# Tests of General Relativity with the Binary Black Hole Signals from the LIGO-Virgo Catalog GWTC-1

The LIGO Scientific Collaboration and the Virgo Collaboration

The detection of gravitational waves by Advanced LIGO and Advanced Virgo provides an opportunity to test general relativity in a regime that is inaccessible to traditional astronomical observations and laboratory tests. We present four tests of the consistency of the data with binary black hole gravitational waveforms predicted by general relativity. One test subtracts the best-fit waveform from the data and checks the consistency of the residual with detector noise. The second test checks the consistency of the low- and high-frequency parts of the observed signals. The third test checks that phenomenological deviations introduced in the waveform model (including in the post-Newtonian coefficients) are consistent with zero. The fourth test constrains modifications to the propagation of gravitational waves due to a modified dispersion relation, including that from a massive graviton. We present results both for individual events and also results obtained by combining together particularly strong events from the first and second observing runs of Advanced LIGO and Advanced Virgo, as collected in the catalog GWTC-1. We do not find any inconsistency of the data with the predictions of general relativity and improve our previously presented combined constraints by factors of 1.1 to 2.4. In particular, we bound the mass of the graviton to be $m_g \leq 5.0 \times 10^{-23}$ eV/$c^2$ (90% credible level), an improvement of a factor of 1.5 over our previously presented results. Additionally, we check that the four gravitational-wave events published for the first time in GWTC-1 do not lead to stronger constraints on alternative polarizations than those published previously.

## I. INTRODUCTION

Einstein's theory of gravity, general relativity (GR), has withstood a large number of experimental tests [1]. With the advent of gravitational-wave (GW) astronomy and the observations by the Advanced LIGO [2] and Advanced Virgo [3] detectors, a range of new tests of GR have become possible. These include both weak field tests of the propagation of GWs, as well as tests of the strong field regime of compact binary sources. See [4–8] for previous applications of such tests to GW data.

We report results from tests of GR on all the confident binary black hole GW events in the catalog GWTC-1 [9], i.e., those from the first and second observing runs of the advanced generation of detectors. Besides all of the events previously announced (GW150914, GW151012, GW151226, GW170104, GW170608, and GW170814) [5–7, 10–13], this includes the four new GW events reported in [14] (GW170729, GW170809, GW170818, and GW170823). We do not investigate any of the marginal triggers in GWTC-1, which have a false-alarm rate (FAR) greater than one per year. Table I displays a complete list of the events we consider. Tests of GR on the binary neutron star event GW170817 are described in [8].

The search results in [14] originate from two modeled searches and one weakly modeled search [5, 11, 14, 15]. The modeled searches use templates based on GR to find candidate events and to assess their significance. However, detection by such searches does not in itself imply full compatibility of the signal with GR [16, 17]. The weakly modeled search relies on coherence of signals between multiple detectors, as expected for an astrophysical source. While it assumes that the morphology of the signal resembles a chirp (whose frequency increases with time), as expected for a compact binary coalescence, it does not assume that the detailed waveform shape agrees with GR. A transient signal strongly deviating from GR would likely be found by the weakly modeled search, even if missed by the modeled searches. So far, however, all significant [FAR $< (1\,\mathrm{yr})^{-1}$] transient signals found by the weakly modeled search were also found by at least one of the modeled searches [14].

At present, there are no complete theories of gravity other than GR that are mathematically and physically viable and provide well defined alternative predictions for the waveforms arising from the coalescence of two black holes (if, indeed, these theories even admit black holes). Thus, we cannot test GR by direct comparison with other specific theories. Instead, we can (i) check the consistency of the GR predictions with the data and (ii) introduce *ad hoc* modifications in GR waveforms to determine the degree to which the values of the deviation parameters agree with GR. These methods are agnostic to any particular choice of alternative theory. For the most part, our results should therefore be interpreted as observational constraints on possible GW phenomenologies, independent of the overall suitability or well-posedness of any specific alternative to GR. These limits are useful in providing a quantitative indication of the degree to which the data is described by GR; they may also be interpreted more specifically in the context of any given alternative to produce constraints, if applicable.

