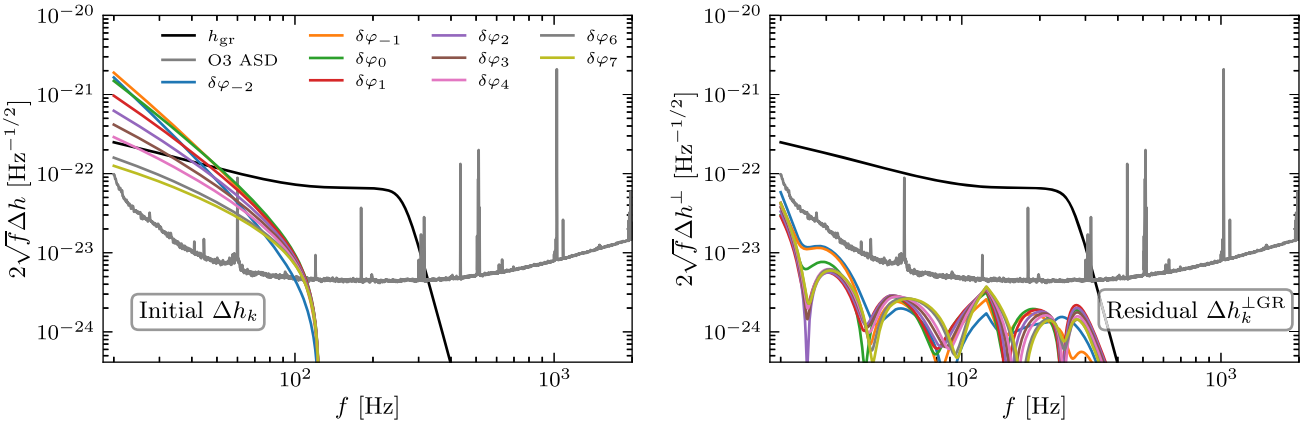

FIG. 3. Residual amplitude for ppE injected deviations from GR for a GW150914-like detection. On the left and right, we include the GR waveform (black) and O3 Livingston ASD (gray). On the left, we show the waveform residuals that would be caused by injection what we show in Fig. 4 from parametrized tests. On the right, we show what the residual deviation is after the stealth biases in the GR parameters are accounted for. This plot is made by tuning the ppE coefficients so that their residuals $\|\Delta h_k^{\perp GR}\| = 1$ are of unit size. This figure illustrates that the primary observable in these parametrized tests is the magnitude of the residual deviation (right), which is markedly reduced compared to the nominal deviations (left) [122].

term $\mathcal{O}(\Delta A \cdot h_{GR}(f))$ if we were adding additional amplitude corrections. In practice since our perturbations from GR are purely phase deviations, this has a negligible effect. In Fig. 4, we show the nominal phase deviations $\Delta\Psi_k$ (left) and the perpendicular phase deviations $\Delta\Psi_k^{\perp}$ (right). One can see that the power law behavior of the $\Delta\Psi \propto f^{\alpha}$ does not exist in the residuals because GR parameters are biased so it has an alternating pattern on the right.

A. Overlap between different ppE orders

In Sec. II D, we studied how good a parametrized template is at capturing the bGR waveform deviation and its relation to the Bayes factor. One can compute the overlap between two different ppE tests as given in Eq. (29).

We compute the overlap between different ppE tests of GR as a simple way to see how important mismodeling is for constraints on bGR theories. In Fig. 5, we compute the overlap between $\Delta h^{\perp}_{k_{inj}}$ injected signal residual and the $\Delta h^{\perp}_{k_{rec}}$ which is the perpendicular template. We compute this plot using $\boldsymbol{\theta} = (\boldsymbol{\theta}_{ext}, \boldsymbol{\theta}_{int}, \delta\varphi_k)$. $\boldsymbol{\theta}_{ext}$ is the extrinsic parameters

$$\boldsymbol{\theta}_{ext} = (d_L, \mathrm{ra}, \mathrm{dec}, \iota, \psi, t_c, \phi_c), \tag{44}$$

which are the luminosity distance, right ascension, declination, inclination, polarization, time of coalescence, and phase of coalescence, respectively. The intrinsic parameters are

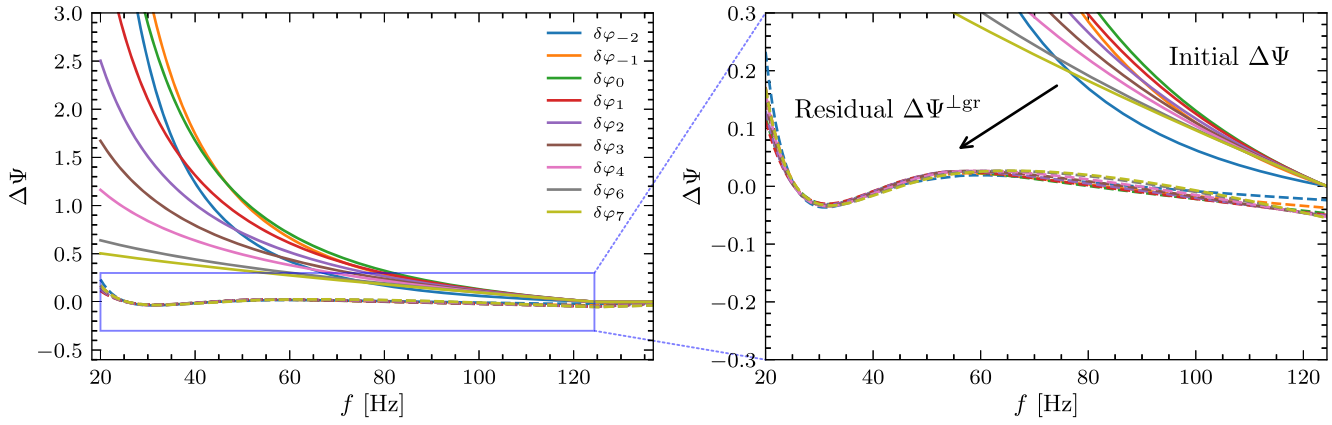

FIG. 4. This figure shows how the ppE behave after GR parameter projection. On the left, we show the ppE dephasing for a GW150914-like detection and with $\delta\varphi_k$ normalized equivalently to Fig. 3. On the right, we show the residual (perpendicular) phase deviation in the injection after the GR biases are taken into account. The total dephasing of the perpendicular waveforms is noticeably smaller with the residual phase deviations $\Delta\Psi^{\perp GR}$ displaying multiple zero crossings, akin to the behavior of an oscillatory polynomial. The near-identical appearance of the ppE dephasing residuals on the right visually demonstrates why leading-order deviations from GR appear similar to one another [122].

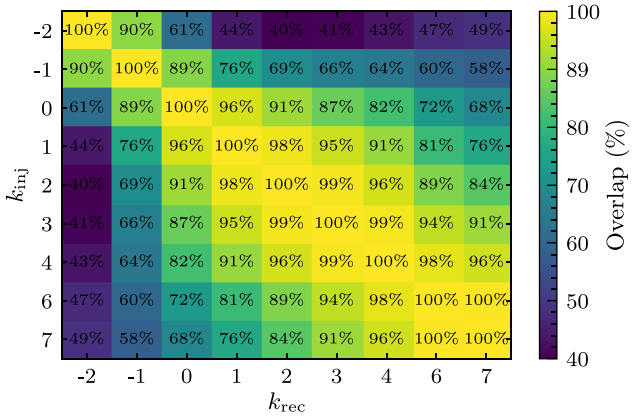

FIG. 5. The overlap between deviations injected (y-axis) and the recovery model (x-axis). We use GW150914 event parameters with three detectors at O3 Livingston sensitivity, all extrinsic parameters measured and intrinsic parameters $(\mathcal{M}_c, q, \chi_{\rm eff}, \chi_{\rm p})$. One can see that injecting a deviation at $k/2$-th PN order fractionally from GR is perfectly captured on the diagonal but has less overlap as the injection and recovery order grow. This means that if there is a true deviation, the significance drops slowly as you search for the wrong PN order $k_{\rm rec} \neq k_{\rm inj}$ because the intrinsic parameters will capture the difference. Note that we do not include $k = 5$ in either of these plots because it is nearly completely degenerate with phase of coalescence [125].

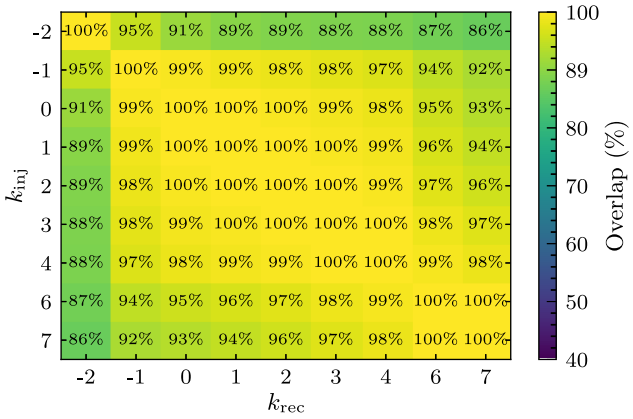

FIG. 6. The overlap between deviations injected (y-axis) and the recovery model (x-axis). We use GW150914 event parameters with three detectors at O3 Livingston sensitivity, all extrinsic parameters measured and intrinsic parameters $(\mathcal{M}_c, q)$. This can be directly compared to Fig. 5 which measures the parameters $(\chi_{\rm eff}, \chi_{\rm p})$ additionally. One can see that ignoring the parameter uncertainty of the spin parameters means that the overlap is higher for these waveforms since the spins cannot bias the waveforms [125].

$$\boldsymbol{\theta}_{\rm int} = (\mathcal{M}_c, \eta, \chi_{\rm eff}, \chi_{\rm p}), \tag{45}$$

which are chirp mass, symmetric mass ratio, effective inspiral spin parameter, and effective precession spin parameter, respectively. If the binary has total mass $M = m_1 + m_2$, the symmetric mass ratio is $\eta = m_1 m_2 / M^2$ and the chirp mass is $\mathcal{M}_c = \eta^{3/5} M$. The effective spin is $\chi_{\rm eff} \equiv (m_1 \chi_{1z} + m_2 \chi_{2z})/M$ where χ_{iz} is the z-component of the spin of body i [123], and effective precession spin $\chi_{\rm p}$ captures the in-plane spin dynamics as defined in Ref. [124]. We used GW150914 event parameters but use Hanford, Livingston, and Virgo network operating at O3 Livingston sensitivity. We compute the residual waveform via Eq. (12) for both the injected and recovery models and then plot the overlap in Fig. 5. One can see that if $k_{\rm rec} = k_{\rm inj}$, the overlap is 1 so there is no bias since we are searching with the same model as we are recovering with. In contrast, if we had a true deviation at $k_{\rm inj} = -2$ and tried to recover with $k_{\rm rec} = 7$, we would miss out in about 50% of the residual SNR. Note that the values of this plot strongly depend on the values of $\boldsymbol{\theta}_{\rm int}$. If we had picked a chirp mass which was in the binary neutron star (BNS) range $\mathcal{M}_c \sim 1.2 M_\odot$, this would change significantly. This case is further discussed in Appendix E where we show the overlap plots.

