

Tests of general relativity with GW230529: A neutron star merging with a lower mass-gap compact object

Elise M. Sanger^{1,*}, Soumen Roy^{2,3,†}, Michalis Agathos^{4,5}, Ofek Birnholtz⁶, Alessandra Buonanno^{1,7},
Tim Dietrich^{8,1}, Maria Haney⁹, Felix-Louis Julie¹, Geraint Pratten⁹, Jan Steinhoff¹, Chris Van Den Broeck^{2,3},
Sylvia Biscoveanu¹⁰, Prasanta Char¹¹, Anna Heffernan¹², Prathamesh Joshi¹³, Atul Kedia¹⁴,
R. M. S. Schofield¹⁵, M. Trevor⁷ and Michael Zevin¹⁶

¹*Max Planck Institute for Gravitational Physics (Albert Einstein Institute),
Am Muhlenberg 1, Potsdam, 14476, Germany*

²*Nikhef—National Institute for Subatomic Physics, Science Park 105,
1098 XG Amsterdam, The Netherlands*

³*Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University,
Princetonplein 1, 3584 CC Utrecht, The Netherlands*

⁴*Queen Mary University of London, Mile End Road, London, E1 4NS, United Kingdom*

⁵*Department of Applied Mathematics and Theoretical Physics, University of Cambridge,
Wilberforce Road CB3 0WA Cambridge, United Kingdom*

⁶*Bar-Ilan University, Ramat Gan, 5290002, Israel*

⁷*Department of Physics, University of Maryland, College Park, Maryland 20742, USA*

⁸*Institut fur Physik und Astronomie, Universitat Potsdam, Haus 28, Karl-Liebknecht-Strae 24/25,
14476, Potsdam, Germany*

⁹*School of Physics and Astronomy and Institute for Gravitational Wave Astronomy,
University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom*

¹⁰*Center for Interdisciplinary Exploration and Research in Astrophysics (CIERA),
Northwestern University, Evanston, Illinois 60201, USA*

¹¹*Universite de Liege, B-4000 Liege, Belgium*

¹²*IAC3—IEEC, Universitat de les Illes Balears, E-07122 Palma, Spain*

¹³*The Pennsylvania State University, University Park, Pennsylvania 16802, USA*

¹⁴*Center for Computational Relativity and Gravitation, Rochester Institute of Technology,
Rochester, New York 14623, USA*

¹⁵*University of Oregon, Eugene, Oregon 97403, USA*

¹⁶*Adler Planetarium, 1300 South DuSable Lake Shore Drive, Chicago, Illinois, 60605, USA*



(Received 5 November 2025; accepted 26 March 2026; published 30 April 2026)

On May 29, 2023, the LIGO Livingston observatory detected the gravitational-wave signal GW230529_181500 from the merger of a neutron star with a lower mass-gap compact object. Its long inspiral signal provides a unique opportunity to test general relativity (GR) in a parameter space previously unexplored by strong-field tests. In this work, we performed parameterized inspiral tests of GR with GW230529_181500. Specifically, we search for deviations in the frequency-domain gravitational wave phase by allowing for agnostic corrections to the post-Newtonian coefficients. We performed tests with the Flexible Theory Independent and Test Infrastructure For General Relativity frameworks using several quasicircular waveform models that capture different physical effects (higher modes, spins, tides). We find that the signal is consistent with GR for all deviation parameters. Assuming the primary object is a black hole, we obtain particularly tight constraints on the dipole radiation at -1 PN order of $|\delta\hat{\varphi}_{-2}| \lesssim 8 \times 10^{-5}$, which is a factor ~ 17 times more stringent than previous bounds from the neutron star–black hole merger GW200115_042309, as well as on the 0.5 PN and 1 PN deviation parameters. We discuss some challenges that arise when analyzing this signal, namely biases due to correlations with tidal effects and the degeneracy between the 0 PN deviation parameter and the chirp mass. To illustrate the importance of GW230529_181500 for tests of GR, we mapped the agnostic -1 PN results to a class of

*Contact author: elise.saenger@aei.mpg.de

†Contact author: soumen.roy@uclouvain.be

Einstein-scalar-Gauss-Bonnet (ESGB) theories of gravity. We also conducted an analysis probing the specific phase deviation expected in ESGB theory and obtain an upper bound on the Gauss-Bonnet coupling of $\ell_{\text{GB}} \lesssim 0.51 M_{\odot}$ ($\sqrt{\alpha_{\text{GB}}} \lesssim 0.28$ km), which is better than any previously reported constraint.

DOI: [10.1103/r43k-51yq](https://doi.org/10.1103/r43k-51yq)

I. INTRODUCTION

On May 29, 2023, at 18:15:00 UTC, a gravitational wave (GW) signal emitted from the merger of two compact objects, most likely a black hole (BH) and a neutron star (NS) [1], was observed by the Advanced LIGO detector [2,3]. This event was named GW230529_181500 and will be referred to as GW230529 for brevity. It was observed only by the LIGO Livingston observatory, with a signal-to-noise ratio (SNR) of 11.6 and a false-alarm rate of less than 1 in 1000 years. Assuming that General Relativity (GR) is the correct theory to describe the GW230529 signal, the follow-up analyses estimated the source component masses to be $3.6^{+0.8}_{-1.1} M_{\odot}$ and $1.4^{+0.6}_{-0.2} M_{\odot}$ with a 90% credible interval. This puts the primary squarely in the lower mass gap of $\sim 2 - 5 M_{\odot}$, where few compact objects were observed [4–8]. This event, with the compact object within the hypothesized mass gap and with a long inspiral signal, provides an opportunity to test GR in a region of parameter space previously unexplored by strong-field tests of GR.

GW230529 is not the first observation to find support for compact objects in the lower mass gap. A number of recent studies that have found evidence for compact objects in the mass gap include observations of noninteracting binary systems [9,10], radio pulsar surveys [11], and GWs from compact binary coalescences [6,8,12,13]. When a binary is observed with component masses measured in the mass-gap range of $\sim 2 - 5 M_{\odot}$, the immediate challenge is to determine whether the component in the gap is a NS or a BH. Such an identification would be remarkable because it would either increase the known maximum mass for a NS [14,15] or decrease the known minimum mass for a BH from low-mass x-ray binary observations [16–18]. However, it has not been conclusively proven whether these components are BHs, NSs, or something else.

Recent theoretical research has suggested several potential exotic compact objects, such as gravastars [19], boson stars [20], or Planck-scale modifications of BH horizons [21,22], which could also fall into this gap. Primordial BHs formed from overdense regions in the early Universe could also fill the lower mass gap [23]. Besides astrophysical and primordial BHs in GR and exotic compact objects, NSs in alternative GR theories could also inhabit this gap. For example, axionic scalar-tensor theory with viable phenomenological equations-of-state (EOSs) can produce NSs with maximum masses larger than $2.5 M_{\odot}$, but below the $3 M_{\odot}$ threshold [24]. Therefore, the GW230529 signal provides an opportunity to perform tests of GR to uncover new

physics, as well as pointing to the existence of potential exotic compact objects.

General relativity is the simplest and most successful theory of gravity to date, and it has been rigorously validated by various experimental tests in our Solar System [25], observations of binary pulsars [26,27], cosmological data [28], GWs [29–34], and massive BHs [35]. Recently, the evidence of a stochastic GW background with pulsar timing arrays [36,37] has also led to some new gravitational tests. Despite its overall success, we do not know how to reconcile GR with quantum mechanics and how (or if) to employ GR to explain certain cosmological phenomena, such as dark matter and dark energy. These limitations have motivated the development of alternative theories of gravity, such as the Brans-Dicke [38], Horndeski scalar-tensor [39], Aether [40], Einstein-Gauss-Bonnet [41,42], and Chern-Simons [43,44] theories, and the effective-field-theory extension of GR [45], all of which have survived a wide range of experimental tests [25,46–49].

The observation of GW signals from coalescences of compact binary systems has opened the avenue for studying the two-body dynamics in strong gravitational fields with high velocities approaching the speed of light. In contrast, other observations typically probe states of either weak gravitational fields—where spacetime curvature is negligible—or where velocities are well below the speed of light. However, in binary pulsars, the strongly gravitating bodies allow for strong-field tests of gravity, albeit for low velocities. Developing an accurate GW waveform model that captures the entire evolution of compact binaries is essential for determining source characteristics, validating the predictions of GR, and exploring potential alternative theories of gravity. However, the development of waveforms within specific alternative theories has not yet reached a level of maturity sufficient to perform model comparisons with GR. While numerical-relativity simulations in beyond-GR theories are becoming more prevalent, either by solving the full equations of motions [50–57] or through approximate treatments [58–60], there are not yet sufficient simulations in one specific theory to calibrate a semianalytic inspiral-merger-ringdown model in that theory. Therefore, semianalytical inspiral-merger-ringdown models in alternative theories of gravity that are compared to GW data either leave the merger like GR [61,62] or they introduce agnostic parameters in the merger that are then marginalized over during parameter estimation [63]. Another possibility is to directly compare numerical-relativity simulations in beyond-GR theories with observational data [59], but those tests are limited by the

lengths of the numerical waveforms and the region of parameter space where simulations are available. Nevertheless, there may be a more accurate alternative theory that we are not aware of.

The alternative models often propose variances in the post-Newtonian (PN) phase coefficients of the GW signal, which are derived solving perturbatively the two-body dynamics and gravitational radiation by a series expansion in the orbital velocity, intertwined with the strength of the gravitational field. This has led to the development of theory-agnostic approaches to parametrized tests of GR [64–68]. The same framework can be used with search pipelines to detect non-GR signals [69,70]. In these parametrized tests, an additional parameter is introduced into the PN coefficients to represent generic deviations from GR and a series of Bayesian parameter estimation analyses is performed for each of these parameters. The resulting posterior distribution of a deviation parameter indicates whether generic features in the GW signal suggest that GR is not the most probable theory of gravity. Importantly, claiming a GR-violation would require performing extensive tests that prove the GR deviation is not due to systematics due to waveforms’ inaccuracy, lack of physical effects, environmental astrophysical effects, and gravitational lensing [71]. The parametrized test in the Bayesian framework has been extensively developed for the LIGO-Virgo-KAGRA (LVK) collaboration to test GR as part of the GW transient catalogs (GWTC) [32–34]. It has been performed using two implementations: the Flexible Theory-Independent (FTI) approach [31,72], which can be applied to any aligned-spin frequency-domain waveform model, and the Test Infrastructure For General Relativity (TIGER) approach [73–76], which is based on a (frequency-domain) phenomenological waveform family. In this study, we perform analyses of GW230529 using both approaches with the SEOBNRv4 [77–81] and IMRPhenomX [82–84] families, respectively.

Since alternative theories often predict the functional form of the beyond-GR correction as a parametric variation of the PN terms, the measurements of phenomenological deviation parameters can be mapped onto specific alternative theories of gravity [46,85,86]. By using the leading order correction of an alternative gravity theory, we typically provide a theory-agnostic bound on the beyond-GR parameters predicted by that theory. In this work, we use this method to provide bounds on the Gauss-Bonnet coupling in the class of Einstein-scalar-Gauss-Bonnet (ESGB) theories of gravity with $f'(0) \neq 0$. We also perform a theory-specific test for ESGB by implementing all known corrections to the PN terms [87–94]. This includes newly computed corrections at 1.5PN order, as given in Appendix C. The corrections to ESGB are larger for smaller BH masses, so GW230529 provides an excellent opportunity to probe relatively small values of the coupling. Previous tests for ESGB have placed constraints on the coupling of $\sqrt{\alpha_{\text{GB}}} \lesssim 1.18$ km by combining

bounds from multiple events [47], and another recent study of an eccentric binary obtained a bound of $\sqrt{\alpha_{\text{GB}}} \lesssim 2.38$ km [95].

The rest of this paper is organized as follows. Section II explains the parametrized inspiral tests of GR used. Section III describes the setup of the analyses performed. The results and constraints on deviations from GR obtained for GW230529 are presented in Sec. IV. Finally, in Sec. V we map the agnostic constraints to a specific modified gravity theory, namely ESGB gravity. In Appendix A, we list the priors used in the analyses. Appendix B discusses possible false violations of GR when using too wide priors on the deviation parameters, which can cause wraparound in the waveform. Lastly, we list in Appendix C the ESGB corrections up to 1.5PN order in the GW phase. Throughout this paper we follow the convention of $G = c = 1$.

II. PARAMETERIZED INSPIRAL TESTS OF GR

In GR, the GW signal from the early inspiral of a quasicircular compact binary can be approximated using the PN formalism, which expands the waveform in powers of the velocity v , where $\mathcal{O}(v^{2n})$ relative to the leading order term is called the n PN order [96,97]. The frequency domain phase of a GW signal can then be obtained using the stationary phase approximation [98]. In GR, it is given by

$$\Psi_{\ell m}^{\text{GR}}(f) = 2\pi f t_c - \phi_c - \frac{\pi}{4} + \frac{3}{128\eta v^5} \frac{m}{2} \sum_{n=0}^7 (\psi_n^{\text{GR}} + \psi_{n(l)}^{\text{GR}} \log v) v^n, \quad (1)$$

where t_c, ϕ_c are the time and phase at coalescence, $v = (2\pi f M/m)^{1/3}$ with f the GW frequency, $M = m_1 + m_2$ is the total mass, and $\eta = m_1 m_2 / M^2$ is the symmetric mass ratio. Here, ψ_n^{GR} and $\psi_{n(l)}^{\text{GR}}$ represent the $(n/2)$ PN coefficients in GR, and they depend only on the intrinsic parameters of the binary. The subscript (l) indicates the logarithmic terms that enter at 2.5PN and 3PN orders. It is also important to note that the 0.5PN term is absent in GR. The subscript ℓm denotes the (ℓ, m) mode in the mode decomposition of the GW signal into spherical harmonics [not to be confused with the subscript (l) denoting the coefficients of the log terms]. The dominant mode is the $(2, 2)$ mode, and other modes are referred to as higher modes.