In particular, with regard to the consistency of the GR predictions (i), we (a) look for residual power after subtracting the best-fitting GR waveform from the data, and (b) evaluate the consistency of the high and low frequency components of the observed signal. With regard to deviations from GR (ii), we separately introduce parametrizations for (a) the emitted waveform, and (b) its propagation. The former could be viewed as representing possible GR modifications in the strong-field region close to the binary, while the latter would correspond to weak-field modifications away from the source. Although we consider these independently, modifications to GW propagation would most likely be accompanied by modifications to GW generation in any given extension of GR. We have also checked that none of the events discussed here provide stronger constraints on models with purely vector and purely scalar GW polarizations than those previously published in [7, 8]. Our analyses do not reveal any inconsistency of the data with the

predictions of GR.

Limits on deviations from GR for individual events are dominated by statistical errors due to detector noise. These errors can be reduced by appropriately combining results from multiple events. Sources of systematic errors, on the other hand, include uncertainties in the detector calibration and power spectral density (PSD) estimation and errors in the modeling of waveforms in GR. Detector calibration uncertainties are modeled as corrections to the measured detector response function and are marginalized over. Studies on the effect of PSD uncertainties are currently ongoing. A full characterization of the systematic errors due to the GR waveform models that we employ is beyond the scope of this study; some investigations can be found in [18–22].

This paper is organized as follows. Section II provides an overview of the data sets employed here, while Sec. III details which GW events are used to produce the individual and combined results presented in this paper. In Sec. IV we explain the gravitational waveforms and data analysis formalisms which our tests of GR are based on, before we present the results in the following sections. Section V contains two signal consistency tests: the residuals test in V A and the inspiral-merger-ringdown consistency test in V B. Results from parameterized tests are given in Sec. VI for GW generation, and in Sec. VII for GW propagation. We briefly discuss the study of GW polarizations in Sec. VIII. Finally, we conclude in Sec. IX. We give results for individual events and some checks on waveform systematics in the Appendix.

The results of each test and associated data products can be found in Ref. [23]. The GW strain data for all the events considered are available at the Gravitational Wave Open Science Center [24].

## II. DATA, CALIBRATION AND CLEANING

The first observing run of Advanced LIGO (O1) lasted from September 12$^{\text{th}}$, 2015 to January 19$^{\text{th}}$, 2016. The second observing run (O2) lasted from November 30$^{\text{th}}$, 2016 to August 25$^{\text{th}}$, 2017, with the Advanced Virgo observatory joining on August 1$^{\text{st}}$, 2017. This paper includes all GW events originating from the coalescence of two black holes found in these two data sets and published in [5, 14].

The GW detector's response to changes in the differential arm length (the interferometer's degree of freedom most sensitive to GWs) must be calibrated using independent, accurate, absolute references. The LIGO detectors use photon recoil (radiation pressure) from auxiliary laser systems to induce mirror motions that change the arm cavity lengths, allowing a direct measure of the detector response [25–27]. Calibration of Virgo relies on measurements of Michelson interference fringes as the main optics swing freely, using the primary laser wavelength as a fiducial length. Subsequent measurements propagate the calibration to arrive at the final detector response [28, 29]. These complex-valued, frequency-dependent measurements of the LIGO and Virgo detectors' response yield the uncertainty in their respective estimated amplitude and phase of the GW strain output. The amplitude and phase correction factors are modeled as cubic splines and marginalized over in the estimation of astrophysical source parameters [14, 30–32]. Additionally, the uncertainty in the time stamping of Virgo data (much larger than the LIGO timing uncertainty, which is included in the phase correction factor) is also accounted for in the analysis.

Post-processing techniques to subtract noise contributions and frequency lines from the data around gravitational-wave events were developed in O2 and introduced in [7, 13, 33], for the astrophysical parameter estimation of GW170608, GW170814, and GW170817. This noise subtraction was achieved using optimal Wiener filters to calculate coupling transfer functions from auxiliary sensors [34]. A new, optimized parallelizable method in the frequency domain [35] allows large scale noise subtraction on LIGO data. All of the O2 analyses presented in this manuscript use the noise-subtracted data set with the latest calibration available. The O1 data set is the same used in previous publications. Reanalysis of the O1 events is motivated by improvements in the parameter estimation pipeline, an improved frequency-dependent calibration, and the availability of new waveform models.