We also stress that the result of this overlap computation depends on which GR parameters are measured. If we take our intrinsic parameters to be just the masses

$$\boldsymbol{\theta}_{\rm int} = (\mathcal{M}_c, \eta), \tag{46}$$

then the size of the overlaps computed would be modified from what is shown Fig. 5. In Fig. 6, we show the result of this computation where we only allow the masses to be biased. You can see that including fewer GR parameters means that the overlaps between mismodeled ppE waveforms are much higher. The parametrized tests thus appear *fractionally* more similar when excluding spin biases—in the sense that the overlap is larger for different ppE tests. However note that the residual SNR $\rho^{\perp}$ is higher if you have fewer total GR parameters since the spins cannot compensate to reduce the residual GW strain.

The results of what we see in Figs. 5 and 6 highlights why it is difficult to do multiparameter tests of GR with the ppE parameters. Since their overlaps are so high, this means that the parameter combinations are highly degenerate with one another. This means that the errors are highly correlated and inflated when the Fisher matrix is inverted. In a later part of this manuscript Sec. IV, we will address how to distinguish the independent information in a data-driven way.

B. Capturing nonviolent nonlocality with ppE

In a previous paper [54], we studied the effects of nonviolent nonlocality on the waveform for stellar mass BBH systems. We found that the ppE parameters were pretty successful at capturing nonviolent nonlocality deviations from GR when stacking across many events. This result is not very surprising in light of what we have presented about the overlap between different ppE tests in

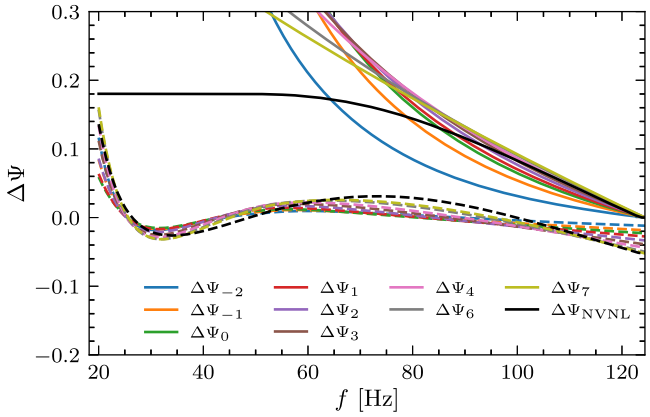

FIG. 7. The ppE phase deviations (colored) and the nonviolent nonlocality deviation (black) for both nominal $\Delta\Psi(f)$ (solid) and the perpendicular signal $\Delta\Psi^{\perp \mathrm{GR}}(f)$ (dashed). While the nominal dephasing differs significantly at low frequencies between the ppE templates and the injected nonviolent nonlocality deviation, this difference can be compensated by adjusting intrinsic parameters such as masses and spins. As a result, the perpendicular dephasing becomes very similar across cases. This plot demonstrates the flexibility of the ppE templates to capture dephasing patterns that are otherwise quite distinct, owing to the bias framework developed in this paper [126].

Sec. III A which for a GW150914-like signal the ppE tests have overlaps of at least 40%. In Ref. [54], we measured the stochastic effects of nonviolent nonlocality hierarchically over a large number of events and the constraint is about a factor of 30% better when we use the optimal waveform compared to just using the parametrized tests. In this section, we will explain why this is successful.

If we examine only the principal component of the stochastic deviation, one can show that nonviolent nonlocality has phase deviation that occurs at order

$$\Delta\Psi_{\mathrm{NVNL}}(f) \sim (\pi M f)^{\alpha} e^{-\frac{1}{8}(\pi M f)^{-4/3}} \tag{47}$$

where α is some generic power parameter.[3] Notice however that Eq. (47) power series has an essential singularity as $Mf \to 0$, so we cannot identify the "leading-order" ppE term. The exponential in Eq. (47) turns on too rapidly to identify such a leading order expression.

While Eq. (47) is manifestly not of the form of a power law in f as prescribed by the parametrized tests, we can use the bias framework developed in this paper to see how well nonviolent nonlocality deviations can be captured by parametrized tests. In Fig. 7, we show the nominal deviations for a GW150914-like event normalizing the waveform residual to one. One can see that while the nominal phase deviations between ppE (colored solid) are not very similar in shape to the nonviolent nonlocality phase deviation (black solid), the residual phase deviations (dashed) overlap and agree to a much larger extent. While nonviolent nonlocality exhibits nonpower-law behavior, the perpendicular phase deviations (dashed black) still shows a remarkable similarity to the ppE perpendicular deviations [$\Delta\Psi_{\perp}(f)$ as defined in Eq. (43)]. For the ppE dephasing, we plot the best fit values of $\delta\varphi_k$ for an injected nonviolent nonlocality deviation.

IV. SVD APPROACH

As we have seen in the previous sections, the ppE waveform deviations look relatively similar after removing the effects of GR biases. One can see that the right side of Fig. 4 has some properties that look similar across each PN order. The consequence of this is that attempting to measure multiple $\delta\varphi_k$ at once would make the covariance blow up and thus it is difficult to perform multiparameter tests of GR. Due to the allure of measuring multiple ppE parameters simultaneously, there has been a large amount of work in this area with the singular value decomposition (SVD) approach [100,101] and with the similar but distinct principal component analysis approach [54,107–110,127,128]. In this section, we will propose an improved SVD test of GR that builds upon the groundwork of the original Pai *et al.* results [100].

In this work, we want to find the common features of $\Delta h_k^{\perp}$ where each k is a different PN order deviation from GR. The SVD finds the best projector [106] such that

$$C(v_\alpha) = \sum_k \|\Delta h_k^{\perp} - P_n \Delta h_k^{\perp}\|^2, \tag{48}$$

where we assume that P_n is an orthogonal projector onto the SVD basis and $\|\cdot\|$ is a norm. The orthogonal projector is equal to

$$P_n h_a = \sum_{\alpha=1}^{n} (v_\alpha | h_a) v_\alpha, \tag{49}$$

and v_α are the SVD basis elements. These elements are orthonormal for the inner product

$$(v_\alpha | v_\beta) = \delta_{\alpha\beta}, \tag{50}$$

We will demonstrate that the SVD described by Pai *et al.* [100] is equivalent to minimizing Eq. (48). Thus, the SVD corresponds to finding an orthogonal basis that best projects all input waveforms.

Let us lay the groundwork for the SVD by noting properties of the noise-weighted inner product. In this section, we describe the general procedure to describe how to *pack* a frequency domain signal which has the symmetry $h(-f) = -h^*(f)$ corresponding to a real time domain signal. Note that the noise-weighted inner product is equal to

[3]This expression is implicitly in $\Delta\Psi(f)$ in Eq. (35) and (36) of [54]. Solving Eq. (35) would result in an equation which is some power in f times $f_{\ell m}(r(f))$ where $f_{\ell m}$ given in Eq. (2) of [54].

$$(h_1|h_2) = 4\mathrm{Re}\int_0^\infty df \frac{h_1^*(f)h_2(f)}{S(f)}, \tag{51}$$

If we break it up into finite differences, the noise-weighted inner product is

$$(h_1|h_2) = \sum_{i=0}^{N_f-1} \frac{h_1^*(f_i)h_2(f_i) + h_1(f_i)h_2^*(f_i)}{2} w^2(f_i), \tag{52}$$

where the weight is defined via

$$w(f_i) = \frac{2\sqrt{\Delta f_i}}{\sqrt{S(f_i)}}, \tag{53}$$

where the frequency spacing is $\Delta f_i = f_{i+1} - f_i$ and $S(f_i)$ is the value of the PSD at frequency bin i. Note that the discrete inner product in Eq. (52) does not have the form of the Hermitian inner product that we know and expect. This is just due to the fact that we simplified the original inner product integral in Eq. (51) using the fact that the time domain signals are real. Therefore, we need to construct an inner product where it is properly packed so that it is of the form that the SVD theorem applies.[4] Let us introduce the packing operator $\mathcal{P}$ as

$$\mathcal{P}[\Delta h] = \begin{bmatrix} \mathrm{Re}\,\Delta h(f_0)w(f_0) \\ \vdots \\ \mathrm{Re}\,\Delta h(f_{N_f-1})w(f_{N_f-1}) \\ \mathrm{Im}\,\Delta h(f_0)w(f_0) \\ \vdots \\ \mathrm{Im}\,\Delta h(f_{N_f-1})w(f_{N_f-1}) \end{bmatrix} \in \mathbb{R}^{2N_f} \tag{54}$$

which is just a vector of length $2N_f$ which consists of the real and imaginary parts of Δh_k times the weight. This works because if we choose $h_{1,2} = a_{1,2} + ib_{1,2}$, then

$$\begin{aligned}(h_1|h_2) &= \sum_{i=0}^{N_f-1} (a_1^i a_2^i + b_1^i b_2^i) w^2(f_i), \\ &\equiv \mathcal{P}[h_1]^T \mathcal{P}[h_2],\end{aligned} \tag{55}$$

where the first equality is via Eq. (52) and second is via the definition of the packing operator in Eq. (54). With these definitions, the packing operator now explicitly has the Euclidean inner product form. Thus, we define the data matrix $\boldsymbol{H}$ via

$$H_{kI} \equiv \mathcal{P}[\Delta h_k^\perp], \tag{56}$$

where $0 \le I < 2N_f$ indexes packed frequency and k indexes PN order. Note that since we are working in the perpendicular basis, the covariance matrix is decoupled from the GR parameters.

Since we use the "snapshots" of the data matrix (appropriately packing them together), the data matrix is

$$\boldsymbol{H} = [\ldots, \Delta h_k^\perp(f)w(f), \ldots] \in \mathbb{C}^{n_k \times 2N_f}, \tag{57}$$

where n_k is the number of ppE waveform deviations and N_f is the number of frequency bins. This snapshot matrix $\boldsymbol{H}$ can be best captured by projectors from the SVD formalism [100,106]

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Omega}\boldsymbol{V}^\dagger, \tag{58}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}^\dagger$ are left and right unitary square matrices of n_k and $2N_f$ dimensions, respectively. $\boldsymbol{\Omega}$ is a diagonal matrix consisting of the singular values of $\boldsymbol{H}$ in descending order. The SVD can be equivalently represented as

$$\boldsymbol{H} = \sum_{\alpha}^{n_k} \omega_\alpha \boldsymbol{u}_\alpha \boldsymbol{v}_\alpha^*, \tag{59}$$

since the rank of the matrix is the number of ppE waveform deviations. The $\boldsymbol{u}_\alpha$ are the left-singular vectors and the $\boldsymbol{v}_\alpha^*$ are the right-singular vectors. This is shown in Fig. 8, which has the $\boldsymbol{H}$ matrix in red, and the left-singular and right-singular vectors denoted.

Let us also give some comments about how to best use the SVD with modified gravity waveforms. In our work, we have been generically assuming that there are only phase deviations. If there are amplitude and phase deviations (e.g., in a true bGR waveform), the deviations are of the form

$$\Delta h^\perp(f_i) = (\Delta A^\perp + i\Delta\Psi^\perp(f_i)A(f_i))e^{i\Psi(f_i)}. \tag{60}$$

It is actually more useful to note that the inner product has invariance corresponding to $(h_1|h_2) = (h_1(f)e^{-i\Psi(f)}|h_2(f)e^{-i\Psi(f)})$. Therefore, one can define the packing operator as

[4]Note that we could have done this by including negative frequencies and doing the SVD on this case which does have a Hermitian inner product $(a|b) \sim a^\dagger b$, but this would need extra transformations to read out the real form of the singular vectors.