In beyond-GR theories, these PN coefficients can be different from the ones in GR [46,85,86]. To test for deviations from GR during the inspiral, we therefore add a correction to the frequency domain phase of the form

$$\delta\Psi_{\ell m}(f) = \frac{3}{128\eta v^5} \frac{m}{2} \left(\sum_{n=-2}^7 \delta\psi_n v^n + \sum_{n=5}^6 \delta\psi_{n(l)} v^n \log v \right), \quad (2)$$

where $\delta\psi_n$ and $\delta\psi_{n(l)}$ are the deviations in the $(n/2)$ PN coefficients.

For parametrized inspiral tests, we introduce deviation parameters $\delta\hat{\phi}_n$ and $\delta\hat{\phi}_{n(l)}$ that are the fractional deviations of the corresponding PN coefficients in GR. We thus have that

$$\delta\psi_n = \delta\hat{\phi}_n \psi_n^{\text{GR}}, \quad (3)$$

$$\delta\psi_{n(l)} = \delta\hat{\phi}_{n(l)} \psi_{n(l)}^{\text{GR}}. \quad (4)$$

When the $(n/2)$ PN coefficient vanishes in GR (i.e., for $n = -2, 1$), we instead let $\delta\psi_n = \delta\hat{\phi}_n$ so that it is an absolute deviation normalized to the Newtonian coefficient. We do not test for $\delta\hat{\phi}_5$ since this would give a constant phase shift and therefore be degenerate with the phase at coalescence. We also do not perform tests at -0.5 PN. In this parametrization, GR is recovered in the limit $\delta\hat{\phi}_n, \delta\hat{\phi}_{n(l)} \rightarrow 0$.

We employ two different frameworks that can perform this type of test of GR: FTI [72] and TIGER [74,75]. These two frameworks differ in the GR waveform models used and the exact implementation of the inspiral test. FTI uses the SEOBNRv4 waveform family [77–81] and is only available for aligned spins, while TIGER uses the IMRPhenomX waveform family [82–84] and can use precessing spin waveforms. The main difference in the implementation occurs at the transition from a non-GR inspiral to the GR merger-ringdown. In FTI, the testing GR corrections are added to the GW phase of an inspiral-merger-ringdown frequency-domain waveform from GR. The corrections to the frequency-domain phase are computed using Eq. (2), and then smoothly tapered off to zero using a windowing function so that the merger-ringdown signal coincides with the GR one (for details on how this is done, see Ref. [72]). This is done for every mode separately and the corrections are then added together. In the TIGER framework, the parametrized deviations are incorporated into the GR phase coefficients of the IMRPhenomXAS model. This model represents the primary spherical harmonic radiation mode for the coalescence of nonprecessing binary black hole (BBH) systems. The inspiral phase of higher-order modes in the time domain is modeled as a scaling of the dominant quadrupole mode, allowing any deviations in the inspiral phase to propagate to the higher-order modes. In the phenomenological model, the separate inspiral and merger-ringdown segments are smoothly connected by ensuring $\mathcal{C}^1$ continuous condition in both phase and amplitude. When we introduce the variations in the phase coefficients, we simultaneously adjust the phase derivative to preserve the continuity. Deviations in the inspiral phase coefficients do not affect the postinspiral portion of the waveform. Although TIGER also has the option to introduce deviations in the postinspiral part of the waveform, we do not use that in this work.

III. ANALYSIS SETUP

As baseline GR models for the FTI framework, we use the waveform models based on SEOBNRv4HM_ROM [79], which is the reduced order, frequency-domain version of the time-domain model SEOBNRv4HM [77,78]. It is an effective-one-body waveform model for BBHs that assumes quasi-circular orbits and aligned spins, and includes the modes: (2, 2), (2, 1), (3, 3), (4, 4), and (5, 5).¹ Its neutron star–black hole (NSBH) version SEOBNRv4_ROM_NRTidalv2_NSBH [80,81] contains only the dominant (2, 2)-mode, and allows for tidal deformability of the secondary object and tidal disruption. We also use its binary neutron star (BNS) version SEOBNRv4_ROM_NRTidalv2, which has tidal effects on both components and contains the (2,2)-mode only. With the TIGER framework, we use the models based on IMRPhenomXPHM [82,83], which is a phenomenological waveform model that allows for precessing spins and includes the higher modes: (2, 2), (2, 1), (3, 3), (3, 2), and (4, 4). Its BNS version IMRPhenomXP_NRTidalv2 [80,84] allows for tides on both component objects and is (2, 2) mode only. To investigate the effect of waveform systematics due to the presence of precession, higher harmonics, or difference in waveform treatment, we also perform the parametrized test using SEOBNRv4_ROM, IMRPhenomXP, and IMRPhenomXHM models. An overview of the models used and the physics they include is given in Table I. The waveforms are generated using Bilby TGR [99] and LALSimulation [100,101].

For parameter estimation, we employ Bayesian inference using Bilby [102–104]. We use the standard likelihood function assuming additive noise that is stationary and Gaussian. The one-dimensional posteriors for the deviation parameters are obtained by marginalizing over all other parameters. The sampling algorithm used is nested sampling with the DYNESTY sampler [105]. We only allow one deviation parameter to vary at a time and repeat the analysis for each deviation parameter. We use uniform priors for the deviation parameters. The priors used are listed in Appendix A.

We analyze 128 s of data from the LIGO Livingston Observatory with a sampling frequency of 4096 Hz. The analysis is done over the frequency range 20 Hz to 1792 Hz when calculating the likelihood. The power spectral density used to describe the noise was produced with BayesWave [106,107].

IV. RESULTS

The posteriors for the different deviation parameters are shown in Fig. 1. The blue violins on the left are obtained with TIGER, while the orange violins on the right are results from FTI. The filled violins do not have tidal effects, while the unfilled violins with dash and continuous lines

¹We do not include the (5, 5)-mode in our FTI analyses, but do use all other modes available.

TABLE I. An overview of the different waveform models that we use in this work and the physics they include [higher modes or (2,2)-mode only, spin precession or aligned spins only, allowing for tidal effects]. We will refer to them throughout this paper using the shortened name and use the color indicated in the plots.

Waveform model	Short name	Color	Higher modes	Spin precession	Tides
SEOBNRv4HM_ROM [77–79]	SEOBHM	Light orange	✓	...	...
SEOBNRv4_ROM [77]	SEOB		...	...	...
SEOBNRv4_ROM_NRTidalv2_NSBH [80,81]	SEOBNSBH	Red	...	...	✓ (NSBH)
SEOBNRv4_ROM_NRTidalv2 [79,80]	SEOBT	Dark orange	...	...	✓ (BNS)
IMRPhenomXPHM [82]	PhenomXPHM	Light blue	✓	✓	...
IMRPhenomXHM [83]	PhenomXHM		✓	...	...
IMRPhenomXP [82]	PhenomXP		...	✓	...
IMRPhenomXP_NRTidalv2 [80,84]	PhenomXPT	Dark blue	...	✓	✓ (BNS)

are for waveform models including tidal effects. Table II shows the values for the deviation parameters found for GW230529 and the quantile corresponding to GR. We see that all results obtained are consistent with GR, except for the 0PN results. We will first discuss in Sec. IV A in more detail the results obtained with BBH waveforms. Then, in Sec. IV B, we explain the effects that tides in the NSBH and BNS models have on the results. Finally, we will discuss the 0PN results in more detail in Sec. IV C.

A. Binary black hole waveforms

The results for the aligned-spin PhenomXHM, spin-precessing (2, 2)-mode PhenomXP, and aligned-spin (2,2)-mode SEOB are not shown in the violin plot, but they are similar to the spin-precessing PhenomXPHM and aligned-spin SEOBHM results, respectively. It is not surprising that the results for spin-precessing PhenomXPHM and aligned-spin

PhenomXHM agree, since for GW230529 there was no evidence found for spin-precession (the posteriors on the effective precessing spin were uninformative [1]). It was also expected that the results for PhenomXPHM and PhenomXP agree well, because for the SNR and source parameters of GW230529, the higher modes are not expected to strongly contribute. The posteriors for SEOB are slightly wider than for SEOBHM, but otherwise agree well. The increase in width is most likely due to a slight loss in SNR when not including the higher-order modes.

Let us compare the FTI and TIGER results for BBH waveform models by comparing PhenomXPHM (blue filled) and SEOBHM (orange filled) results with each other. Differences between these results can come from waveform systematics, for example the details of how the waveforms are constructed and the approximations made therein, or from differences in the implementation of the inspiral test of GR in FTI and TIGER. This is a good consistency check

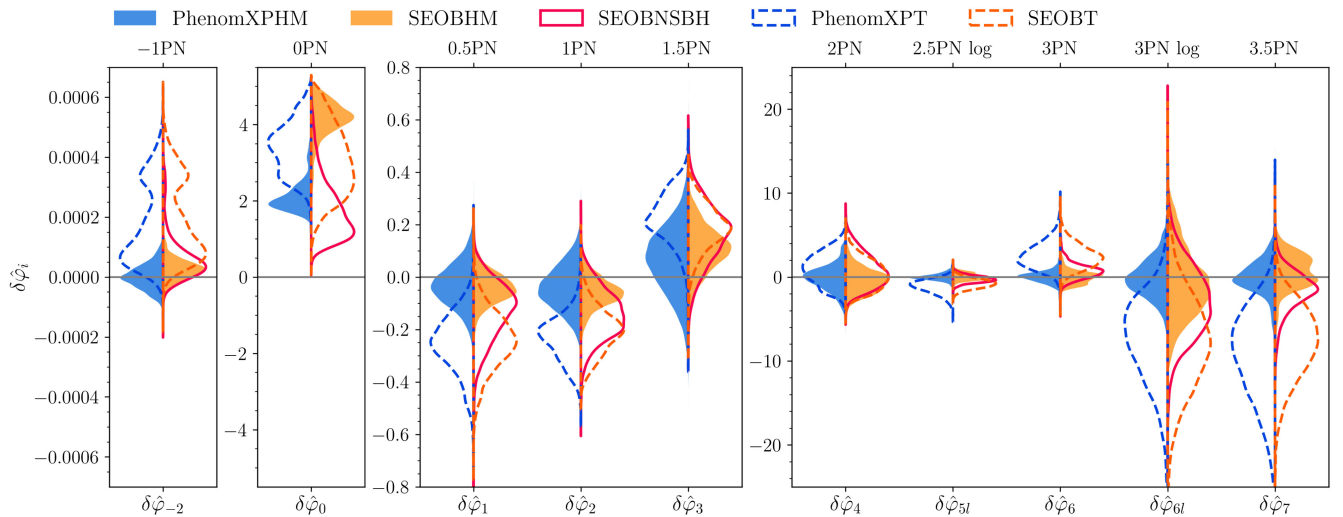


FIG. 1. Posterior distributions for the different deviation parameters $\delta\hat{\phi}_n, \delta\hat{\phi}_{n(l)}$ for GW230529. The blue histograms are obtained with TIGER using the IMRPhenomX waveform family. The orange posteriors are results from FTI using the SEOBNRv4 waveform family. The filled violins are for BBH waveform models, while the dashed lines are BNS models that include tidal effects on both components, and the red solid line is an NSBH model with tides only on the secondary. GR is recovered at $\delta\hat{\phi}_i = 0$, which is indicated by the horizontal gray line. All results are consistent with GR except for 0PN, which will be discussed in more detail in Sec. IV C.

TABLE II. The values for the different deviation parameters $\delta\hat{\phi}_n, \delta\hat{\phi}_{n(l)}$ obtained with GW230529. Displayed are the median values and 90% confidence intervals inferred. We also display the quantile Q_{GR} corresponding to the GR value $\delta\hat{\phi}_i = 0$.

Parameter	SEOBHM		SEOBNSBH		PhenomXPHM		PhenomXPT	
	$\delta\hat{\phi}_i$	$Q_{\text{GR}} (\%)$	$\delta\hat{\phi}_n$	$Q_{\text{GR}} (\%)$	$\delta\hat{\phi}_i$	$Q_{\text{GR}} (\%)$	$\delta\hat{\phi}_i$	$Q_{\text{GR}} (\%)$
$\delta\hat{\phi}_{-2}$	$2.5^{+7.2}_{-6.1} \times 10^{-5}$	23	$4.6^{+22}_{-6.8} \times 10^{-5}$	14	$0.1^{+8.7}_{-6.8} \times 10^{-5}$	49	$14^{+27}_{-12} \times 10^{-5}$	3
$\delta\hat{\phi}_0$	$4.2^{+0.6}_{-1.2}$	0	$1.6^{+1.7}_{-0.8}$	0	$2.0^{+1.2}_{-0.4}$	0	$3.4^{+1.1}_{-1.0}$	0
$\delta\hat{\phi}_1$	$-0.04^{+0.11}_{-0.10}$	75	$-0.12^{+0.15}_{-0.21}$	90	$-0.04^{+0.15}_{-0.15}$	67	$-0.25^{+0.17}_{-0.22}$	99
$\delta\hat{\phi}_2$	$-0.06^{+0.11}_{-0.11}$	82	$-0.14^{+0.16}_{-0.15}$	93	$-0.04^{+0.15}_{-0.14}$	69	$-0.23^{+0.13}_{-0.16}$	99
$\delta\hat{\phi}_3$	$0.11^{+0.15}_{-0.16}$	13	$0.17^{+0.18}_{-0.21}$	9	$0.07^{+0.21}_{-0.23}$	30	$0.23^{+0.15}_{-0.17}$	2
$\delta\hat{\phi}_4$	$-0.4^{+2.7}_{-2.3}$	61	$0.5^{+3.7}_{-2.6}$	40	$-0.1^{+2.8}_{-2.9}$	51	$1.0^{+3.0}_{-3.0}$	32
$\delta\hat{\phi}_{5l}$	$0.01^{+0.79}_{-0.60}$	49	$-0.27^{+0.74}_{-0.68}$	73	$-0.03^{+0.81}_{-0.68}$	54	$-1.1^{+1.2}_{-1.7}$	94
$\delta\hat{\phi}_6$	$0.0^{+1.6}_{-1.3}$	49	$0.7^{+1.8}_{-1.4}$	21	$0.1^{+1.4}_{-1.2}$	46	$2.4^{+2.9}_{-2.1}$	3
$\delta\hat{\phi}_{6l}$	$-0.9^{+8.1}_{-6.1}$	59	$-3.0^{+7.3}_{-6.7}$	75	$-0.2^{+4.5}_{-4.0}$	54	$-7.0^{+7.2}_{-9.9}$	95
$\delta\hat{\phi}_7$	$0.6^{+3.3}_{-4.3}$	40	$-1.5^{+3.4}_{-4.7}$	77	$-0.1^{+3.7}_{-4.7}$	52	$-8.9^{+7.7}_{-10}$	97

to see that both pipelines work as intended. Looking at the filled violins in Fig. 1, we see that the results are consistent with each other (except for $\delta\hat{\phi}_0$). There are sometimes some minor differences between the SEOBHM and PhenomXHM, e.g., one posterior is slightly wider than the other, but there are no major differences.