## III. EVENTS AND SIGNIFICANCE

We present results for all confident detections of binary black hole events in GTWC-1 [9], i.e., all such events detected during O1 and O2 with a FAR lower than one per year, as published in [14]. The central columns of Table I list the FARs of each event as evaluated by the three search pipelines used in [14]. Two of these pipelines (PyCBC and GstLAL) rely on waveform templates computed from binary black hole coalescences in GR. Making use of a measure of significance that assumes the validity of GR could potentially lead to biases in the selection of events to be tested, systematically disfavoring signals in which a GR violation would be most evident. Therefore, it is important to consider the possibilities that (1) there were GW signals with such large deviations from GR that they were missed entirely by the modeled searches, and (2) there were events that were picked up by the modeled searches but classified as marginal (and thus excluded from our analysis) because of their significant deviations from GR.

These worries can largely be dispelled by considering the third GW search pipeline, the coherent WaveBurst (cWB) weakly modeled search presented in [14]. This cWB search [15, 36, 37] was tuned to detect chirping signals—like those that would be expected from compact binary coalescences—but was not tuned to any specific GR predictions.[1] cWB is most sensitive to short signals from high-mass binary black holes. It is still able to detect signals from lower-mass binaries (e.g., GW151226), though with reduced significance compared to the modeled searches. Thus, a signal from a low-mass binary, or a marginal event, with a significant departure from the GR

[1] Chirping signals from compact binary coalescences are a feature of many theories of gravity. All that is required is that the orbital frequency increases as the binary radiates energy and angular momentum in GWs.

predictions (hence not detected by the GR modeled searches) would not necessarily be detected by the cWB search with a FAR $< (1\,\mathrm{yr})^{-1}$. However, if there is a population of such signals, they will not all be weak and/or from low-mass binaries. Thus, one would expect some of the signals in the population to be detected by cWB, even if they evade detection by the modeled searches.

All signals detected by the cWB search with FAR $< (1\,\mathrm{yr})^{-1}$ were also found by at least one modeled search with FAR $< (1\,\mathrm{yr})^{-1}$. Given the above considerations, this is evidence that our analysis does not exclude chirping GW signals that were missed in the modeled searches because of drastic departures from GR. Similarly, this is also evidence against the possibility of marginal events representing a population of GR-deviating signals, as none of them show high significance [FAR $< (1\,\mathrm{yr})^{-1}$] in the cWB search only. Thus, we believe that we have not biased our analysis by considering only the ten events with FAR $< (1\,\mathrm{yr})^{-1}$, as published in [14].

We consider each of the GW events individually, carrying out different analyses on a case-by-case basis. Some of the tests presented here, such as the inspiral-merger-ringdown (IMR) consistency test in Sec. V B and the parameterized tests in Sec. VI, distinguish between the inspiral and the post-inspiral regimes of the signal. The separation between these two regimes is performed in the frequency-domain, choosing a particular cutoff frequency determined by the parameters of the event. Larger-mass systems merge at lower frequencies, presenting a short inspiral signal in band; lower mass systems have longer observable inspiral signals, but the detector's sensitivity decreases at higher frequencies and hence the post-inspiral signal becomes less informative. Therefore, depending on the total mass of the system, a particular signal might not provide enough information within the sensitive frequency band of the GW detectors for all analyses.

As a proxy for the amount of information that can be extracted from each part of the signal, we calculate the signal-to-noise ratio (SNR) of the inspiral and the post-inspiral parts of the signals separately. We only apply inspiral (post-inspiral) tests if the inspiral (post-inspiral) SNR is greater than 6. Each test uses a different inspiral-cutoff frequency, and hence they assign different SNRs to the two regimes (details provided in the relevant section for each test). In Table I we indicate which analyses have been performed on which event, based on this frequency and the corresponding SNR.[2]

In addition to the individual analysis of each event, we derive combined constraints on departures from GR using multiple signals simultaneously. Constraints from individual events are largely dominated by statistical uncertainties due to detector noise. Combining events together can reduce such statistical errors on parameters that take consistent values across all events. However, it is impossible to make joint probabilistic statements from multiple events without prior assumptions about the nature of each observation and how it relates to others in the set. This means that, although there are well-defined statistical procedures for producing joint results, there is no unique way of doing so.

In light of this, we adopt what we take to be the most straightforward strategy, although future studies may follow different criteria. First, in combining events we assume that deviations from GR are manifested equally across events, independent of source properties. This is justified for studies of modified GW propagation, since those effects should not depend on the source.[3] For other analyses, it is quite a strong assumption to take all deviations from GR to be independent of source properties. Such combined tests should not be expected to necessarily reveal generic source-dependent deviations, although they might if the departures from GR are large enough (see, e.g., [38]). Future work may circumvent this issue by combining marginalized likelihood ratios (Bayes factors), instead of posterior probability distributions [39].