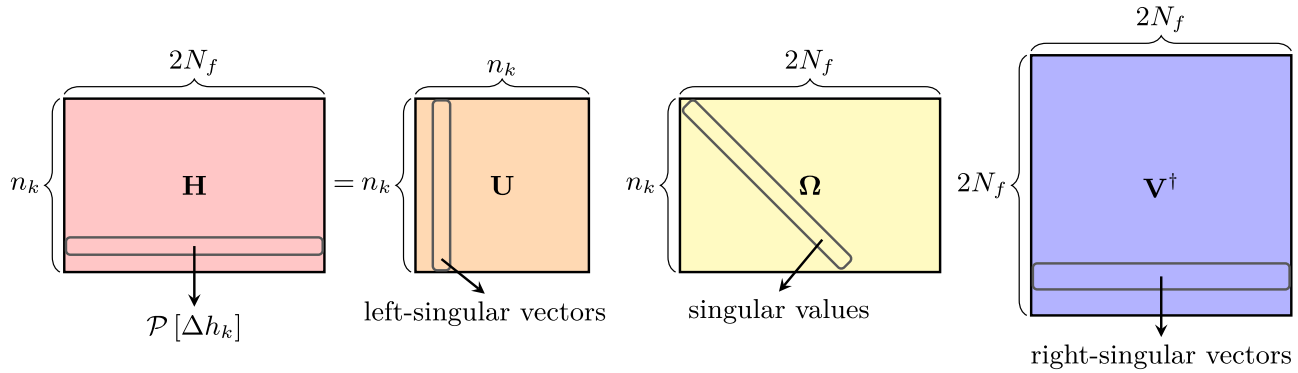

FIG. 8. We show the visualization of the SVD operation. For our choice of the SVD, we use $H_{ik} = \mathcal{P}[\Delta h_k^\perp]$ as given in Eqs. (56) and (61). The matrix $\boldsymbol{U}$ is a square matrix that represents linear redefinitions of $\delta\varphi_k$ while the matrix $\boldsymbol{V}$ is linear combinations of $\Delta h_k^\perp(f_i)w(f_i)$ which diagonalize $\boldsymbol{H}$. Finally, the matrix $\boldsymbol{\Sigma}$ is ordered by singular value that ranks the importance of each principal mode. This figure is written when we assume that we are using a single detector, but in the multiple detector case, the number of $\boldsymbol{H}$ columns would be increased as described in Eq. (63). We use the standard SVD algorithm on the matrix $\boldsymbol{H}$ to get the singular vectors $v_\alpha(f)$ from $\boldsymbol{V}^\dagger$ rows.

$$\mathcal{P}[\Delta h^\perp] = \begin{bmatrix} \vdots \\ \Delta A^\perp(f_i)w(f_i) \\ \vdots \\ \vdots \\ \Delta\Psi^\perp(f_i)A(f_i)w(f_i) \\ \vdots \end{bmatrix}, \tag{61}$$

where we explicitly removed the GR phase. Notice that our definition for Eq. (61) is a good representation of the GW inner product because it is real and well-behaved as stipulated in Eq. (55). Since we have been focusing on ppE phase deviations corrections to GR, the first column actually reduces to zero for this paper, but we give the generic formalism here. Additionally, the case of multiple detectors proceeds straightforwardly.

$$\sum_{\det A}(a|b)_A = \sum_{\det A}\sum_{i=0}^{N_f-1} a_A^*(f_i)b_A(f_i)w_A^2(f_i), \tag{62}$$

where the weights $w_A(f_i)$ depend on the individual detector sensitivity. Therefore, the packing operator for a collection of gravitational wave detectors is just

$$\mathcal{P}[\Delta h^\perp] \equiv \begin{bmatrix} \vdots \\ \mathcal{P}_{\det A}[\Delta h^\perp_{\det A}] \\ \vdots \end{bmatrix} \in \mathbb{R}^{2N_f N_{\det}}, \tag{63}$$

where this formalism now directly generalizes to this case. This also generalizes trivially to detectors of different bandwidths (e.g., LISA + LIGO) where the frequency bins are not the same in each detector.

We have introduced the SVD to identify the common features in the ppE templates. We are specializing this analysis to the case of using all perpendicular ppE waveform templates. $\boldsymbol{U}$ is a unitary matrix that converts between the ppE template parameters $\delta\varphi_k$ to the SVD basis $\delta\varphi_\alpha$. If one looks at the rows of the matrix $\boldsymbol{V}$ and adds the real part of the first half plus imaginary part of the second half, we would have

$$v_\alpha(f_i) = (\Delta A^\perp(f_i) + i\Delta\Psi_\alpha^\perp(f_i)A(f_i))e^{i\Psi(f_i)}w(f_i), \tag{64}$$

which allows unique identification of the dominant parts of the SVD as $\Delta\Psi_\alpha^\perp(f_i)$. Note that $\Delta A^\perp(f_i) = 0$ since we originally started with only phase deviations. Therefore we can easily read off the SVD phase deviation $\Delta\Psi_\alpha^\perp(f_i)$ by looking at the singular vector. Note that when amplitude deviations are set to zero, Eq. (64) is the same from of what Pai *et al.* [100] found. This proves that the SVD operation in [100] is equivalent to minimizing Eq. (48).

Let us now show what the SVD looks like for a particular signal. In Fig. 9, we took the example of GW150914 again and computed the SVD of $\Delta h_k^\perp$ between $k = -2, -1, \ldots, 4$, 6, 7. We normalized each of the perpendicular strains such that they are equally important ($\|\Delta h_k^\perp\| = 1$). On the left panel of Fig. 9, we plot the phase modulation $\Delta\Psi_k^\perp(f)$ in solid, semitransparent lines. We plot the three dominant SVD lines with dot dashed (black, red, blue). You can see that black SVD tracks each line very well while the blue and red ones are orthogonal to it. This demonstrates that there is a main principal feature of the ppE tests that exists in all of them. We plot the three dominant SVD modes because they account for 99.9% of the total variance; consequently, truncating the expansion at these three eigenvectors results in an information loss of only 0.1%.

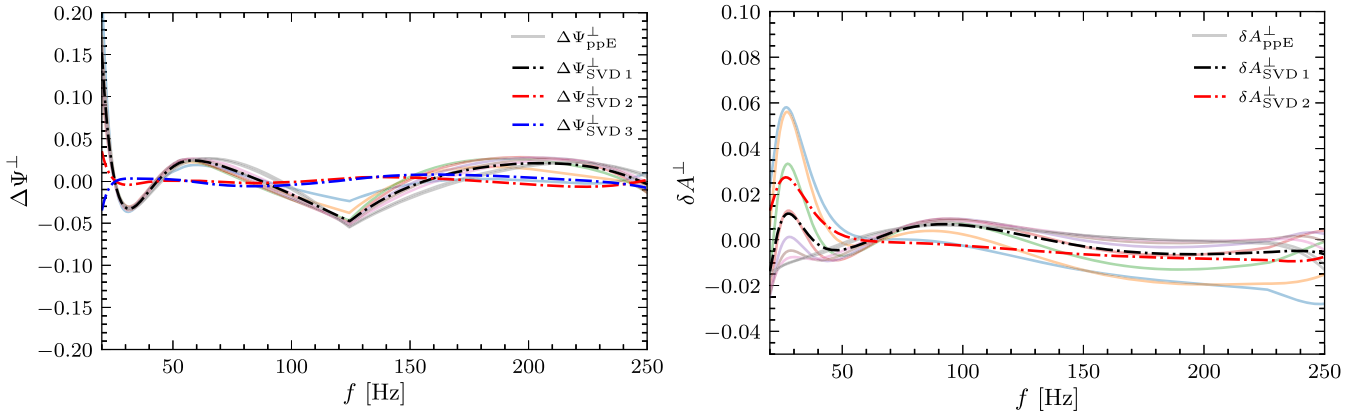

FIG. 9. Demonstration of the SVD on different ppE waveform deviations. In transparent solid, different $\Delta\Psi^{\perp}_{k}$ are shown and the dashed lines are the SVD modes for GW150914-like event. Note that we normalize each $\Delta\Psi^{\perp}_{k}$ so that $\|\Delta h^{\perp}_{k}\| = 1$ so each signal is equally weighted in the SVD. One can see that the dashed black line is the dominant SVD term that best fits each waveform. Additionally the red and blue curves represent subdominant SVD dephasing that are orthogonal to each other. One can see that each of these ppE waveform deviations is captured by only a few SVD modes, which explains how degeneracies would form when using the original parameter space [122].

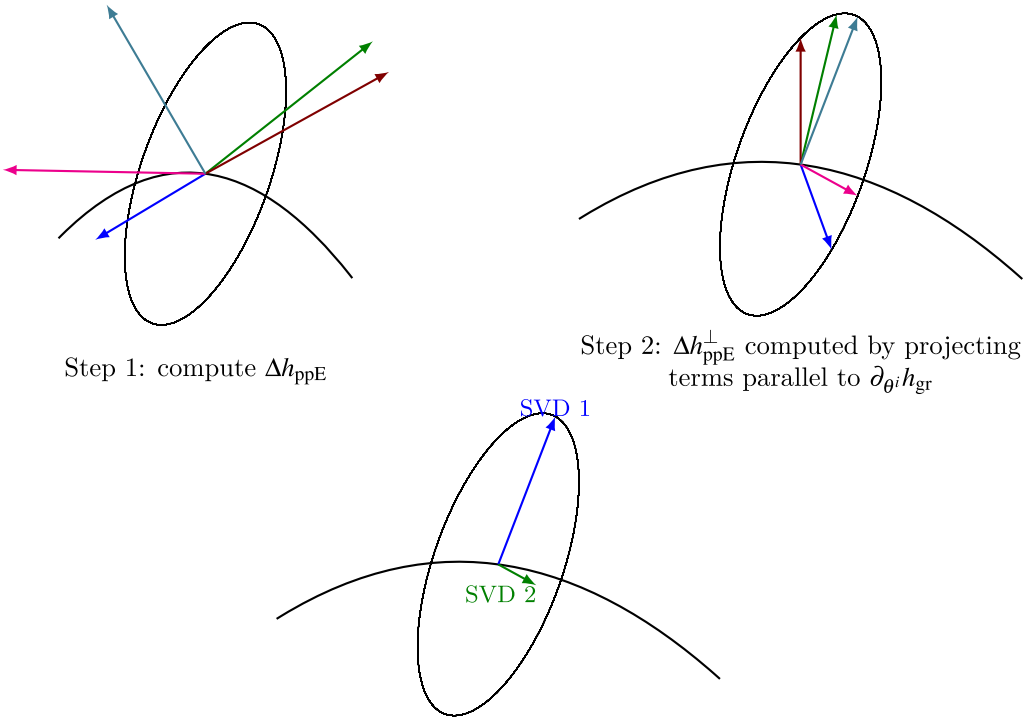

FIG. 10. Construction of SVD multiparameter tests. We start by orthogonalizing the ppE deviations from GR. Next, we construct the SVD basis by choosing the directions that are best fit by looking at the perpendicular ppE components. One can see that the SVD 1 in the third step is perpendicular to SVD 2, but the length is shorter because most of the SVD vectors point upward. This figure highlights our intuition that the natural basis is to do the SVD on $\Delta h^{\perp}_{k}$ and not on the original Δh_{k}. By construction, the SVD vectors lie in the subspace that is perpendicular to GR.