B. Waveforms with tidal effects

When looking at the waveforms with tidal effects in Fig. 1, we see that there are no major differences between PhenomXPT (blue dashed line) and SEOBT (orange dashed line). However, we notice that there are significant differences compared to the waveforms without tidal effects (filled violins). The posteriors for PhenomXPT and SEOBT, which have tides on both compact objects, are shifted away from GR and are wider compared to the results without tidal effects. This also happens to a lesser amount for SEOBNSBH (red solid line), which only has tides on the secondary object.

The NRTidalv2 models [80] modify the BBH waveforms to account for differences in the waveform coming from tides and other NS matter effects. Both the phase and amplitude are modified due to tidal effects, which are modeled using the tidal deformabilities Λ_i of the two objects. For BHs, the tidal deformability is zero, so the NSBH model SEOBNSBH enforces $\Lambda_1 = 0$. We do not expect changes in the amplitude to significantly influence our results because, in the parametrized inspiral tests, we only modify the frequency domain phase. We therefore only expect the change in phase evolution to influence the posteriors on the deviation parameters for the BNS and NSBH models.

The merger-ringdown is modeled differently for BNS and NSBH models. The merger frequency of the PhenomXPT and SEOBT models is calibrated to NR. Above this frequency, the amplitude is smoothly tapered off to zero in an agnostic way [80]. The reason for this is that the merger-ringdown of BNSs is not modeled well enough, and, more importantly, the merger frequency is typically

high enough that the postmerger signal is outside the sensitive frequency band of current detectors, hence, no bias is expected during parameter estimation [108]. When using these models for GW230529, the merger frequency is lower and inside the frequency band analyzed. To check that the amplitude tapering is not influencing our results, we also did an analysis with a maximum frequency of 400 Hz, which is below the merger frequency. No differences were found between the posteriors, so the amplitude tapering is not causing the shift away from GR that we observe when using PhenomXPT and SEOBT.

For SEOBNSBH, the merger-ringdown is modeled differently. For NSBHs, tidal disruption of the NS can happen when the tidal disruption frequency is lower than the ringdown frequency [81]. If this happens, then the amplitude is smoothly tapered off above the tidal disruption frequency. Otherwise, the merger-ringdown is modeled similarly to the merger-ringdown of BBHs, with a slightly suppressed amplitude. For GW230529, the NS most likely gets tidally disrupted [1,109]. However, the tidal disruption frequency is above the maximum frequency used in our analysis [1,81]. We therefore do not expect tidal disruption to influence our results.

During inspiral, both the phase and amplitude are modified due to tidal effects. Because the tidal deformability is always positive (at least for NSs), the change in phase is always in the same direction. The deviation parameters try to cancel this effect by changing the phase in the opposite way. This means that a nonzero Λ will always lead to a specific $\delta\hat{\phi}_n$ to shift toward either positive or negative values (depending on the sign of the PN coefficient in GR), but not both. This explains why the deviation parameters all shift away from zero for the BNS and NSBH models compared to the BBH models. The shift is less strong for the NSBH model since there is only one tidal parameter that is nonzero so the tides have less of an effect. The extra tidal parameters also lead to broadening of the posteriors, especially for low SNR signals where their

values are not well measured. Of course, this explanation is oversimplified due to correlations between the tides and other GR parameters as well as the deviation parameters, but nonetheless it explains why there is a shift away from GR and not just a broadening of the posteriors.

To check that this reasoning is correct, we look at the Λ_i — $\delta\hat{\varphi}_n$ posteriors in Fig. 2. The top plot shows the posteriors from an NSBH run for the 0.5PN deviation parameter (red). We see in the 2D posterior that $\delta\hat{\varphi}_1$ shifts away from 0 for larger values of Λ_2 . This can be seen in the histogram where it has more support for negative values of $\delta\hat{\varphi}_1$ than is the case for the corresponding BBH run (orange). If we restrict the tidal deformability to $\Lambda_2 < 1000$ (green shaded area), which is a more realistic range based on the mass of the secondary and current constraints on the NS EOS [110] and it is consistent with what was found using an astrophysically motivated prior in Ref. [1], then the shift away from GR disappears and the posterior (green histogram) becomes similar to the BBH one.

In the middle and bottom plots of Fig. 2 we look at BNS runs (dark blue) for the -1 PN deviation parameter (middle) and 3.5 PN deviation parameter (bottom). Here we see the same effect where $\delta\hat{\varphi}_{-2}$ and $\delta\hat{\varphi}_7$ are correlated with Λ_1 and they shift away from GR for large values of Λ_1 . If we now restrict the tides to small values of $\Lambda_1 < 300$ ² and the more realistic values of $\Lambda_2 < 1000$ (green shaded area) [1,110], we once again see that this leads to the posterior (green histogram) shifting to the BBH one (light blue). This behavior is also seen for the other deviation parameters.

We conclude that the shift away from GR of the posteriors on the deviation parameters for the SEOBNSBH and especially the PhenomXPT and SEOB waveforms is most likely due to correlations between the deviation parameters and the tidal deformability combined with the latter not being well constrained. Limiting the tidal deformability of the two objects to more realistic values based on their masses removes this shift away from GR. Since the NSBH and BNS results with the tides constrained to realistic values are similar to the BBH results, we will use the BBH results from now on (unless otherwise specified), even though GW230529 is most likely a NSBH [1].

C. Degeneracy between 0PN-deviation parameter and chirp mass

At low frequencies (i.e., during early inspiral), the phase of a GR waveform is dominated by the 0PN term. In our

²More realistically, using the EOS set of [110], the tidal deformability for a NS with masses above $2.5M_\odot$ (if such masses are supported) would be $\Lambda \lesssim 5$. Restricting Λ_1 to such low values would however leave us with too few samples to say anything sensible. We therefore went with a larger cutoff value that is a balance between the number of samples left and keeping the tides realistic.

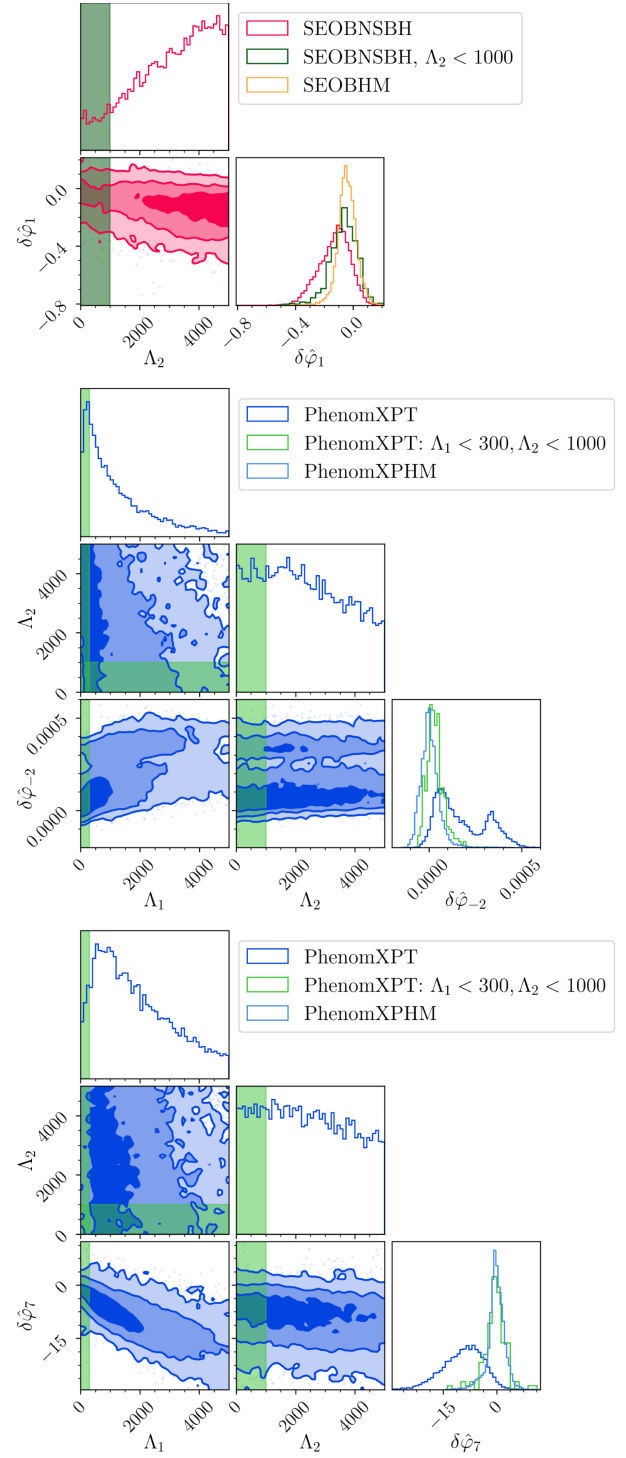


FIG. 2. 2D posteriors between the tidal deformability Λ_i and deviation parameters $\delta\hat{\varphi}_n$ (red and dark blue). The green shaded regions are more realistic ranges for the tides, which, when restricting the posterior to those regions, give the green histograms. The orange and light blue histograms are the posterior obtained with the comparable BBH models. The top plot is for an NSBH waveform, while the other two are with a BNS waveform. We see that there is a correlation between the tides and the deviation parameters, which leads to a shift away from GR. When restricting the tides to more reasonable values, this shift disappears.

testing GR waveform model, the phase contribution from the 0PN term is given by

$$\Psi_{\ell m}^{\text{OPN}}(f) = \frac{3}{128\eta v^5} \frac{m}{2} (1 + \delta\hat{\varphi}_0) \psi_0^{\text{GR}} = C_0 \frac{1 + \delta\hat{\varphi}_0}{\mathcal{M}_c^{5/3}} f^{-5/3}, \quad (5)$$

where C_0 is a constant that does not depend on the intrinsic parameters of the binary and $\mathcal{M}_c = \eta^{3/5} M$ is the chirp mass. This means that in the frequency regime where the 0PN term is dominant, the deviation parameter $\delta\hat{\varphi}_0$ is degenerate with the chirp mass of the binary, so the two are strongly correlated (see also Sec. IV C in Ref. [72]). Assuming that the binary is consistent with GR and that the true chirp mass is equal to $\mathcal{M}_c^{\text{GR}}$, this correlation is given by

$$\delta\hat{\varphi}_0 = \left(\frac{\mathcal{M}_c}{\mathcal{M}_c^{\text{GR}}} \right)^{5/3} - 1. \quad (6)$$

For low chirp-mass systems, the late inspiral and merger-ringdown are outside the frequency band observed by the GW detectors. For such systems, this degeneracy at 0PN means that, when doing inspiral tests of GR at 0PN, the chirp mass can take any value and $\delta\hat{\varphi}_0$ can be used to compensate without significantly altering the waveform. For high enough chirp mass, the late inspiral and merger-ringdown move into the frequency band used for analysis. This eventually breaks the degeneracy, and such high values of the chirp mass would then be disfavored. This means that the $\mathcal{M}_c$ — $\delta\hat{\varphi}_0$ posterior is more or less free to

move along the line of the degeneracy between the two, until it reaches too high values of the chirp mass.

Figure 3 shows the 2D posterior of the chirp mass $\mathcal{M}_c$ and 0PN deviation parameter $\delta\hat{\varphi}_0$ for GW230529 obtained with SEOBHM (orange). The maximum likelihood sample from a GR run with SEOBHM is highlighted in green. This is also the value used in Eq. (6) to compute the correlation (dark green). We see that the posterior closely follows the correlation that is expected from the 0PN degeneracy.

To check that this degeneracy should be present in GW230529-like signals, we compute the mismatch between the maximum likelihood GR waveform and non-GR waveforms using PhenomXPHM. We vary the chirp mass and 0PN deviation coefficient, but keep all other parameters the same, and the mismatch is minimized over time and phase. Figure 4 gives a map of these mismatches. We see that there is a line along which the mismatch is small. This line corresponds to the correlation in Eq. (6). We notice that the low mismatch region becomes wider for larger chirp mass (see also the insets).