Second, we choose to produce combined constraints only from events that were found in both modeled searches (PyCBC [40–42] and GstLAL [43, 44]) with a FAR of at most one per one-thousand years. This ensures that there is a very small probability of inclusion of a non-astrophysical event. The events used for the combined results are indicated with bold names in Table I. The events thus excluded from the combined analysis have low SNR and would therefore contribute only marginally to tightened constraints. Excluding marginal events from our analyses amounts to assigning a null *a priori* probability to the possibility that those data contain any information about the tests in question. This is, in a sense, the most conservative choice.

In summary, we enforce two significance thresholds: FAR $< (1\,\mathrm{yr})^{-1}$, for single-event analyses, and FAR $< (1000\,\mathrm{yr})^{-1}$, for combined results. This two-tiered setup allows us to produce conservative joint results by including only the most significant events, while also providing information about a broader (less significant) set of triggers. This is intended to enable the interested reader to combine individual results with less stringent criteria and under different statistical assumptions, according to their specific needs and tolerance for false positives. In the future, we may adapt our thresholds depending on the rate of detections.

## IV. PARAMETER INFERENCE

The starting point for all the analyses presented here are waveform models that describe the GWs emitted by coalescing black-hole binaries. The GW signature depends on the intrinsic parameters describing the binary as well as the extrinsic parameters specifying the location and orientation of the source with respect to the detector network. The intrinsic parameters for

[2] While we perform these tests on all events with SNR $> 6$ in the appropriate regime, in a few cases the results appear uninformative and the posterior distribution extends across the entire prior considered. Since the results are prior dependent, upper limits should not be set from these individual analyses. See Sec. 3 of the Appendix for details.

[3] Propagation effects do depend critically on source distance. However, this dependence is factored out explicitly, in a way that allows for combining events as we do here (see Sec. VII).

TABLE I. The GW events considered in this paper, separated by observing run. The first block of columns gives the names of the events and lists some of their relevant properties obtained using GR waveforms (luminosity distance $D_{\rm L}$, source frame total mass $M_{\rm tot}$ and final mass $M_{\rm f}$, and dimensionless final spin $a_{\rm f}$). The next block of columns gives the significance, measured by the false-alarm-rate (FAR), with which each event was detected by each of the three searches employed, as well as the matched filter signal-to-noise ratio from the stochastic sampling analyses with GR waveforms. A dash indicates that an event was not identified by a search. The parameters and SNR values give the medians and 90% credible intervals. All the events except for GW151226 and GW170729 are consistent with a binary of nonspinning black holes (when analyzed assuming GR). See [14] for more details about all the events.[a] The last block of columns indicates which GR tests are performed on a given event: RT = residuals test (Sec. V A); IMR = inspiral-merger-ringdown consistency test (Sec. V B); PI & PPI = parameterized tests of GW generation for inspiral and post-inspiral phases (Sec. VI); MDR = modified GW dispersion relation (Sec. VII). The events with bold names are used to obtain the combined results for each test.