Additionally, on the right side of Fig. 9, the fractional amplitude modulations $\delta A_k^{\perp} \equiv \Delta A_k / A_{\rm GR}$ are shown in solid, semitransparent lines while the dominant SVD modes are dot dashed for the SVD vectors. Note that $\delta A_k = 0$ originally—since we only considered phase modulations to the parametrized tests. However $\delta A_k^{\perp} \neq 0$ because the GR parameter biases[5] shift some the phase modulation in Δh_k into a mixture of phase and amplitude modulation in $\Delta h_k^{\perp}$. This figure directly demonstrates why we saw such high overlaps between each of the ppE tests in Fig. 5.

In Fig. 10, we visually demonstrate how our proposal for using the SVD to identify nondegenerate directions for the parametrized tests of GR. One can see that originally, each Δh_k (colored arrow) lies off the GR manifold (black curved line). In step 2, the parametrized tests are projected to be perpendicular to GR manifold while being unit length. In step 3, the singular value decomposition examines these collection of parametrized test templates and identifies the orthogonal basis which is weighted by importance.

V. CONCLUSION

In this work, we have shown that there is a relationship between searches for beyond GR morphologies and the geometric description of noise in parameter estimation. We showed that the ppE parameters are able to capture generic beyond GR deviations due to the bias mechanism described above. We also showed that these parametrized tests should do a good job in real data analysis and there is not too big of a loss by having systematic issues in the beyond GR template. We demonstrated that the techniques for multiparameter tests of GR suffer from degeneracies, so this means measuring multiple ones simultaneously is difficult. Finally, we proposed a new template for searching for deviations from GR using the SVD. The SVD introduced in this paper generalizes previous approaches by allowing for amplitude + phase modulations and multiple detectors.

In this work, we have provided a geometric explanation for why parametrized tests are unexpectedly effective at capturing smooth, monotonic dephasing away from GR. It remains to be seen whether there are some dephasing terms that could not be captured by the parametrized tests. The approaches that are developed in this paper for geometrically understanding how similar small effects such as GR violation are to one another can be applied to other situations that GW astronomy is in. For example, this could help understand when it is possible to disentangle the observational signatures of spin precession, orbital eccentricity, waveform-model systematics, non-Gaussian instrumental glitches, and genuine deviations from GR in a more categorical manner. Many of the most informative measurements in GW astronomy hinge on resolving such subtle effects, and the geometric framework developed here provides a principled way to assess their detectability.

[5]Notably precession parameters such as χ_p are mixing phase and amplitude in the perpendicular waveform.

ACKNOWLEDGMENTS

We thank Michele Vallisneri, Ethan Payne, Taylor Knapp, Max Isi, Daiki Watarai, Bangalore Sathyaprakash, and Katerina Chatziioannou for their insightful discussions. B. S. acknowledges support by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1745301. Y. C. and B. S. acknowledge support from the Brinson Foundation, the Simons Foundation (Award No. 568762), and by NSF Grants No. PHY-2309211 and No. PHY-2309231. J. G. would like to gratefully acknowledge the support from the National Science Foundation through the Grant No. NSF PHY-2207758. We also thank the anonymous referee for their feedback on this work.

DATA AVAILABILITY

The data that support the findings of this article are openly available [116].

APPENDIX A: COMMENT ON NUMERICAL STABILITY

We stated in Eq. (5) that the best way to parametrize the bias equation is in terms of

$$\Delta h_{\rm raw} \equiv h_{\rm s}(\boldsymbol{\theta}_{\rm t}, \boldsymbol{\lambda}_{\rm t}) - h_{\rm m}(\boldsymbol{\theta}_{\rm t}), \tag{A1}$$

and then "raw" bias to the GR parameters is equal to

$$(\Delta\theta^i_{\rm bias})_{\rm raw} = \Sigma^{ij}(\partial_j h_{\rm m}|\Delta h_{\rm raw}), \tag{A2}$$

In fact, as first derived by Cutler-Vallisneri, the expression for Eq. (A2) has a limited range of validity. In fact, it is more numerically stable to use the following expression for Δh as

$$\Delta h_{\rm stab} \equiv [\Delta A + i\Delta\Psi A_{\rm s}]e^{i\Psi_{\rm s}} \tag{A3}$$

then the "stabilized" bias equation is

$$(\Delta\theta^i_{\rm bias})_{\rm stab} = \Sigma^{ij}(\partial_j h_{\rm m}|\Delta h_{\rm stab}), \tag{A4}$$

where $h_s(\boldsymbol{\theta}_{\rm t}, \boldsymbol{\lambda}_{\rm t}) \equiv A_{\rm s} e^{i\Psi_{\rm s}}$,

$$\Delta A \equiv A_{\rm s}(\boldsymbol{\theta}_{\rm t}, \boldsymbol{\lambda}_{\rm t}) - A_{\rm m}(\boldsymbol{\theta}_{\rm t}), \tag{A5}$$

$$\Delta\Psi \equiv \Psi_{\rm s}(\boldsymbol{\theta}_{\rm t}, \boldsymbol{\lambda}_{\rm t}) - \Psi_{\rm m}(\boldsymbol{\theta}_{\rm t}). \tag{A6}$$

This is derived in their paper, but one can show that the error for these expressions is

$$\text{error of}(\Delta\theta^{\text{raw}}_{\text{bias}})_{\text{error}} \sim \Sigma^{ij}(\partial_j h_{\text{m}}|\Delta\Psi^2 h_{\text{s}}) \tag{A7}$$

Let us define alternative expressions for amplitude and phase deviation which are the difference between true and ML point

$$\delta A \equiv A_{\text{s}}(\boldsymbol{\theta}_{\text{t}}, \boldsymbol{\lambda}_{\text{t}}) - A_{\text{m}}(\boldsymbol{\theta}_{\text{ML}}), \tag{A8}$$

$$\delta\Psi \equiv \Psi_{\text{s}}(\boldsymbol{\theta}_{\text{t}}, \boldsymbol{\lambda}_{\text{t}}) - \Psi_{\text{m}}(\boldsymbol{\theta}_{\text{ML}}). \tag{A9}$$

In fact, note that $\delta\Psi$ is small compared to $\Delta\Psi$ so expanding in these has a much greater regime of validity. One can show that

$$\text{error of}(\Delta\theta^{i}_{\text{bias}})_{\text{stab}} \sim \Sigma^{ij}(\partial_j h_{\text{m}}|\delta\Psi^2 h_{\text{s}}), \tag{A10}$$

which is valid over a much larger part of the parameter space.

Note that the discussion above admits a simple geometric interpretation. Using the "raw" expression for the bias equation in Eq. (A2) effectively assumes that the model waveform never departs from the true waveform by more than ~ 1 radian in phase, $\Delta\Psi \lesssim 1$, which corresponds schematically to requiring $\|\Delta h\| \lesssim 1$. However, what actually controls the validity of the expansion is not the total model error, but the *residual* error after refitting the GR parameters. In other words, it is sufficient that the ML phase only deviates from the true phase by less than a radian, $\delta\Psi \lesssim 1$, which corresponds to $\|\Delta h_{\perp}\| \lesssim 1$, where $\Delta h_{\perp}$ is the component of the model error orthogonal to the GR tangent space $\{\partial_i h_{\text{m}}\}$.

Therefore, the regime of validity of the expressions in this paper is set by the smallness of the residual mismatch, $\|\Delta h_{\perp}\|$, rather than the total mismatch, $\|\Delta h\|$. By working with the "stabilized" expression in Eq. (A4), the bias formula remains accurate even in situations with $\|\Delta h\| \gtrsim 1$, i.e., in the presence of substantial stealth bias [73], as long as the residual component $\Delta h_{\perp}$ remains small.

APPENDIX B: BIAS FORMULA FOR MULTIPLE μ^a PARAMETERS

This can generically be proven in the following way. Assume $\boldsymbol{\mu} = \mu^a$ and $\boldsymbol{\lambda} = \lambda^A$. Then the estimate for μ^a is

$$\Gamma_{IJ}\mu^{J}_{\text{bias}} = (\partial_I h_{\text{m}}|\Delta h). \tag{B1}$$

Let us now write this equation as a matrix equation in block form

$$\begin{pmatrix} \Gamma_{ij} & \Gamma_{ib} \\ \Gamma_{ai} & \Gamma_{ab} \end{pmatrix} \begin{pmatrix} \Delta\theta^{i}_{\text{bias}} \\ \Delta\mu^{b}_{\text{bias}} \end{pmatrix} = \begin{pmatrix} (\partial_i h_{\text{m}}|\Delta h) \\ (\partial_a h_{\text{m}}|\Delta h) \end{pmatrix}. \tag{B2}$$

Using properties of the Schur complement, we can solve for $\Delta\mu^{a}_{\text{bias}}$ where the degeneracy with $\boldsymbol{\theta}$ is removed

$$[\Gamma_{ab} - \Gamma_{ai}(\Gamma_{ij})^{-1}\Gamma_{jb}]\Delta\mu^{a}_{\text{bias}} = (\partial_a h_{\text{m}}|\Delta h) - \Gamma_{ai}(\Gamma_{ij})^{-1}(\partial_j h_{\text{m}}|\Delta h). \tag{B3}$$

The part which is perpendicular to $\boldsymbol{\theta}$ is $(\partial_a h_{\text{m}})^{\perp\boldsymbol{\theta}} = \partial_a h_{\text{m}} - \Gamma_{aj}(\Gamma_{ij})^{-1}\partial_i h_{\text{m}}$ (this can be proven by a simple calculation). We also note that the reduced Fisher matrix is defined to be

$$\Gamma^{\text{red}}_{ab} = [\Gamma_{ab} - \Gamma_{ai}(\Gamma_{ij})^{-1}\Gamma_{jb}] = ((\partial_a h_{\text{m}})^{\perp\boldsymbol{\theta}}|(\partial_b h_{\text{m}})^{\perp\boldsymbol{\theta}}). \tag{B4}$$

Thus the reduced Fisher matrix is the inner product of the derivatives of μ^a which have been made orthogonal to $\boldsymbol{\theta}$. Therefore, the geometric value for the bias to $\boldsymbol{\mu}$ is

$$\Delta\mu^{a}_{\text{bias}} = \Sigma^{ab}_{\text{red}}((\partial_b h_{\text{m}})^{\perp\boldsymbol{\theta}}|(\Delta h)^{\perp\boldsymbol{\theta}}). \tag{B5}$$

This equation describes how biases can affect our measurement of small parameters. For example, a true deviation from GR Δh at a particular PN order can find evidence of deviation at a different PN order. Additionally, it explains how eccentricity and precession can be confused with one another.

APPENDIX C: DERIVATION OF BAYES FACTORS

In this appendix, we will compute the Bayes factor in two different cases. The first is if you have a bGR theory, and you know it *a priori* and search with the correct waveform which was originally done in [72]. The second case, we will generalize this equation to the case where we are searching for a violation with an incorrect model (for example a parametrized test of GR). Generally, the Bayes factor is a ratio of the evidences

$$\mathcal{B}^{\mathcal{M}_2}_{\mathcal{M}_1} \equiv \frac{p(d|\mathcal{M}_2)}{p(d|\mathcal{M}_1)}, \tag{C1}$$

where d is the data and $\mathcal{M}_i$ are the models. The evidence is equal to the integral over all the parameters

$$p(d|\mathcal{M}) = \int d^n\theta\, p(\theta|\mathcal{M})\, p(d|\theta, \mathcal{M}), \tag{C2}$$

and n is the number of parameters in the model.