As an extra check that this strong correlation should indeed be present in a GR signal with similar source parameters as GW230925, we inject the maximum likelihood GR waveform using SEOBHM into a zero-noise realization of the LIGO Livingston detector with the same power spectral density as at the time of GW230529. We then evaluate the likelihood for random values of the chirp mass and 0PN deviation parameter, while keeping the other parameters the same as for the injected waveform. This gives the likelihood map shown in Fig. 5. The red star indicates the injected values, and the red line shows the

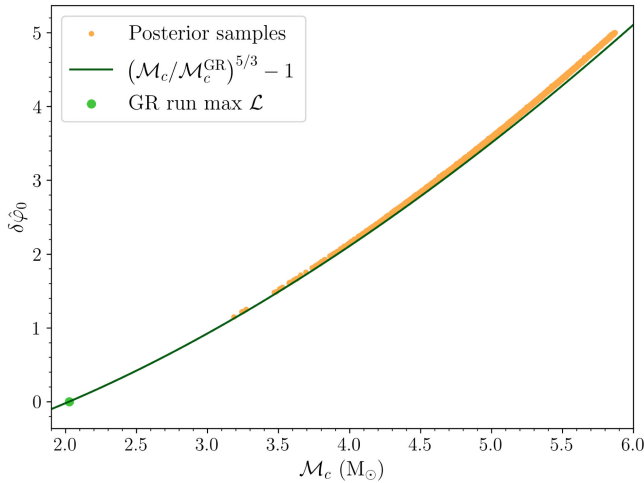


FIG. 3. 2D posterior between the chirp mass $\mathcal{M}_c$ and the 0PN deviation parameter $\delta\hat{\varphi}_0$ using the SEOBHM waveform model (orange). The green dot indicates the maximum likelihood values obtained with a GR run. The dark green line is the expected degeneracy between $\mathcal{M}_c$ and $\delta\hat{\varphi}_0$ based on the GR maximum likelihood chirp mass $\mathcal{M}_c^{\text{GR}}$ as per Eq. (6). We see that the posterior follows the expected correlation.

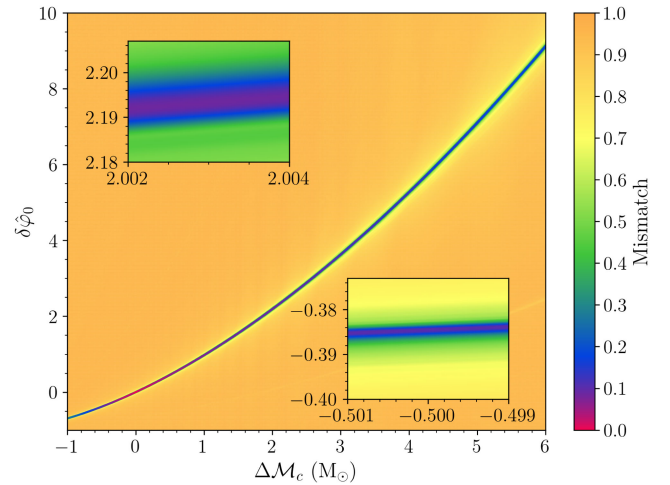


FIG. 4. Mismatch between the maximum likelihood waveform from a GR run using PhenomXPHM and non-GR waveforms as a function of chirp mass $\mathcal{M}_c$ and the 0PN deviation parameter $\delta\hat{\varphi}_0$. We clearly see the presence of the degeneracy from Eq. (6). The insets show a zoom-in on two regions around the degeneracy line. We notice that the low mismatch region becomes wider for larger chirp mass.

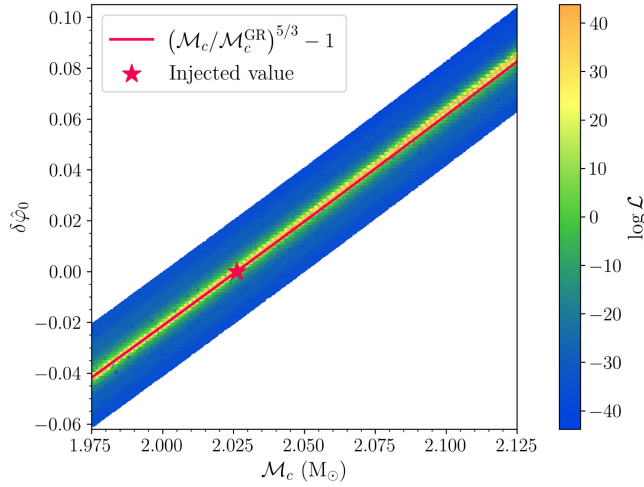


FIG. 5. Likelihood for non-GR waveforms as a function of chirp mass $\mathcal{M}_c$ and the OPN deviation parameter $\delta\hat{\varphi}_0$ when injecting the maximum likelihood waveform from a GR run using SEOBH. The injected values are indicated by the red star, and the red line is the expected degeneracy according to Eq. (6). We see that the high likelihood region is sharply peaked and follows the expected correlation.

expected degeneracy in the waveform according to Eq. (6). We see that the maximum likelihood region (yellow colors) indeed follows this line. There is no visible gradient along the line, which confirms that the waveform does not significantly change along the line (at least not within this chirp mass range).

When looking at Fig. 3, we also notice that the posterior is away from the GR value of $\delta\hat{\varphi}_0 = 0$. There could be multiple causes for this shift away from GR, which we will explain below. Most likely, a combination of them are playing a role for GW230529.

Because the likelihood barely changes along the line of the correlation, small changes in the likelihood due to, for example, noise could push the posterior to certain values. Normally when having multiple detectors observing a signal, the expectation is that such fluctuations due to noise cancel each other out between the detectors since the noise is mostly uncorrelated. For single-detector events, the correlation between detectors cannot be used to prevent extra testing GR parameters being fitted to noise features. Because this event is single detector, the noise might indeed be playing a part in the shift of the posterior toward higher chirp mass and $\delta\hat{\varphi}_0$.

It is known that noise effects such as nonstationarity, non-Gaussianity, and glitches can impact tests of GR [34,71,111,112] and even parameter estimation in GR [113–116]. More studies would be needed to see if biases like the shift away from GR at OPN can be produced by noise and to see if biases in tests of GR due to noise effects are indeed enhanced for single-detector events. There was no transient noise found around GW230529 [1], so we do not expect the noise to be majorly impacting our results.

The choice of prior could also significantly influence the posterior when there is such a degeneracy. For testing GR parameters that are strongly correlated with astrophysical parameters, this is a known issue, and it is therefore important to choose the correct priors [117]. We used a prior for the chirp mass that is uniform in the component masses, which translates to a preference for higher chirp-mass values. Since the waveform (and thus the likelihood) does not change much along this degeneracy, the posterior is likely shifted toward higher chirp mass and $\delta\hat{\varphi}_0$ because of this choice of prior.

Payne *et al.* [117] show that an astrophysically informed prior can help mitigate biases in tests of GR. To check the effect the choice of prior has on the OPN results for GW230529, we also performed an analysis with a prior that resembles their astrophysical prior for the primary mass and mass ratio [see Eqs. (5) and (6) in Ref. [117]]. Instead of inferring the powers α and β , we fix their values to $\alpha = 3$ and $\beta = 3$, which is consistent with their results for the OPN test. Note that we extend their prior to lower masses, where the power law is no longer representative of the astrophysical population, but does exemplify a simple analytic prior that prefers lower masses and more equal masses. We also performed an analysis where we use a prior that is uniform in the chirp mass and mass ratio. The results are shown in Fig. 6. We see that the low-mass-inclined power law prior (blue) gives results that are consistent with GR. The results also show more equal

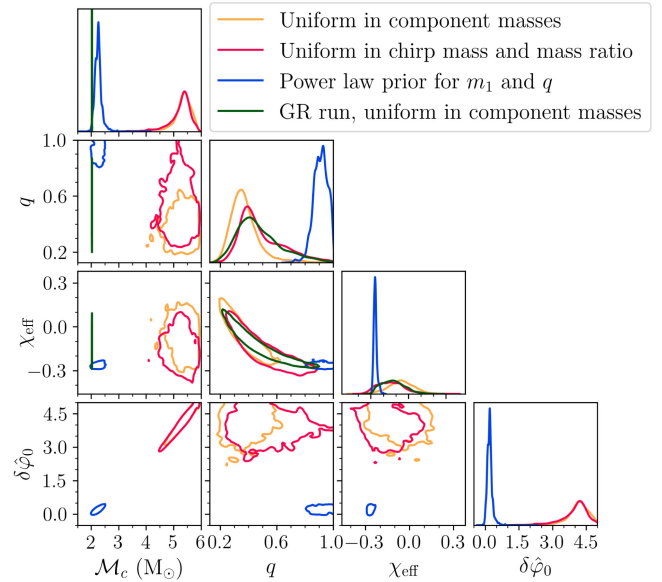


FIG. 6. Results for the OPN analysis using different priors for the masses obtained with SEOBH. We see that the low-mass-inclined power law prior (blue) gives results that are consistent with GR as opposed to the default uniform in component masses prior (orange) and the uniform in chirp mass and mass ratio prior (red). We also show the posteriors for a GR run using the default prior and SEOBH.

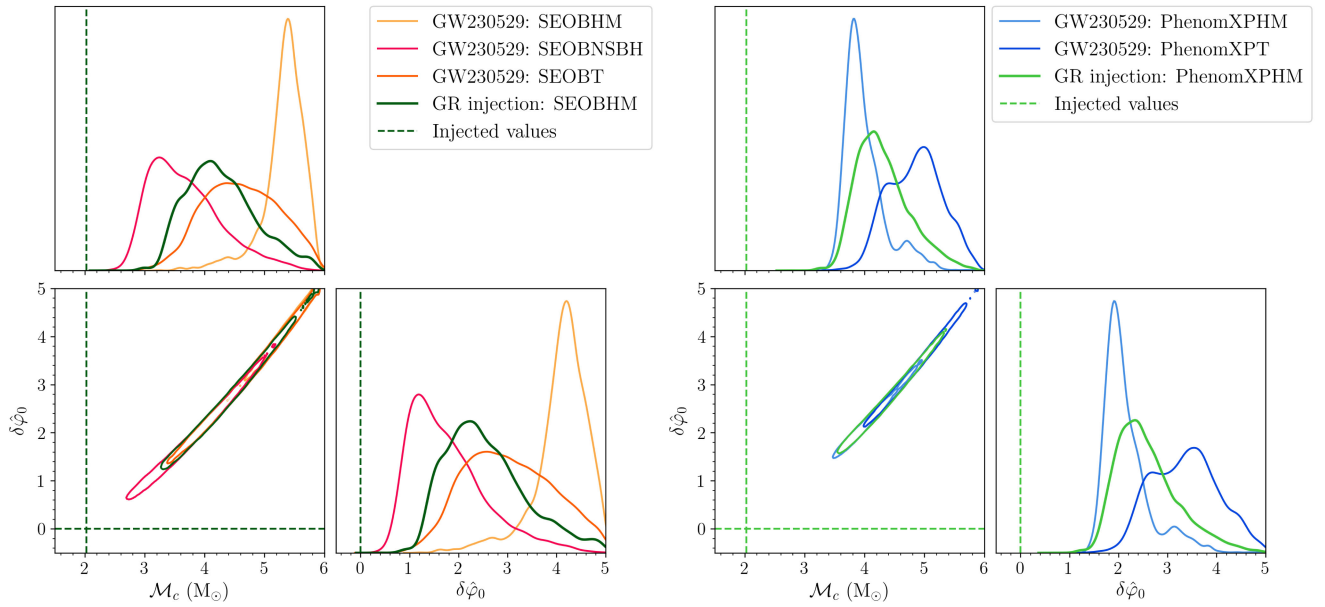


FIG. 7. Posteriors for the chirp mass $\mathcal{M}_c$ and OPN deviation parameter $\delta\hat{\varphi}_0$ for zero-noise injections of the maximum likelihood waveform from the corresponding GR run (green), compared to the posteriors obtained for GW230529 with the different waveform models. Left shows the posteriors obtained with FTI and right shows the results from TIGER. We see that the results from the GR injections are biased toward higher $\mathcal{M}_c$ and positive $\delta\hat{\varphi}_0$ and are similar to the GW230529 results.

masses, which is expected considering the prior prefers this, and a negative effective spin. This shift toward more equal masses and negative effective spin also happens for GR when using certain population-informed priors [1]. The results with the prior uniform in chirp mass and mass ratio (red) agree well with the results using our default prior that is uniform in the component masses (orange). This confirms that the apparent deviation from GR for $\delta\hat{\varphi}_0$ is prior dependent and most likely not a true deviation.

Another effect that could be causing the shift of the posterior towards higher chirp mass is some sampling issue. As can be seen in Fig. 5, the likelihood is sharply peaked around the correlation line (i.e., it is extremely narrow and elongated). It is often difficult for the sampler to find such narrow features in the likelihood. We see in Fig. 4 that the low mismatch region becomes wider for high chirp mass, which means that the high-likelihood region also becomes wider for high chirp mass. This makes it easier for the sampler to find that part of the high-likelihood region. This could be part of the cause for the shift of the posterior toward high chirp mass and $\delta\hat{\varphi}_0$. Changing the parametrization and priors used could make it easier for the sampler to find the high likelihood region in the lower chirp mass regime.

Lastly, GW230529 being detected by only the LIGO Livingston detector might also lead to biases. For single-detector events, the extrinsic parameters are not well measured. Because of correlations, biases on extrinsic parameters could propagate to intrinsic ones, negatively affecting how well they are measured compared to the case

of an event detected by multiple detectors. The chirp mass, luminosity distance, and inclination are correlated through the amplitude of the signal. In GR, the chirp mass is being well measured in the phase and the amplitude is not needed to constrain it. Since the chirp mass is free to change together with $\delta\hat{\varphi}_0$ in the phasing for low-mass systems, the amplitude would help constrain the chirp mass. But for single-detector events, the inclination and distance are not well measured (multiple detectors are needed for that). Indeed, we also see that the luminosity distance is wider and shifted toward larger values compared to the GR run. This increase in correlations for single-detector events could be contributing to the bias in chirp mass and $\delta\hat{\varphi}_0$. This could also explain why this shift away from GR is not observed as strongly for the BNS GW170817 [31] and the NSBH GW200115_042309 [34], which were both observed in three detectors.