1. Bayes factor for a correctly modeled theory

We will first derive the Bayes factor for the case of a beyond-GR (bGR) injection where we have perfectly modeled the bGR morphology. If we inject $s_{\text{bGR}} = h_{\text{bGR}}(\boldsymbol{\theta}_{\text{t}}, \lambda_{\text{t}})$, in the Fisher matrix limit the likelihood is

$$p(s_{\text{bGR}}|\delta\theta^{\mu}) = \mathcal{N} e^{-|n|^2/2 + (G^{-1})^{\mu\nu}(n|h_{\mu})(n|h_{\nu})/2 - G_{\mu\nu}\delta\theta^{\mu}\delta\theta^{\nu}/2}, \tag{C3}$$

where $G_{\mu\nu} = (h_\mu|h_\nu)$ is the $(m+1)$ dimensional bGR Fisher matrix and Greek letters (μ,ν) range over (θ^i,λ) parameters. The value of $\delta\theta^\mu$ is defined as

$$\delta\theta^\mu = (G^{-1})^{\mu\nu}(n|h_\nu). \tag{C4}$$

If we assume flat priors, the evidence is

$$\begin{aligned} p(s_{\rm bGR}|{\rm bGR}) &= \int p(\theta^\mu|{\rm bGR})p(\delta\theta^\mu|s_{\rm bGR}), \\ &= \frac{(2\pi)^{(m+1)/2}\sqrt{|G^{-1}|}}{\prod_\mu \Delta\theta^\mu_{\rm prior}} \\ &\quad \times \mathcal{N} e^{-|n|^2/2+(G^{-1})^{\mu\nu}(n|h_\mu)(n|h_\nu)/2}. \end{aligned} \tag{C5}$$

Next, we will compute the evidence for GR when we have beyond GR morphology

$$\begin{aligned} p(s_{\rm bGR}|{\rm GR}) &= \frac{(2\pi)^{m/2}\sqrt{|F^{-1}|}}{\prod_i \Delta\theta^i_{\rm prior}} \\ &\quad \times \mathcal{N} e^{-|n+\Delta h|^2/2+(F^{-1})^{ij}(n+\Delta h|h_i)(n+\Delta h|h_j)/2}, \end{aligned} \tag{C6}$$

where F_{ij} is the (m) dimensional GR Fisher matrix and (i,j) range of θ^i. After further simplification, the GR evidence can be compactly written as

$$\begin{aligned} p(s_{\rm bGR}|{\rm GR}) &= \frac{(2\pi)^{m/2}\sqrt{|F^{-1}|}}{\prod_i \Delta\theta^i_{\rm prior}} \\ &\quad \times \mathcal{N} e^{-|n|^2/2-|\Delta h_{\perp{\rm GR}}|^2/2-x|\Delta h_{\perp{\rm GR}}|+(F^{-1})^{ij}n_i n_j}, \end{aligned} \tag{C7}$$

where $\Delta h = h_{\rm bGR}(\boldsymbol{\theta}_{\rm t},\lambda_{\rm t}) - h_{\rm GR}(\boldsymbol{\theta}_{\rm t})$. Therefore the Bayes factor is equal to

$$\mathcal{B}^{\rm bGR}_{\rm GR} = \frac{(2\pi)^{1/2}\Delta\lambda_{\rm est}}{\Delta\lambda^{\rm bGR}_{\rm prior}} e^{|\Delta h_{\perp{\rm GR}}|^2/2+x|\Delta h_{\perp{\rm GR}}|+x^2/2}, \tag{C8}$$

where $x = (\Delta h_\perp, n)/|\Delta h_\perp|$ is a normal random variable and the standard deviation on λ from the observation is equal to

$$\begin{aligned} \Delta\lambda_{\rm est} &\equiv \sqrt{|F|/|G|}, \\ &= \frac{1}{\|(\partial_\lambda h_{\rm bGR})^{\perp{\rm GR}}\|}, \end{aligned} \tag{C9}$$

where $(\partial_\lambda h_{\rm bGR})^{\perp{\rm GR}}$ is the component of the derivative that is perpendicular to GR. This equation matches the results of [72] in Eq. (11). One can see that the prefactor $\Delta\lambda_{\rm est}/\Delta\lambda^{\rm bGR}_{\rm prior}$ outside Eq. (C8) is an Occam factor that disfavors more complicated models while the evidence for beyond-GR morphology comes from the term $\Delta h_{\perp{\rm GR}}$ in the exponential.

2. Bayes factor for a mismodeled theory

Let us now turn to the case that we are using a ppE model to search for beyond GR morphology which has mismodeled it. The GR evidence is the same as in Eq. (C6) while the evidence for ppE is equal to

$$\begin{aligned} p(s_{\rm GR}|{\rm ppE}) &= \frac{(2\pi)^{(m+1)/2}\sqrt{|H^{-1}|}}{\prod_\alpha \Delta\theta^\alpha_{\rm prior}} \\ &\quad \times \mathcal{N} e^{-|n|^2/2-|\Delta h_{\perp{\rm ppE}}|^2/2-x|\Delta h_{\perp{\rm ppE}}|+(H^{-1})^{\alpha\beta}n_\alpha n_\beta}, \end{aligned} \tag{C10}$$

where $H^{\alpha\beta}$ is the $(m+1)$ dimensional ppE Fisher matrix and (α,β) range over (θ^i,μ). We are using notation $\Delta h_{\perp{\rm ppE}}$ to mean perpendicular to both GR and $\partial_\mu h$. The odds ratio between ppE and GR for a bGR injection is thus

$$\begin{aligned} \mathcal{B}^{\rm ppE}_{\rm GR}\Big|_{s_{\rm bGR}} &= \frac{(2\pi)^{1/2}\Delta\mu_{\rm est}}{\Delta\mu^{\rm ppE}_{\rm prior}} \exp\Big[-|\Delta h_{\perp{\rm ppE}}|^2/2 - y|\Delta h_{\perp{\rm ppE}}| \\ &\quad + y^2/2 + |\Delta h_{\perp{\rm GR}}|^2/2 + x|\Delta h_{\perp{\rm GR}}|\Big], \\ &= \frac{(2\pi)^{1/2}\Delta\mu_{\rm est}}{\Delta\mu^{\rm ppE}_{\rm prior}} \exp\left[\frac{1}{2}(\rho^\perp_{\rm ppE})^2 - \rho^\perp_{\rm ppE}y + \frac{1}{2}y^2\right], \end{aligned} \tag{C11}$$

where covariance on the ppE parameter is equal to

$$\begin{aligned} \Delta\mu_{\rm est} &= \sqrt{\frac{|F|}{|H|}}, \\ &= \frac{1}{\|(\partial_\mu h_{\rm ppE})^{\perp{\rm GR}}\|}. \end{aligned} \tag{C12}$$

The quantities (x,y) are random unit variables

$$\begin{aligned} x &\equiv (\Delta h_{\perp{\rm GR}}|n)/|\Delta h_{\perp{\rm GR}}|, \\ y &\equiv (\Delta h_{\perp{\rm ppE}}|n)/|\Delta h_{\perp{\rm ppE}}|, \end{aligned}$$

note that (x,y) have some nonvanishing correlation. The *captured residual SNR* is

$$\rho^{\rm ppE}_\perp = \mathcal{O}(\Delta h^{\perp{\rm GR}}_{\rm ppE}, \Delta h^{\perp{\rm GR}}_{\rm bGR})\rho_\perp, \tag{C13}$$

and the *overlap* between the bGR and ppE waveforms is defined via

$$\mathcal{O} \equiv \frac{(\Delta h^{\perp{\rm GR}}|\partial_\mu h^{\perp{\rm GR}})}{\|\Delta h^{\perp{\rm GR}}\|\|\partial_\mu h^{\perp{\rm GR}}\|}. \tag{C14}$$

In this derivation we used the block matrices to show that the random unit variables are related via

$$x\|\Delta h_{\perp \mathrm{GR}}\| = y\|\Delta h_{\perp \mathrm{ppE}}\| + \mathcal{O}(\Delta h^{\perp \mathrm{GR}}_{\mathrm{ppE}}, \Delta h^{\perp \mathrm{GR}}_{\mathrm{bGR}})(\Delta h^{\perp \mathrm{GR}}_{\mathrm{ppE}}|n), \quad \text{(C15)}$$

which allowed simplification of the equation above. This can be written in a simpler form in the following way

$$x = \sqrt{1-\mathcal{O}^2}y + \mathcal{O}z \quad \text{(C16)}$$

this follows from $\|\Delta h_{\perp \mathrm{ppE}}\| = \sqrt{1-\mathcal{O}^2}\rho_\perp$ and defining $z \equiv (\Delta h^{\perp \mathrm{GR}}_{\mathrm{ppE}}|n)/\|\Delta h^{\perp \mathrm{GR}}_{\mathrm{ppE}}\|$. Looking at Fig. 2, we see that $\Delta h^{\perp \mathrm{GR}}_{\mathrm{ppE}}$ is the blue line, $\Delta h^{\perp \mathrm{GR}}_{\mathrm{bGR}}$ is the red line, and $\Delta h_{\perp \mathrm{ppE}}$ is the brown line. Note that x, y, z are correlated with one another, one can show that $E[xz] = \mathcal{O}$ directly by using $E[(a|n)(b|n)] = (a|b)$ identify for Gaussian noise. Using this property, $E[yz] = 0$ and $E[xy] = \sqrt{1-\mathcal{O}^2}$.

3. Bayes factor between two beyond GR theories

Finally, let us compute the Bayes factor that compares a beyond GR signal to that of a parametrized model as we did in Eq. (31). This can be done using

$$\mathcal{B}^{\mathrm{bGR}}_{\mathrm{ppE}}\Big|_{s_{\mathrm{bGR}}} = \frac{\mathcal{B}^{\mathrm{GR}}_{\mathrm{bGR}}}{\mathcal{B}^{\mathrm{GR}}_{\mathrm{ppE}}}\Big|_{s_{\mathrm{bGR}}}, \quad \text{(C17)}$$

since the Bayes factors are the ratio of the evidences. This results in the expression

$$\mathcal{B}^{\mathrm{bGR}}_{\mathrm{ppE}}\Big|_{s_{\mathrm{bGR}}} = \frac{\Delta\lambda_{\mathrm{stat}}\Delta\mu_{\mathrm{prior}}}{\Delta\lambda_{\mathrm{prior}}\Delta\mu_{\mathrm{stat}}} e^{(x+\rho_\perp)^2/2-(y+\mathcal{O}\rho_\perp)^2/2}, \quad \text{(C18)}$$

where x, y are correlated as discussed in Eq. (C16) and the Bayes factors on the right side were derived in previous Eqs. (C8) and (C2). While we have written down an expression for the Bayes factor, in this section we will (a) perform a rotation so it can be written as sum of decorrelated unit normals and (b) perform a high SNR expansion which approximates the z-score for how big the bias is.

Let us start with the log Bayes factor and decorrelate it [excluding constant Occam factor terms of Eq. (C18)], which we write as

$$\ell = \frac{1}{2}\rho_\perp^2 + \rho_\perp x + \frac{1}{2}x^2 - \left(\frac{1}{2}\mathcal{O}^2\rho_\perp^2 + \rho_\perp \mathcal{O} y + \frac{1}{2}y^2\right), \quad \text{(C19)}$$

where $\ell \equiv \mathcal{B}^{\mathrm{bGR}}_{\mathrm{ppE}}|_{s_{\mathrm{bGR}}}$ + priors. Since x, y are correlated it is convenient to convert to y, z as given in Eq. (C16). The quadratic form of the log Bayes factor can be thus written as

$$\ell = \frac{1}{2}(1-\mathcal{O}^2)\rho_\perp^2 + b^T v + \frac{1}{2}v^T H v \quad \text{(C20)}$$

where $v = (y, z)$ and

$$b = \rho_\perp \begin{bmatrix} \sqrt{1-\mathcal{O}^2} - \mathcal{O} \\ \mathcal{O} \end{bmatrix}, \quad \text{(C21)}$$

$$H = \begin{bmatrix} -\mathcal{O}^2 & \mathcal{O}\sqrt{1-\mathcal{O}^2} \\ \mathcal{O}\sqrt{1-\mathcal{O}^2} & \mathcal{O}^2 \end{bmatrix}. \quad \text{(C22)}$$

Since H is a symmetric matrix, there exists a rotation matrix which diagonalizes H. One can show that the likelihood equals

$$\ell = \frac{1}{2}\mathcal{O}\left[(w_1+\mu_1)^2 - (w_2+\mu_2)^2\right], \quad \text{(C23)}$$

where w_1, w_2 are unit Gaussian random variables that are uncorrelated which are a rotated basis from (y, z).[6] The rotation matrix which accomplishes this decorrelation and whitening operation is

$$R = \begin{bmatrix} \sqrt{\frac{1-\mathcal{O}}{2}} & \sqrt{\frac{1+\mathcal{O}}{2}} \\ -\sqrt{\frac{1+\mathcal{O}}{2}} & \sqrt{\frac{1-\mathcal{O}}{2}} \end{bmatrix}, \quad \text{(C24)}$$

and the parameters μ_1, μ_2 are defined via $(\mu_1, -\mu_2) \equiv \frac{1}{\mathcal{O}} R \cdot b$ and are equal to

$$\mu_1 \equiv \frac{\sqrt{1+\mathcal{O}} - \mathcal{O}\sqrt{1-\mathcal{O}}}{\sqrt{2}\mathcal{O}}, \quad \text{(C25)}$$

$$\mu_2 \equiv \frac{\sqrt{1-\mathcal{O}} - \mathcal{O}\sqrt{1+\mathcal{O}}}{\sqrt{2}\mathcal{O}}. \quad \text{(C26)}$$

By doing this rotation and whitening operation, we have written this in a simple form. The results of Eq. (C23) demonstrate that the log Bayes factor can be written as a difference between two noncentral chi-square distributions where the noncentrality parameters are $\lambda_1 = \mu_1^2$ and $\lambda_2 = \mu_2^2$. Therefore, we have derived the expression for the Bayes factor between two theories

$$\mathcal{B}^{\mathrm{bGR}}_{\mathrm{ppE}}\Big|_{s_{\mathrm{bGR}}} = \frac{\Delta\lambda_{\mathrm{stat}}\Delta\mu_{\mathrm{prior}}}{\Delta\lambda_{\mathrm{prior}}\Delta\mu_{\mathrm{stat}}} e^{(w_1+\mu_1)^2/2-(w_2+\mu_2)^2/2}. \quad \text{(C27)}$$

[6] Explicitly $(w_1, w_2) \equiv R \cdot (y, z)$.

Since the difference between two noncentral chi-square random variables is not a well-known distribution, this is the most simplified form for the Bayes factor here.

While Eq. (C27) depends on two random variables, we can write this as a simplified distribution if we assume that the residual SNR is large ($\rho_\perp \gg 1$). If the residual SNR is large (and correspondingly $|\mu_1|, |\mu_2| \gg 1$[7]) then the log Bayes factor can be written as

$$\ell \approx \frac{1}{2}(1-\mathcal{O}^2)\rho_\perp^2 + \mathcal{O}(\mu_1 w_1 - \mu_2 w_2). \tag{C28}$$

Note that we can define a new unit normal random variable

$$W \equiv \frac{\mu_1 w_1 - \mu_2 w_2}{\sqrt{\mu_1^2+\mu_2^2}} \sim \mathcal{N}(0,1), \tag{C29}$$

thus the log Bayes factor is distributed as

$$\ell \approx \frac{1}{2}(1-\mathcal{O}^2)\rho_\perp^2 + \mathcal{O}\sqrt{\mu_1^2+\mu_2^2}W \tag{C30}$$

This means the z-score is equal to

$$\begin{aligned} Z &= \frac{(\rho_\perp^{\rm diff})^2/2}{|\mathcal{O}|\sqrt{\mu_1^2+\mu_2^2}}, \\ &= \frac{1-\mathcal{O}^2}{2\sqrt{1+\mathcal{O}^2-2\mathcal{O}\sqrt{1-\mathcal{O}^2}}}\rho_\perp. \end{aligned} \tag{C31}$$

One can see that the z-score is linearly proportional to the residual SNR, but also depends on the overlap between the ppE and beyond GR waveforms. Explicit evaluation of the prefactor yields an answer about $Z/\rho_\perp \sim 0.3$–0.6 when $\mathcal{O} \lesssim 0.7$ while for $\mathcal{O} \sim 1$ then $Z/\rho_\perp = 0$. This makes sense that the z-score is zero and there is no way to discern between these two theories for perfectly aligned theories (i.e., the angle between the manifolds is zero in Fig. 2). Note that the expression in Eq. (C31) is probably only exactly valid in the case of an exceedingly loud beyond GR signal ($\rho_\perp \gg 1$), so in realistic scenarios one must use Eq. (C27) which contains two random variables.

APPENDIX D: MULTIDETECTOR GEOMETRY

Let us use terminology that is from the polarization part of the LVK GWTC-3 paper [10]. A multidetector signal looks like

$$d^A(f) = F^{A\alpha} h_\alpha(f, \boldsymbol{\theta}_{\rm t}) + n^A(f), \tag{D1}$$

where A is the detector index and $\alpha = (+, \times)$ is the polarization index. The antenna responses are $F^{A\alpha} = D_A^{ij}\epsilon_{ij}^\alpha$ where D_A^{ij} is the Ath detector tensor and ϵ_{ij}^α is the GW polarization tensor of the α polarization. If we introduce the square brackets to represent the sum over detectors

$$[a|b] = \sum_A (a_A|b_A)_A, \tag{D2}$$

then the maximum likelihood point is the solution to

$$\Gamma_{IJ}\Delta\Theta^J = [\partial_I h_{\rm m}|\Delta h] + [\partial_I h_{\rm m}|n], \tag{D3}$$

where the network Fisher information matrix is $\Gamma_{IJ} \equiv [\partial_I h_{\rm m}|\partial_J h_{\rm m}]$. The equations for the statistical error and bias straightforwardly generalize from Eq. (7) and are equal to

$$\Delta\Theta^I_{\rm stat} = \Sigma^{IJ}[\partial_I h_{\rm m}|n], \tag{D4}$$

$$\Delta\Theta^I_{\rm bias} = \Sigma^{IJ}[\partial_I h_{\rm m}|\Delta h]. \tag{D5}$$

Thus, the effect of searching for deviations $\Delta\mu^a$ in a waveform described by true parameters $(\theta^i_{\rm t}, \lambda^A_{\rm t})$ is

$$\Delta\mu^a_{\rm bias} = \Sigma^{ab}_{\rm red}[(\partial_b h_{\rm m})^{\perp\boldsymbol{\theta}}|(\Delta h)^{\perp\boldsymbol{\theta}}], \tag{D6}$$

where $\Sigma^{ab}_{\rm red}$ is the reduced network covariance matrix and the sum is understood.

APPENDIX E: OVERLAPS FOR GW170817

As we discussed in the main part of the paper in Sec. III A, the results of how strong the overlaps are depends on the values of the intrinsic GR parameters for the event. In this appendix, we will provide context for this by giving an example for a GW170817-like event.

The left panel of Fig. 11 shows the overlap between the ppE tests computed using only the masses as measured GR parameters, analogous to Fig. 6. In contrast, the right panel allows $\chi_{\rm eff}$ and $\chi_{\rm p}$ to vary, making it analogous to Fig. 5 in the main text. The long inspiral time of a BNS event changes how well the ppE tests can compensate for one another. One can see that generally, the overlaps between the theories are smaller than what we saw in Sec. III A. For both of these plots, we have not included the $k = 0$ case for the overlap computation, which is because this 0PN test is nearly perfectly correlated with the chirp mass in the Fisher matrix. This degeneracy arises because current ground-based GW detectors do not measure much of the merger phase for BNS systems, and has been highlighted in the literature before in Appendix B of [129]. One can also see that in the right panel of Fig. 11, the overlaps for the $k = 3, 4$—corresponding to 1.5PN and 2PN terms—are significantly reduced compared to the nonspinning case. This reduction occurs because $\chi_{\rm eff}$ and $\chi_{\rm p}$ first enter the waveform at these PN orders, respectively.

[7]Note that for $\mathcal{O} = 1$, $\mu_1 = -\mu_2 = \rho_\perp$ while for smaller $\mathcal{O}$ this can be less valid as μ_2 can get small and cross zero. Thus the validity of $\rho_\perp \gg 1 \not\Rightarrow |\mu_2| \gg 1$, but for most signals the overlap is large enough that this is a good heuristic.

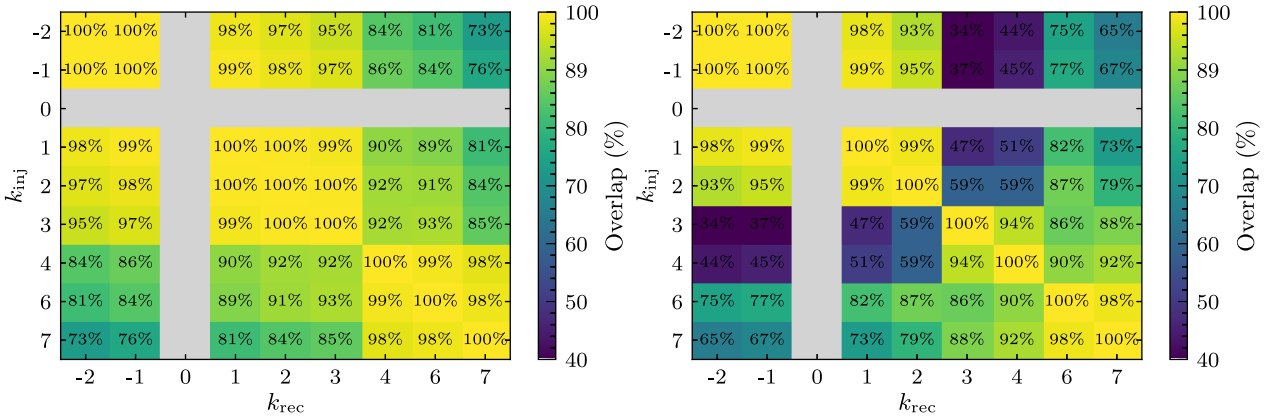

FIG. 11. Residual amplitude for ppE injected deviations from GR for a GW170817-like detection. On the left is case of nonspinning inference while on the right is the spinning case. Compared to the BBH case given in Sec. III A, this figure highlights how the overlap depends on the choice of the GR parameters. One can see that the fact that $\chi_{\rm eff}$ enters the BNS waveform primarily at 1.5PN while $\chi_{\rm p}$ enters at 2PN causes the overlaps to reduce on the right side. Due to the fact that BNS systems do not have much merger phenomenology in current ground-based detectors, the $k = 0$ term is highly degenerate with the chirp mass which is thus excluded from this plot [125].