As a check to see that this shift away from GR is not a true deviation from GR, we performed parameter estimation on zero-noise injections of the maximum likelihood waveform from the corresponding GR run using SEOBHM and PhenomXPHM. Figure 7 shows that the posteriors obtained for the chirp mass and $\delta\hat{\varphi}_0$ (green) are again shifted toward higher values away from GR and the true chirp mass (green dashed lines), but they are different from the ones obtained for GW230529 using the same waveform models (light orange and light blue). The differences are of the same order as those we get between individual waveform models. From this we can conclude that the noise is most likely not the dominant factor in the shift away from

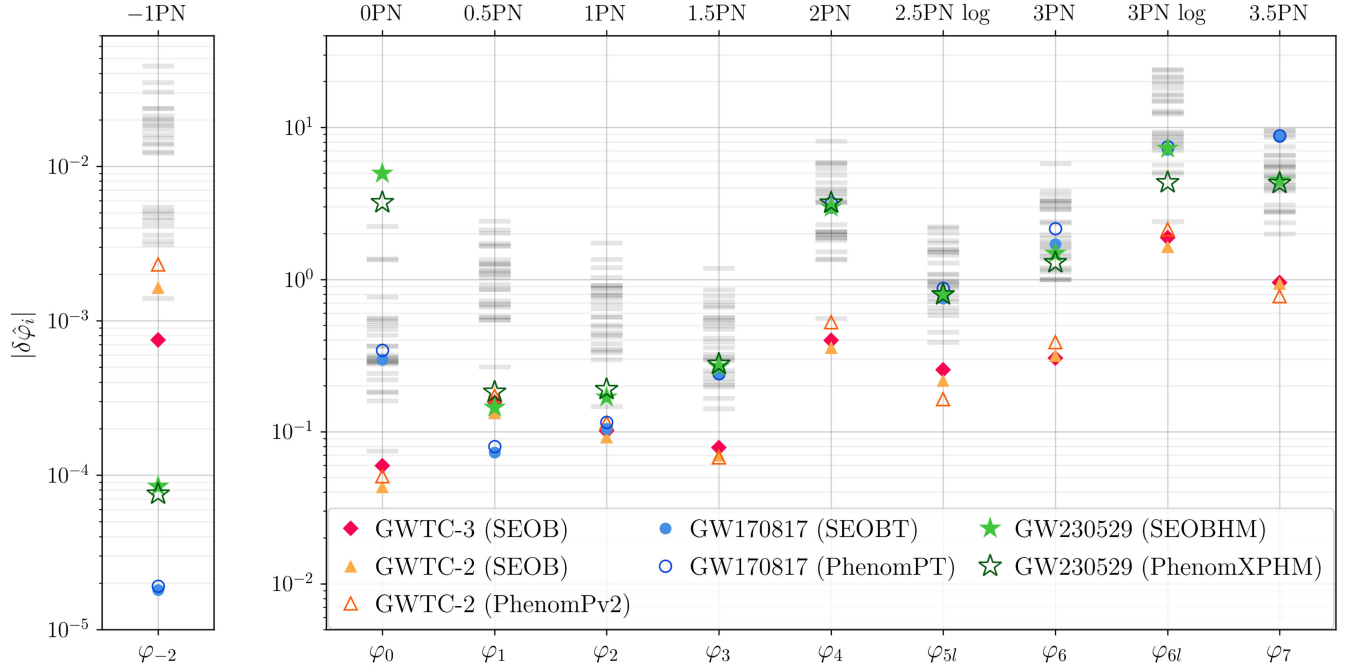


FIG. 8. The bounds for the different deviation parameters $\delta\hat{\phi}_n, \delta\hat{\phi}_{n(l)}$ obtained with GW230529 (green stars) compared to previously obtained bounds. We show the results obtained with the BBH waveforms as a proxy for an NSBH with realistic tides, as discussed in Sec. IV B. The orange triangles and red diamonds are bounds from GWTC-2 and GWTC-3, respectively, obtained by combining the posteriors for different events together (using SEOBnr4_ROM and IMRPhenomPv2). The blue circles are the bounds obtained with binary NS GW170817 (where they used SEOBnr4_ROM_NRTidalv2 and IMRPhenomPv2_NRTidalv2). The grey stripes are bounds from individual events in GWTC-2 and GWTC-3. We see that GW230529 gives exceptionally tight bounds for the -1PN , 0.5PN , and 1PN deviation parameters. (Adapted from Ref. [34]).

GR, but we cannot fully exclude that it might be playing a part.

We conclude that for GW230529 the results show a strong correlation between the 0PN deviation parameter and the chirp mass, and that this correlation matches the expected degeneracy that would be there for a GR signal with such low masses. Even though the posterior excludes the GR values, we do not consider the 0PN results for GW230529 as evidence against consistency with GR, because the shift away from zero is along the line of this degeneracy. We also observe a similar shift away from GR for a zero-noise injection with the maximum likelihood GR waveform, which further confirms that the 0PN results for GW230529 are most likely not a true deviation from GR.

D. Comparison with previous events

Figure 8 shows the 90% upper bounds on the deviation coefficients found for GW230529 (green stars), as well as the combined bounds from testing GR with GWTC-2 [33] (orange triangles) and GWTC-3 [34] (red diamonds). The plot also shows for comparison the bounds from GW170817 [31] (blue circles) and other bounds from individual events (grey stripes). Most noteworthy is the bound on the -1PN deviation of $|\delta\hat{\phi}_{-2}| \lesssim 8 \times 10^{-5}$. This is a factor ~ 17 better than previous bounds from the NSBH

GW200115_042309. It is also tighter than the combined constraint from GWTC-3 of $|\delta\hat{\phi}_{-2}| \lesssim 8 \times 10^{-4}$, which combined the results from all NSBHs and BBHs in GWTC-3 for which the LVK performed tests of GR. The only previous -1PN bound that is tighter than our constraints from GW230529, is the results for BNS GW170817, from which a bound of $|\delta\hat{\phi}_{-2}| \lesssim 2 \times 10^{-5}$ was obtained. The bounds found at 0.5PN and 1PN are also relatively tight, with $|\delta\hat{\phi}_1| \lesssim 0.2$ even being similar to the previous combined bounds.

We would expect the degeneracy at 0PN to also be there for other low-mass events. This is indeed the case for the BNS GW170817 [31] and the NSBH GW200115_042309 [34]. Those two events also show the shift away from GR toward positive $\delta\hat{\phi}_0$, however, not as strong as for GW230529 and they still have support for GR. This is most likely due to those events being observed in three detectors, so there are fewer degeneracies with extrinsic parameters.

V. TESTS OF ESGB GRAVITY

To illustrate the importance of the -1PN bounds from GW230529 for constraints on specific alternative theories of gravity, we have here a look at ESGB [41,42]. ESGB is a modified gravity theory where a scalar field is coupled to

the Gauss-Bonnet density. It is described by the action

$$S = \frac{1}{16\pi} \int d^4x \sqrt{-g} (R - 2(\partial\phi)^2 + \ell_{\text{GB}}^2 f(\phi) \mathcal{G}). \quad (7)$$

Here, g is the metric determinant, R is the Ricci scalar, ϕ is a scalar field with kinetic term $(\partial\phi)^2 = g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi$, and $\mathcal{G}$ is the Gauss-Bonnet invariant

$$\mathcal{G} = R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma} - 4R^{\mu\nu} R_{\mu\nu} + R^2, \quad (8)$$

where $R^{\mu\nu\rho\sigma}$ is the Riemann tensor and $R^{\mu\nu}$ the Ricci tensor. The integral over a four-dimensional spacetime $\int d^4x \sqrt{-g} \mathcal{G}$ is a boundary term [118]. The function $f(\phi)$, thus defined modulo a constant, specifies the theory. As for ℓ_{GB} , it is the Gauss-Bonnet coupling strength with dimension of length. We assume that matter fields are minimally coupled to $g_{\mu\nu}$.

In this paper, we consider the class of theories such that $f'(0) \neq 0$. Without loss of generality, the function $f(\phi)$ can be expanded as follows:

$$f(\phi) = 2\phi + \mathcal{O}(\phi^2). \quad (9)$$

At leading order in ϕ , the action (7) reduces to that of shift-symmetric ESGB theories [89], which is invariant under the shift $\phi \rightarrow \phi + \Delta\phi$, where $\Delta\phi$ is a constant. This is equivalent to the small- ϕ approximation of Einstein-dilaton-Gauss-Bonnet for which $f(\phi) = e^{\gamma\phi}$ with a constant γ (the value of γ can be absorbed into the coupling constant). Here, we use the convention of Refs. [89,93,94]. Another convention denotes the Gauss-Bonnet coupling by α_{GB} [47], which is related to our ℓ_{GB} as $\alpha_{\text{GB}}^2 = \ell_{\text{GB}}^4 / (16\pi)$.

It is possible to find the ESGB corrections to the GW phase during inspiral using the PN formalism. The leading-order correction to the frequency domain phase in ESGB appears at -1PN due to scalar dipole radiation. In the class of ESGB theories (9) and at $\mathcal{O}(\ell_{\text{GB}}^4)$, it reads [47]

$$\delta\hat{\varphi}_{-2} = -5\ell_{\text{GB}}^4 \frac{(m_1^2 s_2 - m_2^2 s_1)^2}{168m_1^4 m_2^4}, \quad (10)$$

where $m_{1,2}$ are the masses of the compact objects, while $s_{1,2}$ were introduced in Refs. [46,47]. Here, s is defined as the scalar monopole of a body at $\mathcal{O}(\ell_{\text{GB}}^2)$, divided by that of a nonspinning BH with the same mass [89]. The higher-order corrections have been fully computed up to 1PN (2PN relative order) for nonspinning binaries in the past [87–91,93,94]. These higher-order corrections are given in Appendix C. We also provide the full 1.5PN correction in Eq. (C1e), which is new to this paper.

In the class of ESGB theories (9), NSs cannot carry scalar monopoles at $\mathcal{O}(\ell_{\text{GB}}^2)$ so $s = 0$ [47,87,119], and for nonspinning BHs, $s = 1$ [89]. So for a binary consisting of a NS and a nonspinning BH, Eq. (10) can be simplified to

$$\delta\hat{\varphi}_{-2} = \frac{-5\ell_{\text{GB}}^4}{168m_1^4}, \quad (11)$$

where m_1 is the mass of the BH. When computing the ESGB corrections for GW230529, we assume the secondary to be a NS independently of the waveform model used (remember that we use the BBH results as a proxy for the NSBH results with realistic tides).

A. Constraints on ESGB by mapping agnostic results

It is possible to get approximate constraints on the ESGB coupling from theory-agnostic tests. This is done by mapping the -1PN agnostic constraints to the ESGB coupling using the leading order correction. Inverting Eq. (11) gives

$$\ell_{\text{GB}} = \left(\frac{-168m_1^4 \delta\hat{\varphi}_{-2}}{5} \right)^{1/4}. \quad (12)$$

Note that we need to restrict ourselves to negative values of $\delta\hat{\varphi}_{-2}$. When mapping posteriors, we also need to reweight the samples to account for differing priors. To map from a prior uniform in $\delta\hat{\varphi}_{-2}$ to a prior uniform in ℓ_{GB} , the new sample weights w become

$$w = \frac{\partial \ell_{\text{GB}}}{\partial (\delta\hat{\varphi}_{-2})} = \frac{\ell_{\text{GB}}}{4\delta\hat{\varphi}_{-2}}. \quad (13)$$

The posteriors obtained by mapping the agnostic posteriors for $\delta\hat{\varphi}_{-2}$ to ℓ_{GB} are shown in the top panel in Fig. 9. We see that the posteriors for results obtained with SEOBHM (orange), SEOBNSBH (red), and PhenomXPHM (blue) do not significantly differ. The gap between $\ell_{\text{GB}} \sim 0.1$ and 0 is not due to lack of support for GR. It is rather an effect of the sampling and reweighting due to the ℓ_{GB}^4 dependence of $\delta\hat{\varphi}_{-2}$. Values of $\ell_{\text{GB}} \lesssim 0.1$ in this case require $\delta\varphi_{-2} \lesssim 10^{-8}$, which is four orders of magnitude smaller than the bound found. The limited number of samples is therefore causing a gap for small values. This is also the reason why the lower end of the posterior looks so ragged; there are only a few samples that have a relatively large weight. We would need many more samples to close this gap and make the posterior smooth for small ℓ_{GB} .

The 90% upper bound on the ESGB coupling (indicated with dashed lines in Fig. 9) is $\ell_{\text{GB}} \lesssim 0.67M_\odot$ or $\sqrt{\alpha_{\text{GB}}} \lesssim 0.37 \text{ km}$ (for SEOBHM and SEOBNSBH, for PhenomXPHM the bound is $\ell_{\text{GB}} \lesssim 0.72M_\odot$ or $\sqrt{\alpha_{\text{GB}}} \lesssim 0.40 \text{ km}$). This bound can be improved by including higher PN order effects in ESGB. In the next section, we include these effects in a theory-specific inspiral test for ESGB.