[1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Observation of gravitational waves from a binary black hole merger, Phys. Rev. Lett. **116**, 061102 (2016).

[2] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW170817: Observation of gravitational waves from a binary neutron star inspiral, Phys. Rev. Lett. **119**, 161101 (2017).

[3] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), 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**, 031040 (2019).

[4] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, Phys. Rev. X **11**, 021053 (2021).

[5] R. Abbott *et al.* (KAGRA, Virgo, and LIGO Scientific Collaborations), GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, Phys. Rev. X **13**, 041039 (2023).

[6] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Tests of general relativity with GW150914, Phys. Rev. Lett. **116**, 221101 (2016); **121**, 129902(E) (2018).

[7] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Tests of general relativity with GW170817, Phys. Rev. Lett. **123**, 011102 (2019).

[8] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Tests of general relativity with the binary black hole signals from the LIGO-Virgo catalog GWTC-1, Phys. Rev. D **100**, 104036 (2019).

[9] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog, Phys. Rev. D **103**, 122002 (2021).

[10] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Tests of general relativity with GWTC-3, Phys. Rev. D **112**, 084080 (2025).

[11] N. Yunes, K. Yagi, and F. Pretorius, Theoretical physics implications of the binary black-hole mergers GW150914 and GW151226, Phys. Rev. D **94**, 084002 (2016).

[12] R. Nair, S. Perkins, H. O. Silva, and N. Yunes, Fundamental physics implications for higher-curvature theories from binary black hole signals in the LIGO-Virgo catalog GWTC-1, Phys. Rev. Lett. **123**, 191101 (2019).

[13] H. O. Silva, A. M. Holgado, A. Cárdenas-Avendaño, and N. Yunes, Astrophysical and theoretical physics implications from multimessenger neutron star observations, Phys. Rev. Lett. **126**, 181101 (2021).

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

[15] F. Acernese *et al.* (Virgo Collaboration), Advanced Virgo: A second-generation interferometric gravitational wave detector, Classical Quantum Gravity **32**, 024001 (2015).

[16] T. Akutsu *et al.* (KAGRA Collaboration), Overview of KAGRA: Detector design and construction history, Prog. Theor. Exp. Phys. **2021**, 05A101 (2021).

[17] B. Iyer, T. Souradeep, C. Unnikrishnan, S. Dhurandhar, S. Raja, and A. Sengupta, LIGO-India: Proposal of the consortium for Indian initiative in gravitational-wave observations (IndIGO), Technical Report No. LIGO-M1100296-v2, LIGO, 2013, https://dcc.ligo.org/LIGO-M1100296-v2/public.

[18] M. Saleem *et al.*, The science case for LIGO-India, Classical Quantum Gravity **39**, 025004 (2022).

[19] P. Fritschel *et al.*, Report of the LSC post-O5 study group, LIGO Document No. LIGO-T2200287, LVK DCC, 2022.

[20] R. X. Adhikari *et al.*, LIGO Voyager upgrade concept, LIGO Document No. LIGO-T1400226, 2024.

[21] R. X. Adhikari *et al.* (LIGO Collaboration), A cryogenic silicon interferometer for gravitational-wave detection, Classical Quantum Gravity **37**, 165003 (2020).
[22] D. Reitze *et al.*, Cosmic Explorer: The U.S. contribution to gravitational-wave astronomy beyond LIGO, Bull. Am. Astron. Soc. **51**, 035 (2019).
[23] M. Evans *et al.*, A horizon study for Cosmic Explorer: Science, observatories, and community, arXiv:2109.09882.
[24] M. Punturo *et al.*, The Einstein Telescope: A third-generation gravitational wave observatory, Classical Quantum Gravity **27**, 194002 (2010).
[25] M. Maggiore *et al.* (ET Collaboration), Science case for the Einstein Telescope, J. Cosmol. Astropart. Phys. 03 (2020) 050.
[26] M. Branchesi *et al.*, Science with the Einstein Telescope: A comparison of different designs, J. Cosmol. Astropart. Phys. 07 (2023) 068.
[27] P. Amaro-Seoane *et al.* (LISA Collaboration), Laser Interferometer Space Antenna (2017).
[28] P. Auclair *et al.* (LISA Cosmology Working Group), Cosmology with the laser interferometer space antenna, Living Rev. Relativity **26**, 5 (2023).
[29] J. Luo *et al.* (TianQin Collaboration), TianQin: A space-borne gravitational wave detector, Classical Quantum Gravity **33**, 035010 (2016).
[30] W.-R. Hu and Y.-L. Wu, The Taiji program in space for gravitational wave physics and the nature of gravity, Natl. Sci. Rev. **4**, 685 (2017).
[31] W.-H. Ruan, Z.-K. Guo, R.-G. Cai, and Y.-Z. Zhang, Taiji program: Gravitational-wave sources, Int. J. Mod. Phys. A **35**, 2050075 (2020).
[32] N. Seto, S. Kawamura, and T. Nakamura, Possibility of direct measurement of the acceleration of the universe using 0.1-Hz band laser interferometer gravitational wave antenna in space, Phys. Rev. Lett. **87**, 221103 (2001).
[33] S. Sato *et al.*, The status of DECIGO, J. Phys. Conf. Ser. **840**, 012010 (2017).
[34] S. Kawamura *et al.*, Current status of space gravitational wave antenna DECIGO and B-DECIGO, Prog. Theor. Exp. Phys. **2021**, 05A105 (2021).
[35] K. A. Kuns, H. Yu, Y. Chen, and R. X. Adhikari, Astrophysics and cosmology with a decihertz gravitational-wave detector: TianGO, Phys. Rev. D **102**, 043001 (2020).
[36] K. A. Kuns, Future networks of gravitational wave detectors: Quantum noise and space detectors, Ph.D. thesis, University of California, Santa Barbara, 2019.
[37] N. Yunes and F. Pretorius, Fundamental theoretical bias in gravitational wave astrophysics and the parametrized post-Einsteinian framework, Phys. Rev. D **80**, 122003 (2009).
[38] 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**, 062003 (2011).
[39] T. G. F. Li, W. Del Pozzo, S. Vitale, C. Van Den Broeck, M. Agathos, J. Veitch, K. Grover, T. Sidery, R. Sturani, and A. Vecchio, Towards a generic test of the strong field dynamics of general relativity using compact binary coalescence, Phys. Rev. D **85**, 082003 (2012).
[40] 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**, 082001 (2014).
[41] 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**, 044033 (2018).
[42] N. Yunes, X. Siemens, and K. Yagi, Gravitational-wave tests of general relativity with ground-based detectors and pulsar-timing arrays, Living Rev. Relativity **28**, 3 (2025).
[43] K. Chatziioannou, N. Yunes, and N. Cornish, Model-independent test of general relativity: An extended post-Einsteinian framework with complete polarization content, Phys. Rev. D **86**, 022004 (2012); **95**, 129901(E) (2017).
[44] N. Loutrel, P. Pani, and N. Yunes, Parametrized post-Einsteinian framework for precessing binaries, Phys. Rev. D **107**, 044046 (2023).
[45] S. Mezzasoma and N. Yunes, Theory-agnostic framework for inspiral tests of general relativity with higher-harmonic gravitational waves, Phys. Rev. D **106**, 024026 (2022).
[46] 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**, 044020 (2023).
[47] S. A. Bhat, P. Saini, M. Favata, C. Gandevikar, C. K. Mishra, and K. G. Arun, Parametrized tests of general relativity using eccentric compact binaries, Phys. Rev. D **110**, 124062 (2024).
[48] G. S. Bonilla, P. Kumar, and S. A. Teukolsky, Modeling compact binary merger waveforms beyond general relativity, Phys. Rev. D **107**, 024015 (2023).
[49] 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**, 024043 (2023).
[50] D. Watarai, A. Nishizawa, and K. Cannon, Physically consistent gravitational waveform for capturing beyond general relativity effects in the compact object merger phase, Phys. Rev. D **109**, 084058 (2024).
[51] D. Watarai, A. Nishizawa, H. Takeda, H. Imafuku, and K. Cannon, Observational constraints on the nonlinear regime of gravity with a parametrized beyond-GR gravitational waveform model, Phys. Rev. D **113**, 024015 (2026).
[52] 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**, 024036 (2024).
[53] L. Sampson, N. Cornish, and N. Yunes, Mismodeling in gravitational-wave astronomy: The trouble with templates, Phys. Rev. D **89**, 064037 (2014).
[54] B. C. Seymour and Y. Chen, Gravitational-wave signatures of nonviolent non-locality, Phys. Rev. Lett. **136**, 011401 (2026).
[55] E. E. Flanagan and S. A. Hughes, Measuring gravitational waves from binary black hole coalescences: 2. The waves' information and its extraction, with and without templates, Phys. Rev. D **57**, 4566 (1998).

[56] L. Lindblom, B. J. Owen, and D. A. Brown, Model waveform accuracy standards for gravitational wave data analysis, Phys. Rev. D **78**, 124020 (2008).
[57] P. Kumar, K. Barkett, S. Bhagwat, N. Afshari, D. A. Brown, G. Lovelace, M. A. Scheel, and B. Szilágyi, Accuracy and precision of gravitational-wave models of inspiraling neutron star-black hole binaries with spin: Comparison with matter-free numerical relativity in the low-frequency regime, Phys. Rev. D **92**, 102001 (2015).
[58] K. Chatziioannou, A. Klein, N. Yunes, and N. Cornish, Constructing gravitational waves from generic spin-precessing compact binary inspirals, Phys. Rev. D **95**, 104004 (2017).
[59] M. Pürrer and C.-J. Haster, Gravitational waveform accuracy requirements for future ground-based detectors, Phys. Rev. Res. **2**, 023151 (2020).
[60] Q. Hu and J. Veitch, Assessing the model waveform accuracy of gravitational waves, Phys. Rev. D **106**, 044042 (2022).
[61] S. T. McWilliams, B. J. Kelly, and J. G. Baker, Observing mergers of non-spinning black-hole binaries, Phys. Rev. D **82**, 024014 (2010).
[62] A. Toubiana and J. R. Gair, Indistinguishability criterion and estimating the presence of biases, arXiv:2401.06845.
[63] T. Knapp, K. Chatziioannou, K. Mitman, M. A. Scheel, M. Boyle, L. E. Kidder, and H. Pfeiffer, A comprehensive look into the accuracy of SpEC binary black hole waveforms, Phys. Rev. D **113**, 124001 (2026).
[64] J. E. Thompson, C. Hoy, E. Fauchon-Jones, and M. Hannam, On the use and interpretation of signal-model indistinguishability measures for gravitational-wave astronomy, Phys. Rev. D **112**, 064011 (2025).
[65] C. Cutler and M. Vallisneri, LISA detections of massive black hole inspirals: Parameter extraction errors due to inaccurate template waveforms, Phys. Rev. D **76**, 104018 (2007).
[66] Q. Hu and J. Veitch, Accumulating errors in tests of general relativity with gravitational waves: Overlapping signals and inaccurate waveforms, Astrophys. J. **945**, 103 (2023).
[67] C. B. Owen, C.-J. Haster, S. Perkins, N. J. Cornish, and N. Yunes, Waveform accuracy and systematic uncertainties in current gravitational wave observations, Phys. Rev. D **108**, 044018 (2023).
[68] V. Kapil, L. Reali, R. Cotesta, and E. Berti, Systematic bias from waveform modeling for binary black hole populations in next-generation gravitational wave detectors, Phys. Rev. D **109**, 104043 (2024).
[69] A. Dhani, S. H. Völkel, A. Buonanno, H. Estelles, J. Gair, H. P. Pfeiffer, L. Pompili, and A. Toubiana, Systematic biases in estimating the properties of black holes due to inaccurate gravitational-wave models, Phys. Rev. X **15**, 031036 (2025).
[70] L. Capuano, M. Vaglio, R. S. Chandramouli, C. L. Pitte, A. Kuntz, and E. Barausse, Systematic bias in LISA ringdown analysis due to waveform inaccuracy, Phys. Rev. D **112**, 104031 (2025).
[71] S. H. Völkel and A. Dhani, Quantifying systematic biases in black hole spectroscopy, Phys. Rev. D **112**, 084076 (2025).
[72] M. Vallisneri, Testing general relativity with gravitational waves: A reality check, Phys. Rev. D **86**, 082001 (2012).
[73] M. Vallisneri and N. Yunes, Stealth bias in gravitational-wave parameter estimation, Phys. Rev. D **87**, 102002 (2013).
[74] S. Vitale and W. Del Pozzo, How serious can the stealth bias be in gravitational wave parameter estimation?, Phys. Rev. D **89**, 022002 (2014).
[75] C. J. Moore, E. Finch, R. Buscicchio, and D. Gerosa, Testing general relativity with gravitational-wave catalogs: The insidious nature of waveform systematics, iScience **24**, 102577 (2021).
[76] S. Perkins and N. Yunes, Are parametrized tests of general relativity with gravitational waves robust to unknown higher post-Newtonian order effects?, Phys. Rev. D **105**, 124047 (2022).
[77] M. Favata, Systematic parameter errors in inspiraling neutron star binaries, Phys. Rev. Lett. **112**, 101101 (2014).
[78] R. S. Chandramouli, K. Prokup, E. Berti, and N. Yunes, Systematic biases in parametrized tests of general relativity due to waveform mismodeling: The impact of neglecting spin precession and higher modes, Phys. Rev. D **111**, 044026 (2025).
[79] A. Gupta *et al.*, Possible causes of false general relativity violations in gravitational wave observations, SciPost Phys. Comm. Rep. 5 (2024).
[80] P. Saini, M. Favata, and K. G. Arun, Systematic bias on parametrized tests of general relativity due to neglect of orbital eccentricity, Phys. Rev. D **106**, 084031 (2022).
[81] S. A. Bhat, P. Saini, M. Favata, and K. G. Arun, Systematic bias on the inspiral-merger-ringdown consistency test due to neglect of orbital eccentricity, Phys. Rev. D **107**, 024009 (2023).
[82] P. Saini, S. A. Bhat, M. Favata, and K. G. Arun, Eccentricity-induced systematic error on parametrized tests of general relativity: Hierarchical Bayesian inference applied to a binary black hole population, Phys. Rev. D **109**, 084056 (2024).
[83] P. Narayan, N. K. Johnson-McDaniel, and A. Gupta, Effect of ignoring eccentricity in testing general relativity with gravitational waves, Phys. Rev. D **108**, 064003 (2023).
[84] 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. R. Astron. Soc. **535**, 3283 (2024).
[85] S. Kejriwal, L. Speri, and A. J. K. Chua, Impact of correlations on the modeling and inference of beyond vacuum–general relativistic effects in extreme-mass-ratio inspirals, Phys. Rev. D **110**, 084060 (2024).
[86] N. Yunes, M. Coleman Miller, and J. Thornburg, The effect of massive perturbers on extreme mass-ratio inspiral waveforms, Phys. Rev. D **83**, 044030 (2011).
[87] K. Chamberlain, C. J. Moore, D. Gerosa, and N. Yunes, Frequency-domain waveform approximants capturing Doppler shifts, Phys. Rev. D **99**, 024025 (2019).
[88] T. Robson, N. J. Cornish, N. Tamanini, and S. Toonen, Detecting hierarchical stellar systems with LISA, Phys. Rev. D **98**, 064012 (2018).
[89] H. Yu and Y. Chen, Direct determination of supermassive black hole properties with gravitational-wave radiation

from surrounding stellar-mass black hole binaries, Phys. Rev. Lett. **126**, 021101 (2021).
[90] H. Yu, Y. Wang, B. Seymour, and Y. Chen, Detecting gravitational lensing in hierarchical triples in galactic nuclei with space-borne gravitational-wave observatories, Phys. Rev. D **104**, 103011 (2021).
[91] A. Laeuger, B. Seymour, Y. Chen, and H. Yu, Measuring supermassive black hole properties via gravitational radiation from eccentrically orbiting stellar mass black hole binaries, Phys. Rev. D **109**, 064086 (2024).
[92] B. Deme, B.-M. Hoang, S. Naoz, and B. Kocsis, Detecting Kozai–Lidov imprints on the gravitational waves of intermediate-mass black holes in galactic nuclei, Astrophys. J. **901**, 125 (2020).
[93] R. S. Chandramouli and N. Yunes, Ready-to-use analytic model for gravitational waves from a hierarchical triple with Kozai-Lidov oscillations, Phys. Rev. D **105**, 064009 (2022).
[94] P. Gupta, H. Suzuki, H. Okawa, and K.-i. Maeda, Gravitational waves from hierarchical triple systems with Kozai-Lidov oscillation, Phys. Rev. D **101**, 104053 (2020).
[95] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914, Classical Quantum Gravity **33**, 134001 (2016).
[96] S. Ghonge, J. Brandt, J. M. Sullivan, M. Millhouse, K. Chatziioannou, J. A. Clark, T. Littenberg, N. Cornish, S. Hourihane, and L. Cadonati, Assessing and mitigating the impact of glitches on gravitational-wave parameter estimation: A model agnostic approach, Phys. Rev. D **110**, 122002 (2024).
[97] 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**, 024066 (2022).
[98] G. Ashton, Gaussian processes for glitch-robust gravitational-wave astronomy, Mon. Not. R. Astron. Soc. **520**, 2983 (2023).
[99] A. Spadaro, R. Buscicchio, D. Vetrugno, A. Klein, S. Vitale, R. Dolesi, W. J. Weber, and M. Colpi, Glitch systematics on the observation of massive black-hole binaries with LISA, Phys. Rev. D **108**, 123029 (2023).
[100] A. Pai and K. G. Arun, Singular value decomposition in parametrised tests of post-Newtonian theory, Classical Quantum Gravity **30**, 025011 (2013).
[101] K. G. Arun and A. Pai, Tests of general relativity and alternative theories of gravity using gravitational wave observations, Int. J. Mod. Phys. D **22**, 1341012 (2013).
[102] K. Cannon, A. Chapman, C. Hanna, D. Keppel, A. C. Searle, and A. J. Weinstein, Singular value decomposition applied to compact binary coalescence gravitational-wave signals, Phys. Rev. D **82**, 044025 (2010).
[103] K. Cannon, C. Hanna, and D. Keppel, Efficiently enclosing the compact binary parameter space by singular-value decomposition, Phys. Rev. D **84**, 084003 (2011).
[104] K. Cannon, C. Hanna, and D. Keppel, Interpolating compact binary waveforms using the singular value decomposition, Phys. Rev. D **85**, 081504 (2012).
[105] D. Keppel, Use of singular-value decomposition in gravitational-wave data analysis, Int. J. Mod. Phys. Conf. Ser. **23**, 01113 (2013).
[106] M. Tiglio and A. Villanueva, Reduced order and surrogate models for gravitational waves, Living Rev. Relativity **25**, 2 (2022).
[107] S. Datta, M. Saleem, K. G. Arun, and B. S. Sathyaprakash, Multiparameter tests of general relativity using a principle component analysis with next-generation gravitational-wave detectors, Phys. Rev. D **109**, 044036 (2024).
[108] M. Saleem, S. Datta, K. G. Arun, and B. S. Sathyaprakash, Parametrized tests of post-Newtonian theory using principal component analysis, Phys. Rev. D **105**, 084062 (2022).
[109] S. Datta, Enhancing the performance of multiparameter tests of general relativity with LISA using principal component analysis, arXiv:2303.04399.
[110] Z.-C. Ma, R. Niu, and W. Zhao, The multi-parameter test of gravitational wave dispersion with principal component analysis, Res. Astron. Astrophys. **24**, 055012 (2024).
[111] P. Mahapatra *et al.*, Confronting general relativity with principal component analysis: Simulations and results from GWTC-3 events, Phys. Rev. D **112**, 104007 (2025).
[112] A. A. Shoom, P. K. Gupta, B. Krishnan, A. B. Nielsen, and C. D. Capano, Testing the post-Newtonian expansion with GW170817, Gen. Relativ. Gravit. **55**, 55 (2023).
[113] K. G. Arun, B. R. Iyer, M. S. S. Qusailah, and B. S. Sathyaprakash, Testing post-Newtonian theory with gravitational wave observations, Classical Quantum Gravity **23**, L37 (2006).
[114] K. G. Arun, B. R. Iyer, M. S. S. Qusailah, and B. S. Sathyaprakash, Probing the non-linear structure of general relativity with black hole binaries, Phys. Rev. D **74**, 024006 (2006).
[115] S. H. Völkel, N. Franchini, and E. Barausse, Theory-agnostic reconstruction of potential and couplings from quasinormal modes, Phys. Rev. D **105**, 084046 (2022).
[116] B. C. Seymour, J. Golomb, and Y. Chen, waveform-geometry-testing-gr, version v1.0.0, Zenodo, 10.5281/zenodo.20931755 (2026).
[117] 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**, 2658 (1994).
[118] M. Vallisneri, Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects, Phys. Rev. D **77**, 042001 (2008).
[119] C. L. Rodriguez, B. Farr, W. M. Farr, and I. Mandel, Inadequacies of the Fisher information matrix in gravitational-wave parameter estimation, Phys. Rev. D **88**, 084013 (2013).
[120] L. S. Finn, Detection, measurement and gravitational radiation, Phys. Rev. D **46**, 5236 (1992).
[121] C. K. Mishra, K. G. Arun, B. R. Iyer, and B. S. Sathyaprakash, Parametrized tests of post-Newtonian theory using Advanced LIGO and Einstein Telescope, Phys. Rev. D **82**, 064010 (2010).
[122] https://github.com/BrianCSeymour/waveform-geometry-testing-gr/blob/main/ppe-perpendicular-plots.ipynb.
[123] P. Ajith *et al.*, Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins, Phys. Rev. Lett. **106**, 241101 (2011).

[124] P. Schmidt, F. Ohme, and M. Hannam, Towards models of gravitational waveforms from generic binaries II: Modelling precession effects with a single effective precession parameter, Phys. Rev. D **91**, 024043 (2015).
[125] https://github.com/BrianCSeymour/waveform-geometry-testing-gr/blob/main/overlap-plots.ipynb.
[126] https://github.com/BrianCSeymour/waveform-geometry-testing-gr/blob/main/ppe-nvnl-overlap.ipynb.
[127] F. Ohme, A. B. Nielsen, D. Keppel, and A. Lundgren, Statistical and systematic errors for gravitational-wave inspiral signals: A principal component analysis, Phys. Rev. D **88**, 042002 (2013).
[128] 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**, 201101 (2020).
[129] E. Payne, M. Isi, K. Chatziioannou, and W. M. Farr, Fortifying gravitational-wave tests of general relativity against astrophysical assumptions, Phys. Rev. D **108**, 124060 (2023).