B. Theory-specific test for ESGB

Due to the flexibility of the FTI framework, we can use it not only to do a theory-agnostic inspiral test, but also for theory-specific inspiral tests [72]. This is done by explicitly

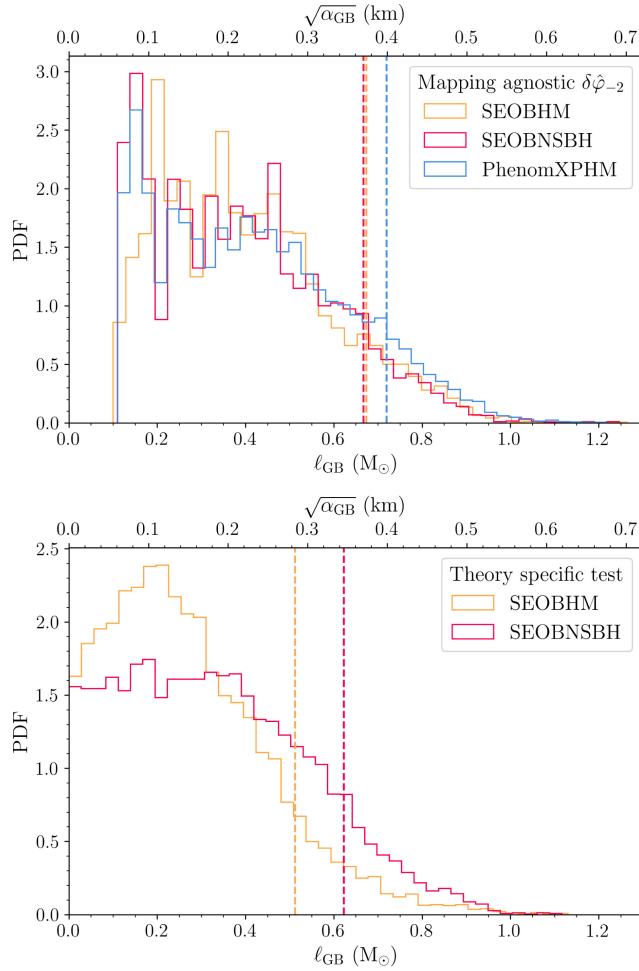


FIG. 9. Top: posteriors for the ESGB coupling ℓ_{GB} obtained by reweighting the $\delta\hat{\varphi}_{-2}$ posterior. Bottom: posteriors for the Einstein-scalar-Gauss Bonnet coupling ℓ_{GB} from the theory-specific test. The top axes show a slightly different definition of the ESGB coupling $\sqrt{\alpha_{\text{GB}}} = \ell_{\text{GB}}/(2\pi^{1/4})$ that is also commonly used.

using the corrections to the frequency-domain phase from a specific theory. This means that we can not only use the leading-order corrections, but also the higher-order corrections. For ESGB, we can thus use all corrections up to 1.5PN, as given in Appendix C. We only add corrections to the dominant (2, 2)-mode. We can then do Bayesian inference sampling over the ESGB coupling ℓ_{GB} , where we use a prior uniform in ℓ_{GB} . A posteriori, we check that the requirement $(\ell_{\text{GB}}/m_1)^4 \ll 1$ holds, which comes from the small coupling approximation that is made in the derivation of the coefficients.

The results from the theory-specific ESGB test are shown in the bottom panel of Fig. 9. Note that for this case the posterior looks much smoother for small ℓ_{GB} than was the case for the mapping and there is no gap close to zero. That is because we are now sampling directly on ℓ_{GB} so there is no problems with undersampling low values of the coupling.

The constraint on the ESGB coupling found using SEOBHM (orange) is $\ell_{\text{GB}} \lesssim 0.51M_{\odot}$ or $\sqrt{\alpha_{\text{GB}}} \lesssim 0.28$ km, and using SEOBNSBH gives slightly worse constraints of $\ell_{\text{GB}} \lesssim 0.62M_{\odot}$ or $\sqrt{\alpha_{\text{GB}}} \lesssim 0.35$ km (red). We notice that the bounds obtained with the theory-specific test are better than the bounds found from mapping the agnostic -1PN results. We note that in this work, the higher-order corrections to the phase for testing ESGB theory are incorporated only in the FTI framework and not in the waveforms for TIGER. This is because the implementation with FTI is more flexible, and we do not expect the results to change, as the -1PN mapping-based results are consistent.

These constraints on the ESGB coupling are better than any previously obtained constraints using GWs. From NSBHs, the best constraint so far was $\sqrt{\alpha_{\text{GB}}} \lesssim 1.33$ km for GW200115_042309 [47]. The best overall constraint was obtained with GW190814 and was $\sqrt{\alpha_{\text{GB}}} \lesssim 0.37$ km (assuming it is a BBH; its secondary lies in the lower mass gap) [47]. These bounds are also better than constraints obtained using low-mass x-ray binaries of $\sqrt{\alpha_{\text{GB}}} \lesssim 1.9$ km [120] and NS EOS of $\sqrt{\alpha_{\text{GB}}} \lesssim 1.29$ km [121]. Laboratory tests for ESGB in the weak field limit provide bounds that are at least 12 orders of magnitude weaker than strong-field tests [122].

VI. CONCLUSION

In this work we have performed inspiral tests of GR with GW230529, the merger of two compact objects with masses $3.6^{+0.8}_{-1.1}M_{\odot}$ and $1.4^{+0.6}_{-0.2}M_{\odot}$. The tests used are parametrized inspiral tests that modify the phase in the frequency domain by allowing generic modifications of the PN coefficients.

We used two different versions of the parametrized inspiral test that employ different waveform approximants: FTI [31,72] using the SEOBNSv4 waveform family [77–81] and TIGER [74,75] using the IMRPhenomX waveform family [82–84]. We compared the results of both tests. The various BBH waveform approximants gave consistent results, all of which are also consistent with GR, except for the 0PN deviation coefficient. The results are shown in Fig. 1.

At 0PN, there is a degeneracy between the chirp mass and the deviation parameter. This means that the chirp mass and the 0PN deviation parameter are free to move along the correlation given by Eq. (6) without significantly changing the waveform. For higher mass systems, this degeneracy is broken by higher PN contributions in the late inspiral and by the merger-ringdown. For GW230529, however, the merger-ringdown is outside the analyzed frequency band, so there is a strong correlation between $\mathcal{M}_c$ and $\delta\hat{\varphi}_0$ in the results, as can be seen in Fig. 3. We notice that the posterior is shifted away from GR and toward higher chirp mass. Since this is also observed in a zero-noise GR injection, it is very likely a false deviation from GR. The cause of this

shift is probably a combination of the choice of priors, sampling issues due to a sharply peaked likelihood, the event being single detector, and noise features.

Since the source of GW230529 is most likely an NSBH, we also performed analyses with waveforms that include tidal effects. The deviation-parameter posteriors obtained with these waveforms are broader and shifted away from GR compared to their BBH counterparts. This is due to correlations between the tidal deformability and the deviation parameters. For GW230529, the tidal deformabilities are not well constrained and even allow unrealistically high values [1], leading to the broad and shifted posteriors of the $\delta\hat{\varphi}_i$. By constraining the tides to more realistic values, the shift and broadening (mostly) disappear (see Fig. 2). We therefore use the results from the BBH waveform models when quoting bounds below.

The bounds obtained on the low-PN deviation parameters (except for 0PN) are some of the best to date. The bound obtained on dipole radiation (-1 PN deviation) for GW230529 of $|\delta\hat{\varphi}_{-2}| \lesssim 8 \times 10^{-5}$ is an order of magnitude tighter than the combined bounds from GWTC-2 [33] and GWTC-3 [34]. The only tighter bound obtained with GWs is for the binary NS GW170817 [31]. The bounds obtained at 0.5 PN and 1 PN are also relatively tight compared to previous constraints. Figure 8 compares the bounds for GW230529 with previous constraints from GWs.

The bounds on -1 PN deviations allow us to constrain scalar-tensor theories that have their leading-order modification at -1 PN. In particular, we have mapped the posteriors obtained for $\delta\hat{\varphi}_{-2}$ to the ESGB coupling ℓ_{GB} using Eq. (10). This gives us an upper bound on the ESGB coupling of $\ell_{\text{GB}} \lesssim 0.67M_{\odot}$. We have also performed a theory-specific test for ESGB gravity by implementing all corrections up to 1.5 PN in the FTI framework. This allows us to sample directly over the ESGB coupling ℓ_{GB} and obtain a slightly better bound of $\ell_{\text{GB}} \lesssim 0.51M_{\odot}$. Figure 9 shows the posteriors obtained from both the mapping and the theory-specific test. These constraints on the ESGB coupling are the best obtained from GW signals to date.

In this work, we found no evidence for deviations from GR for GW230529. We obtained particularly tight constraints on deviations from GR at low-PN orders (with the exception of 0PN), demonstrating the importance of signals similar to GW230529 for tests of GR. We also discussed some challenges that arise when analyzing signals from low-mass systems such as GW230529, namely tidal effects and the degeneracy between the chirp mass and the 0PN deviation parameter. Future, more detailed studies of these effects will hopefully lead to ways to better account for, and perhaps even mitigate, their systematic biases.

When completing this paper, we became aware of an independent study [123] which obtained constraints on

Einstein-dilaton-Gauss-Bonnet gravity using GW230529. They obtain a bound of $\sqrt{\alpha_{\text{GB}}} \lesssim 0.260$ km when including higher-order corrections and using IMRPhenomXPHM, which is similar to our bound of $\sqrt{\alpha_{\text{GB}}} \lesssim 0.28$ km using the theory-specific test for ESGB with SEOBNRv4HM_ROM.

ACKNOWLEDGMENTS

The authors would like to thank N. V. Krishnendu for reviewing the manuscript and providing useful feedback. We thank Aditya Vijaykumar for his work on developing and maintaining the Bilby TGR package. S. R. was supported by the research program of the Netherlands Organization for Scientific Research (NWO). The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459. The authors are also grateful for the computational resources provided by Cardiff University supported by STFC grant ST/I006285/1, as well as computational resources from the Max Planck Institute for Gravitational Physics in Potsdam. This research has made use of data or software obtained from the Gravitational Wave Open Science Center [124], a service of the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation, as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

DATA AVAILABILITY

The data that support the findings of this article are openly available [125].

APPENDIX A: PRIORS USED

The priors used in the analyses of GW230529 are given in the tables below. The names of the priors refer to the

TABLE III. Priors used for the deviation parameter and the chirp mass for the different runs. The priors on the deviation parameter are uniform between the indicated range. The priors on the chirp mass are uniform in the component masses and between the indicated range.

Deviation parameter	Prior	Chirp mass prior
φ_{-2}	Uniform($-0.1, 0.1$)	UniformInComponentsChirpMass(1.9, 2.2)
φ_0	Uniform($-5, 5$) ^a	UniformInComponentsChirpMass(1.8, 6.0)
φ_1	Uniform($-2, 2$)	UniformInComponentsChirpMass(1.9, 2.2)
φ_2	Uniform($-2, 2$)	UniformInComponentsChirpMass(2.00, 2.05)
φ_3	Uniform($-2, 2$)	UniformInComponentsChirpMass(2.00, 2.05)
φ_4	Uniform($-20, 20$)	UniformInComponentsChirpMass(2.00, 2.05)
φ_{5l}	Uniform($-20, 20$)	UniformInComponentsChirpMass(2.00, 2.05)
φ_6	Uniform($-20, 20$)	UniformInComponentsChirpMass(2.00, 2.05)
φ_{6l}	Uniform($-30, 30$)	UniformInComponentsChirpMass(2.00, 2.05)
φ_7	Uniform($-30, 30$)	UniformInComponentsChirpMass(2.00, 2.05)

^aThe prior was limited to Uniform($-0.5, 5$) for PhenomXPT due to problems with waveform wraparound like described in Appendix B.

TABLE IV. Priors on the spin parameters used for the different waveform approximants. The ALIGNEDSPIN prior gives the prior distribution of the aligned spin component based on the generic spin priors.

Waveform approximant	Parameter	Prior
SEOBHM, SEOB, PhenomXHM	χ_1	AlignedSpin(0, 0.99)
	χ_2	AlignedSpin(0, 0.99)
SEOBNSBH	χ_1	AlignedSpin(0, 0.9)
	χ_2	AlignedSpin(0, 0.05)
PhenomXPHM	a_1	Uniform(0, 0.99)
	a_2	Uniform(0, 0.99)
	θ_1	Sine(0, π)
	θ_2	Sine(0, π)
	ϕ_{12}	Uniform(0, 2π)
	ϕ_{jl}	Uniform(0, 2π)
PhenomXP, PhenomXPT	a_1	Uniform(0, 0.99)
	a_2	Uniform(0, 0.05)
	θ_1	Sine(0, π)
	θ_2	Sine(0, π)
	ϕ_{12}	Uniform(0, 2π)
	ϕ_{jl}	Uniform(0, 2π)

names in the Bilby package [102]. The numbers indicate the range used. The priors are chosen to be agnostic, which means most of them are uniform in the parameter sampled or uniform in volume (e.g., for angles). The priors used for the deviation parameters and chirp masses are listed in Table III. The increase in width of the chirp-mass priors for low-PN deviations is because the chirp-mass posterior becomes broader due to correlations with the deviation parameters at low-PN order. The priors used for the spins are listed in Table IV and differ between different waveform approximants. This is due to differing spin descriptions (aligned spins versus precessing spins) and due to

TABLE V. Priors used for the other GR parameters for all runs and waveforms. The prior on the mass ratio is such that it is uniform in the component masses. The prior on the geocentric time is a window of 0.2 seconds around the merger time. The tidal deformabilities are set to zero for BBH waveforms, and $\Lambda_1 = 0$ for SEOBNSBH.

Parameter	Prior
q	UniformInComponentsMassRatio(0.125, 1)
d_L	UniformSourceFrame(1, 1000)
δ	Cosine($-\pi/2, \pi/2$)
α	Uniform(0, 2π)
θ_{jn}	Sine(0, π)
ψ	Uniform(0, π)
ϕ_{ref}	Uniform(0, 2π)
t_c	Uniform(1369419318.6460938, 1369419318.8460937)
$\Lambda_{1,2}$	Uniform(0, 5000)

limitations in the validity of the models. Lastly, all other priors are listed in Table V.

APPENDIX B: FALSE VIOLATIONS OF GR WHEN USING TOO WIDE PRIORS

Large values of the deviation parameters can lead to significant changes in the length of the waveform. For more extreme cases, this can lead to the waveform wrapping around due to the choice of segment length of the data analyzed. In the case of GW230529, this problem can arise for, e.g., the 0.5PN deviation coefficient, where, for extreme negative values of $\delta\hat{\varphi}_1$, the post-trigger duration becomes more than the 2 s used in the analysis. When using wider priors on the deviation parameters, this can lead to false deviations of GR. For example, the posteriors for the

0.5PN deviation parameter would be peaking around $\delta\hat{\varphi}_1 \sim -9$. Normally, these unphysical waveforms that lead to wraparound would be penalized enough by having a low likelihood so as not to get any support in the posterior. Because GW230529 is a single-detector event, we suspect it instead picked up on some noise feature or artifact at the edge of the data segment analyzed, which lead to a high likelihood. When increasing the post-trigger duration to include a longer segment of data after the signal ends, this false violation of GR disappears. For the results presented in this work, we instead restricted the priors on the deviation parameters to avoid the unphysical waveforms that lead to wraparound.

APPENDIX C: HIGHER-ORDER ESGB CORRECTIONS IN THE GW PHASE

The corrections to the frequency-domain phase in ESGB can be computed using the PN formalism. The leading order corrections appear at -1 PN (compared to GR), and the phase corrections have been fully computed up to 1PN (2PN relative order) for nonspinning binaries [47,88–91,93]. We calculated as well the complete 1.5PN corrections (C1e), which are new to this paper, using the recent fluxes at relative 2.5PN order of Ref. [92]. In the formalism described in Sec. II, the corrections are given by

$$\delta\hat{\varphi}_{-2} = -5\ell_{\text{GB}}^4 \frac{(m_1^2 s_2 - m_2^2 s_1)^2}{168m_1^4 m_2^4}, \quad (\text{C1a})$$

$$\delta\hat{\varphi}_0 = -5\ell_{\text{GB}}^4 \frac{659m_1^4 s_2^2 + 1370m_1^2 m_2^2 s_1 s_2 + 659m_2^4 s_1^2 + 728\eta(m_1^2 s_2 - m_2^2 s_1)^2}{16128m_1^4 m_2^4}, \quad (\text{C1b})$$

$$\delta\hat{\varphi}_1 = 25\pi\ell_{\text{GB}}^4 \frac{(m_1^2 s_2 - m_2^2 s_1)^2}{56m_1^4 m_2^4}, \quad (\text{C1c})$$

$$\delta\hat{\varphi}_2 = \ell_{\text{GB}}^4 \left\{ \frac{5m_1^4 s_2^2 [-13792267 + 5588352\delta - 17640\eta(743 + 594\eta)]}{290304m_1^4 m_2^4 (743 + 924\eta)} - \frac{5m_2^4 s_1^2 [13792267 + 5588352\delta + 17640\eta(743 + 594\eta)]}{290304m_1^4 m_2^4 (743 + 924\eta)} + \frac{2m_1^2 m_2^2 s_1 s_2 [56018615 + 3528\eta(12239 + 14850\eta)]}{290304m_1^4 m_2^4 (743 + 924\eta)} \right\}, \quad (\text{C1d})$$

$$\delta\hat{\varphi}_3 = \ell_{\text{GB}}^4 \frac{m_1^4 s_2^2 (-14363 + 1792\delta - 4564\eta) + 2m_1^2 m_2^2 s_1 s_2 (-3557 + 4564\eta) - m_2^4 s_1^2 (14363 + 1792\delta + 4564\eta)}{43008m_1^4 m_2^4}. \quad (\text{C1e})$$

Here, ℓ_{GB} is the ESGB coupling, m_1, m_2 are the masses of the two objects that also appear in the combinations $\eta = m_1 m_2 / (m_1 + m_2)^2$ and $\delta = (m_1 - m_2) / (m_1 + m_2)$. For NSs, $s = 0$ [47] and for nonspinning BHs, $s = 1$ [89,93].

The ESGB corrections to the phase above are complete up to 1.5PN for nonspinning binaries. By contrast, Ref. [123] uses the corrections presented in Ref. [47], which are complete up to 1PN only. Indeed, at 1.5PN,

the corrections of Ref. [47] depend on the free parameter f_3^{GB} to represent the contributions from the fluxes at relative 2.5PN order, which were unknown at the time. Similarly, the 2PN corrections of Ref. [47] depend on f_4^{GB} to represent contributions from the binding energy at 3PN, and from the (still unknown) fluxes at relative 3PN order. Note that in practice, both f_3^{GB} and f_4^{GB} are set to zero in Ref. [47].

[1] A. G. Abac *et al.* (LIGO Scientific, KAGRA, and Virgo Collaborations), Observation of gravitational waves from the coalescence of a $2.5 - 4.5M_\odot$ compact object and a neutron star, *Astrophys. J. Lett.* **970**, L34 (2024).

[2] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, and Virgo Collaborations), Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, *Living Rev. Relativity* **21**, 3 (2018).

- [3] J. Aasi *et al.* (LIGO Scientific Collaboration), Advanced LIGO, *Classical Quantum Gravity* **32**, 074001 (2015).
- [4] M. Fishbach, R. Essick, and D. E. Holz, Does matter matter? Using the mass distribution to distinguish neutron stars and black holes, *Astrophys. J. Lett.* **899**, L8 (2020).
- [5] A. M. Farah, M. Fishbach, R. Essick, D. E. Holz, and S. Galaudage, Bridging the gap: Categorizing gravitational-wave events at the transition between neutron stars and black holes, *Astrophys. J.* **931**, 108 (2022).
- [6] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW190814: Gravitational waves from the coalescence of a 23 solar mass black hole with a 2.6 solar mass compact object, *Astrophys. J. Lett.* **896**, L44 (2020).
- [7] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Population of merging compact binaries inferred using gravitational waves through GWTC-3, *Phys. Rev. X* **13**, 011048 (2023).
- [8] 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).
- [9] T. A. Thompson *et al.*, A noninteracting low-mass black hole–giant star binary system, *Science* **366**, 637 (2019).
- [10] T. Jayasinghe *et al.*, A unicorn in monoceros: The $3M_{\odot}$ dark companion to the bright, nearby red giant V723 Mon is a non-interacting, mass-gap black hole candidate, *Mon. Not. R. Astron. Soc.* **504**, 2577 (2021).
- [11] E. D. Barr *et al.*, A pulsar in a binary with a compact object in the mass gap between neutron stars and black holes, *Science* **383**, 275 (2024).
- [12] 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).
- [13] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW190425: Observation of a compact binary coalescence with total mass $\sim 3.4M_{\odot}$, *Astrophys. J. Lett.* **892**, L3 (2020).
- [14] V. Kalogera and G. Baym, The maximum mass of a neutron star, *Astrophys. J. Lett.* **470**, L61 (1996).
- [15] L. Rezzolla, E. R. Most, and L. R. Weih, Using gravitational-wave observations and quasi-universal relations to constrain the maximum mass of neutron stars, *Astrophys. J. Lett.* **852**, L25 (2018).
- [16] C. D. Bailyn, R. K. Jain, P. Coppi, and J. A. Orosz, The mass distribution of stellar black holes, *Astrophys. J.* **499**, 367 (1998).
- [17] W. M. Farr, N. Sravan, A. Cantrell, L. Kreidberg, C. D. Bailyn, I. Mandel, and V. Kalogera, The mass distribution of stellar-mass black holes, *Astrophys. J.* **741**, 103 (2011).
- [18] F. Ozel, D. Psaltis, R. Narayan, and J. E. McClintock, The black hole mass distribution in the galaxy, *Astrophys. J.* **725**, 1918 (2010).
- [19] P. O. Mazur and E. Mottola, Gravitational vacuum condensate stars, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9545 (2004).
- [20] S. L. Liebling and C. Palenzuela, Dynamical boson stars, *Living Rev. Relativity* **26**, 1 (2023).
- [21] O. Lunin and S. D. Mathur, AdS/CFT duality and the black hole information paradox, *Nucl. Phys.* **B623**, 342 (2002).
- [22] V. Cardoso, S. Hopper, C. F. B. Macedo, C. Palenzuela, and P. Pani, Gravitational-wave signatures of exotic compact objects and of quantum corrections at the horizon scale, *Phys. Rev. D* **94**, 084031 (2016).
- [23] B. Carr, F. Kuhnel, and M. Sandstad, Primordial black holes as dark matter, *Phys. Rev. D* **94**, 083504 (2016).
- [24] V. K. Oikonomou, Phenomenology of axionic static neutron stars with masses in the mass-gap region, *Classical Quantum Gravity* **41**, 085008 (2024).
- [25] C. M. Will, The confrontation between general relativity and experiment, *Living Rev. Relativity* **17**, 4 (2014).
- [26] N. Wex, Testing relativistic gravity with radio pulsars, [arXiv:1402.5594](https://arxiv.org/abs/1402.5594).
- [27] L. Shao, N. Sennett, A. Buonanno, M. Kramer, and N. Wex, Constraining nonperturbative strong-field effects in scalar-tensor gravity by combining pulsar timing and laser-interferometer gravitational-wave detectors, *Phys. Rev. X* **7**, 041025 (2017).
- [28] P. G. Ferreira, Cosmological tests of gravity, *Annu. Rev. Astron. Astrophys.* **57**, 335 (2019).
- [29] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Tests of general relativity with GW150914, *Phys. Rev. Lett.* **116**, 221101 (2016); *Phys. Rev. Lett.* **121**, 129902(E) (2018).
- [30] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Binary black hole mergers in the first Advanced LIGO observing run, *Phys. Rev. X* **6**, 041015 (2016); *Phys. Rev. X* **8**, 039903(E) (2018).
- [31] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Tests of general relativity with GW170817, *Phys. Rev. Lett.* **123**, 011102 (2019).
- [32] 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).
- [33] 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); *Phys. Rev. D* **112**, 049902(E) (2025).
- [34] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Tests of general relativity with GWTC-3, *Phys. Rev. D* **112**, 084080 (2025).
- [35] R. Genzel, F. Eisenhauer, and S. Gillessen, Experimental studies of black holes: Status and future prospects, *Astron. Astrophys. Rev.* **32**, 3 (2024).
- [36] R. C. Bernardo and K.-W. Ng, Beyond the Hellings–Downs curve: Non-Einsteinian gravitational waves in pulsar timing array correlations, *Astron. Astrophys.* **691**, A126 (2024).
- [37] E. Cannizzaro, G. Franciolini, and P. Pani, Novel tests of gravity using nano-Hertz stochastic gravitational-wave background signals, *J. Cosmol. Astropart. Phys.* **04** (2023) 056.
- [38] C. Brans and R. H. Dicke, Mach’s principle and a relativistic theory of gravitation, *Phys. Rev.* **124**, 925 (1961).
- [39] G. W. Horndeski, Second-order scalar-tensor field equations in a four-dimensional space, *Int. J. Theor. Phys.* **10**, 363 (1974).

- [40] T. Jacobson and D. Mattingly, Einstein-Aether waves, *Phys. Rev. D* **70**, 024003 (2004).
- [41] S. Nojiri, S. D. Odintsov, and M. Sasaki, Gauss-Bonnet dark energy, *Phys. Rev. D* **71**, 123509 (2005).
- [42] A. De Felice and S. Tsujikawa, $f(R)$ theories, *Living Rev. Relativity* **13**, 3 (2010).
- [43] R. Jackiw and S. Y. Pi, Chern-Simons modification of general relativity, *Phys. Rev. D* **68**, 104012 (2003).
- [44] S. Alexander and N. Yunes, Chern-Simons modified general relativity, *Phys. Rep.* **480**, 1 (2009).
- [45] S. Endlich, V. Gorbenko, J. Huang, and L. Senatore, An effective formalism for testing extensions to general relativity with gravitational waves, *J. High Energy Phys.* **09** (2017) 122.
- [46] 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).
- [47] Z. Lyu, N. Jiang, and K. Yagi, Constraints on Einstein-dilation-Gauss-Bonnet gravity from black hole-neutron star gravitational wave events, *Phys. Rev. D* **105**, 064001 (2022); **106**, 069901(E) (2022).
- [48] H. Takeda, S. Tsujikawa, and A. Nishizawa, Gravitational-wave constraints on scalar-tensor gravity from a neutron star and black-hole binary GW200115, *Phys. Rev. D* **109**, 104072 (2024).
- [49] H. O. Silva, A. Ghosh, and A. Buonanno, Black-hole ringdown as a probe of higher-curvature gravity theories, *Phys. Rev. D* **107**, 044030 (2023).
- [50] W. E. East and J. L. Ripley, Evolution of Einstein-scalar-Gauss-Bonnet gravity using a modified harmonic formulation, *Phys. Rev. D* **103**, 044040 (2021).
- [51] M. Bezares, R. Aguilera-Miret, L. ter Haar, M. Crisostomi, C. Palenzuela, and E. Barausse, No evidence of kinetic screening in simulations of merging binary neutron stars beyond general relativity, *Phys. Rev. Lett.* **128**, 091103 (2022).
- [52] P. Figueras and T. França, Black hole binaries in cubic Horndeski theories, *Phys. Rev. D* **105**, 124004 (2022).
- [53] L. Aresté Saló, K. Clough, and P. Figueras, Well-posedness of the four-derivative scalar-tensor theory of gravity in singularity avoiding coordinates, *Phys. Rev. Lett.* **129**, 261104 (2022).
- [54] L. Aresté Saló, K. Clough, and P. Figueras, Puncture gauge formulation for Einstein-Gauss-Bonnet gravity and four-derivative scalar-tensor theories in $d + 1$ spacetime dimensions, *Phys. Rev. D* **108**, 084018 (2023).
- [55] L. Aresté Saló, S. E. Brady, K. Clough, D. Doneva, T. Evstafyeva, P. Figueras, T. França, L. Rossi, and S. Yao, GRFolres: A code for modified gravity simulations in strong gravity, *J. Open Source Software* **9**, 6369 (2024).
- [56] M. Corman, J. L. Ripley, and W. E. East, Nonlinear studies of binary black hole mergers in Einstein-scalar-Gauss-Bonnet gravity, *Phys. Rev. D* **107**, 024014 (2023).
- [57] M. Corman and W. E. East, Black hole-neutron star mergers in Einstein-scalar-Gauss-Bonnet gravity, *Phys. Rev. D* **110**, 084065 (2024).
- [58] M. Okounkova, L. C. Stein, M. A. Scheel, and S. A. Teukolsky, Numerical binary black hole collisions in dynamical Chern-Simons gravity, *Phys. Rev. D* **100**, 104026 (2019).
- [59] M. Okounkova, Numerical relativity simulation of GW150914 in Einstein dilaton Gauss-Bonnet gravity, *Phys. Rev. D* **102**, 084046 (2020).
- [60] R. Cayuso, P. Figueras, T. França, and L. Lehner, Modelling self-consistently beyond general relativity, *Phys. Rev. Lett.* **131**, 111403 (2023).
- [61] Z. Carson and K. Yagi, Probing beyond-Kerr spacetimes with inspiral-ringdown corrections to gravitational waves, *Phys. Rev. D* **101**, 084050 (2020).
- [62] Z. Carson and K. Yagi, Probing Einstein-dilaton Gauss-Bonnet gravity with the inspiral and ringdown of gravitational waves, *Phys. Rev. D* **101**, 104030 (2020).
- [63] F.-L. Julié, L. Pompili, and A. Buonanno, Inspiral-merger-ringdown waveforms in Einstein-scalar-Gauss-Bonnet gravity within the effective-one-body formalism, *Phys. Rev. D* **111**, 024016 (2025).
- [64] L. Blanchet and B. S. Sathyaprakash, Detecting the tail effect in gravitational wave experiments, *Phys. Rev. Lett.* **74**, 1067 (1995).
- [65] 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).
- [66] 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).
- [67] N. Yunes and F. Pretorius, Fundamental theoretical bias in gravitational wave astrophysics and the parameterized post-Einsteinian framework, *Phys. Rev. D* **80**, 122003 (2009).
- [68] 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).
- [69] H. Narola, S. Roy, and A. S. Sengupta, Beyond general relativity: Designing a template-based search for exotic gravitational wave signals, *Phys. Rev. D* **107**, 024017 (2023).
- [70] A. Sharma, S. Roy, and A. S. Sengupta, Template bank to search for exotic gravitational wave signals from astrophysical compact binaries, *Phys. Rev. D* **109**, 124049 (2024).
- [71] A. Gupta *et al.*, Possible causes of false general relativity violations in gravitational wave observations, *SciPost Phys. Community Rep.* **5** (2025).
- [72] 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).
- [73] 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).
- [74] 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).

- [75] 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).
- [76] S. Roy, M. Haney, G. Pratten, P. T. H. Pang, and C. Van Den Broeck, Improved parametrized test of general relativity using the IMRPhenomX waveform family: Including higher harmonics and precession, *Phys. Rev. D* **113**, 024016 (2026).
- [77] A. Bohé *et al.*, Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors, *Phys. Rev. D* **95**, 044028 (2017).
- [78] R. Cotesta, A. Buonanno, A. Bohé, A. Taracchini, I. Hinder, and S. Ossokine, Enriching the symphony of gravitational waves from binary black holes by tuning higher harmonics, *Phys. Rev. D* **98**, 084028 (2018).
- [79] R. Cotesta, S. Marsat, and M. Pürrer, Frequency domain reduced order model of aligned-spin effective-one-body waveforms with higher-order modes, *Phys. Rev. D* **101**, 124040 (2020).
- [80] T. Dietrich, A. Samajdar, S. Khan, N. K. Johnson-McDaniel, R. Dudi, and W. Tichy, Improving the NRTidal model for binary neutron star systems, *Phys. Rev. D* **100**, 044003 (2019).
- [81] A. Matas *et al.*, Aligned-spin neutron-star-black-hole waveform model based on the effective-one-body approach and numerical-relativity simulations, *Phys. Rev. D* **102**, 043023 (2020).
- [82] G. Pratten *et al.*, Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes, *Phys. Rev. D* **103**, 104056 (2021).
- [83] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries, *Phys. Rev. D* **102**, 064002 (2020).
- [84] M. Colleoni, F. A. Ramis Vidal, N. K. Johnson-McDaniel, T. Dietrich, M. Haney, and G. Pratten, New gravitational waveform model for precessing binary neutron stars with double-spin effects, *Phys. Rev. D* **111**, 064025 (2025).
- [85] E. Berti, K. Yagi, and N. Yunes, Extreme gravity tests with gravitational waves from compact binary coalescences: (I) inspiral-merger, *Gen. Relativ. Gravit.* **50**, 46 (2018).
- [86] S. Tahura and K. Yagi, Parameterized post-Einsteinian gravitational waveforms in various modified theories of gravity, *Phys. Rev. D* **98**, 084042 (2018); *Phys. Rev. D* **101**, 109902(E) (2020).
- [87] K. Yagi, L. C. Stein, N. Yunes, and T. Tanaka, Post-Newtonian, quasi-circular binary inspirals in quadratic modified gravity, *Phys. Rev. D* **85**, 064022 (2012); *Phys. Rev. D* **93**, 029902(E) (2016).
- [88] N. Sennett, S. Marsat, and A. Buonanno, Gravitational waveforms in scalar-tensor gravity at 2PN relative order, *Phys. Rev. D* **94**, 084003 (2016).
- [89] F.-L. Julié and E. Berti, Post-Newtonian dynamics and black hole thermodynamics in Einstein-scalar-Gauss-Bonnet gravity, *Phys. Rev. D* **100**, 104061 (2019).
- [90] B. Shiralilou, T. Hinderer, S. Nissanke, N. Ortiz, and H. Witek, Nonlinear curvature effects in gravitational waves from inspiralling black hole binaries, *Phys. Rev. D* **103**, L121503 (2021).
- [91] B. Shiralilou, T. Hinderer, S. M. Nissanke, N. Ortiz, and H. Witek, Post-Newtonian gravitational and scalar waves in scalar-Gauss-Bonnet gravity, *Classical Quantum Gravity* **39**, 035002 (2022).
- [92] L. Bernard, L. Blanchet, and D. Trestini, Gravitational waves in scalar-tensor theory to one-and-a-half post-Newtonian order, *J. Cosmol. Astropart. Phys.* **08** (2022) 008.
- [93] F.-L. Julié, H. O. Silva, E. Berti, and N. Yunes, Black hole sensitivities in Einstein-scalar-Gauss-Bonnet gravity, *Phys. Rev. D* **105**, 124031 (2022).
- [94] F.-L. Julié, V. Baibhav, E. Berti, and A. Buonanno, Third post-Newtonian effective-one-body Hamiltonian in scalar-tensor and Einstein-scalar-Gauss-Bonnet gravity, *Phys. Rev. D* **107**, 104044 (2023).
- [95] S. Roy and J. Janquart, Testing modified gravity with the eccentric neutron star-black hole merger GW200105, *Phys. Rev. D* **113**, 024056 (2026).
- [96] L. Blanchet, Gravitational radiation from post-Newtonian sources and inspiralling compact binaries, *Living Rev. Relativity* **17**, 2 (2014).
- [97] A. Buonanno, B. Iyer, E. Ochsner, Y. Pan, and B. S. Sathyaprakash, Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors, *Phys. Rev. D* **80**, 084043 (2009).
- [98] B. S. Sathyaprakash and S. V. Dhurandhar, Choice of filters for the detection of gravitational waves from coalescing binaries, *Phys. Rev. D* **44**, 3819 (1991).
- [99] G. Ashton *et al.*, Bilby TGR (Zenodo, 2024), [10.5281/zenodo.10940210](https://zenodo.org/record/10940210).
- [100] LIGO Scientific, Virgo, and KAGRA Collaborations, LVK Algorithm Library—LALSuite, Free software (GPL) (2018), [10.7935/GT1W-FZ16](https://doi.org/10.7935/GT1W-FZ16).
- [101] K. Wette, SWIGLAL: Python and Octave interfaces to the LALSuite gravitational-wave data analysis libraries, *SoftwareX* **12**, 100634 (2020).
- [102] G. Ashton *et al.*, Bilby: A user-friendly Bayesian inference library for gravitational-wave astronomy, *Astrophys. J. Suppl. Ser.* **241**, 27 (2019).
- [103] I. M. Romero-Shaw *et al.*, Bayesian inference for compact binary coalescences with Bilby: Validation and application to the first LIGO–Virgo gravitational-wave transient catalogue, *Mon. Not. R. Astron. Soc.* **499**, 3295 (2020).
- [104] R. J. E. Smith, G. Ashton, A. Vajpeyi, and C. Talbot, Massively parallel Bayesian inference for transient gravitational-wave astronomy, *Mon. Not. R. Astron. Soc.* **498**, 4492 (2020).
- [105] J. S. Speagle, DYNESTY: A dynamic nested sampling package for estimating Bayesian posteriors and evidences, *Mon. Not. R. Astron. Soc.* **493**, 3132 (2020).
- [106] N. J. Cornish and T. B. Littenberg, BayesWave: Bayesian inference for gravitational wave bursts and instrument glitches, *Classical Quantum Gravity* **32**, 135012 (2015).
- [107] T. B. Littenberg and N. J. Cornish, Bayesian inference for spectral estimation of gravitational wave detector noise, *Phys. Rev. D* **91**, 084034 (2015).

- [108] R. Dudi, F. Pannarale, T. Dietrich, M. Hannam, S. Bernuzzi, F. Ohme, and B. Brügmann, Relevance of tidal effects and post-merger dynamics for binary neutron star parameter estimation, *Phys. Rev. D* **98**, 084061 (2018).
- [109] K. Chandra, I. Gupta, R. Gamba, R. Kashyap, D. Chattopadhyay, A. Gonzalez, S. Bernuzzi, and B. S. Sathyaprakash, On the origins, remnant, and multimessenger prospects of the compact binary merger GW230529, *Astrophys. J.* **977**, 167 (2024).
- [110] S. Huth *et al.*, Constraining neutron-star matter with microscopic and macroscopic collisions, *Nature (London)* **606**, 276 (2022).
- [111] 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).
- [112] 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).
- [113] K. Chatziioannou, C.-J. Haster, T. B. Littenberg, W. M. Farr, S. Ghonge, M. Millhouse, J. A. Clark, and N. Cornish, Noise spectral estimation methods and their impact on gravitational wave measurement of compact binary mergers, *Phys. Rev. D* **100**, 104004 (2019).
- [114] N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, BayesWave analysis pipeline in the era of gravitational wave observations, *Phys. Rev. D* **103**, 044006 (2021).
- [115] O. Edy, A. Lundgren, and L. K. Nuttall, Issues of mis-modeling gravitational-wave data for parameter estimation, *Phys. Rev. D* **103**, 124061 (2021).
- [116] R. Macas, J. Pooley, L. K. Nuttall, D. Davis, M. J. Dyer, Y. Lecoeuche, J. D. Lyman, J. McIver, and K. Rink, Impact of noise transients on low latency gravitational-wave event localization, *Phys. Rev. D* **105**, 103021 (2022).
- [117] 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).
- [118] R. C. Myers, Higher derivative gravity, surface terms and string theory, *Phys. Rev. D* **36**, 392 (1987).
- [119] K. Yagi, L. C. Stein, and N. Yunes, Challenging the presence of scalar charge and dipolar radiation in binary pulsars, *Phys. Rev. D* **93**, 024010 (2016).
- [120] K. Yagi, A new constraint on scalar Gauss-Bonnet gravity and a possible explanation for the excess of the orbital decay rate in a low-mass x-ray binary, *Phys. Rev. D* **86**, 081504 (2012).
- [121] A. Saffer and K. Yagi, Tidal deformabilities of neutron stars in scalar-Gauss-Bonnet gravity and their applications to multimessenger tests of gravity, *Phys. Rev. D* **104**, 124052 (2021).
- [122] B. Elder and J. Sakstein, Mapping the weak field limit of scalar-Gauss-Bonnet gravity, *Phys. Rev. D* **107**, 044006 (2023).
- [123] B. Gao, S.-P. Tang, H.-T. Wang, J. Yan, and Y.-Z. Fan, Constraints on Einstein-dilation-Gauss-Bonnet gravity and the electric charge of compact binary systems from GW230529, *Phys. Rev. D* **110**, 044022 (2024).
- [124] <https://www.gwosc.org>.
- [125] E. M. Sängér, S. Roy, M. Agathos, O. Birnholtz, A. Buonanno, T. Dietrich, M. Haney, F.-L. Julié, G. Pratten, J. Steinhoff, C. Van Den Broeck, S. Biscoveanu, P. Char, A. Heffernan, P. Joshi, A. Kedia, R. M. S. Schofield, M. Trevor, and M. Zevin, Data release for “Tests of general relativity with GW230529: A neutron star merging with a lower mass-gap compact object”, [10.5281/zenodo.19387685](https://doi.org/10.5281/zenodo.19387685) (2026).