| Event | Properties | | | | FAR | | | SNR | GR tests performed | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $D_{\rm L}$ [Mpc] | $M_{\rm tot}$ [$M_\odot$] | $M_{\rm f}$ [$M_\odot$] | $a_{\rm f}$ | PyCBC [yr$^{-1}$] | GstLAL [yr$^{-1}$] | cWB [yr$^{-1}$] | | RT | IMR | PI | PPI | MDR |
| **GW150914**[b] | $430^{+150}_{-170}$ | $66.2^{+3.7}_{-3.3}$ | $63.1^{+3.3}_{-3.0}$ | $0.69^{+0.05}_{-0.04}$ | $< 1.5\times10^{-5}$ | $< 1.0\times10^{-7}$ | $< 1.6\times10^{-4}$ | $25.3^{+0.1}_{-0.2}$ | ✓ | ✓ | ✓ | ✓ | ✓ |
| GW151012[b] | $1060^{+550}_{-480}$ | $37.3^{+10.6}_{-3.9}$ | $35.7^{+10.7}_{-3.8}$ | $0.67^{+0.13}_{-0.11}$ | 0.17 | $7.9\times10^{-3}$ | – | $9.2^{+0.3}_{-0.4}$ | ✓ | – | – | ✓ | ✓ |
| **GW151226**[b,c] | $440^{+180}_{-190}$ | $21.5^{+6.2}_{-1.5}$ | $20.5^{+6.4}_{-1.5}$ | $0.74^{+0.07}_{-0.05}$ | $< 1.7\times10^{-5}$ | $< 1.0\times10^{-7}$ | 0.02 | $12.4^{+0.2}_{-0.3}$ | ✓ | – | ✓ | – | ✓ |
| **GW170104** | $960^{+440}_{-420}$ | $51.3^{+5.3}_{-4.2}$ | $49.1^{+5.2}_{-4.0}$ | $0.66^{+0.08}_{-0.11}$ | $< 1.4\times10^{-5}$ | $< 1.0\times10^{-7}$ | $2.9\times10^{-4}$ | $14.0^{+0.2}_{-0.3}$ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **GW170608** | $320^{+120}_{-110}$ | $18.6^{+3.1}_{-0.7}$ | $17.8^{+3.2}_{-0.7}$ | $0.69^{+0.04}_{-0.04}$ | $< 3.1\times10^{-4}$ | $< 1.0\times10^{-7}$ | $1.4\times10^{-4}$ | $15.6^{+0.2}_{-0.3}$ | ✓ | – | ✓ | ✓ | ✓ |
| GW170729[d] | $2760^{+1380}_{-1340}$ | $85.2^{+15.6}_{-11.1}$ | $80.3^{+14.6}_{-10.2}$ | $0.81^{+0.07}_{-0.13}$ | 1.4 | 0.18 | 0.02 | $10.8^{+0.4}_{-0.5}$ | ✓ | ✓ | – | ✓ | ✓ |
| **GW170809** | $990^{+320}_{-380}$ | $59.2^{+5.4}_{-3.9}$ | $56.4^{+5.2}_{-3.7}$ | $0.70^{+0.08}_{-0.09}$ | $1.4\times10^{-4}$ | $< 1.0\times10^{-7}$ | – | $12.7^{+0.2}_{-0.3}$ | ✓ | ✓ | – | ✓ | ✓ |
| **GW170814** | $580^{+160}_{-210}$ | $56.1^{+3.4}_{-2.7}$ | $53.4^{+3.2}_{-2.4}$ | $0.72^{+0.07}_{-0.05}$ | $< 1.2\times10^{-5}$ | $< 1.0\times10^{-7}$ | $< 2.1\times10^{-4}$ | $17.8^{+0.3}_{-0.3}$ | ✓ | ✓ | ✓ | ✓ | ✓ |
| GW170818 | $1020^{+430}_{-360}$ | $62.5^{+5.1}_{-4.0}$ | $59.8^{+4.8}_{-3.8}$ | $0.67^{+0.07}_{-0.08}$ | – | $4.2\times10^{-5}$ | – | $11.9^{+0.3}_{-0.4}$ | ✓ | ✓ | – | ✓ | ✓ |
| **GW170823** | $1850^{+840}_{-840}$ | $68.9^{+9.9}_{-7.1}$ | $65.6^{+9.4}_{-6.6}$ | $0.71^{+0.08}_{-0.10}$ | $< 3.3\times10^{-5}$ | $< 1.0\times10^{-7}$ | $2.1\times10^{-3}$ | $12.1^{+0.2}_{-0.3}$ | ✓ | ✓ | – | ✓ | ✓ |

[a] The parameters given in this table differ slightly from those in v2 of the arXiv version of [14] since they correct for a small issue in the application of priors for spin parameters.

[b] The FARs for these events differ from those in [5] because the data were re-analyzed with the new pipeline statistics used in O2 (see [14] for more details).

[c] At least one black hole has dimensionless spin > 0.28 (99% credible level).

[d] This event has a higher significance in the unmodeled search than in the modeled searches. Additionally, at least one black hole has dimensionless spin > 0.27 (99% credible level).

circularized black-hole binaries in GR are the two masses $m_i$ of the black holes and the two spin vectors $\vec{S}_i$ defining the rotation of each black hole, where $i \in \{1, 2\}$ labels the two black holes. We assume that the binary has negligible orbital eccentricity, as is expected to be the case when the binary enters the band of ground-based detectors [45, 46] (except in some more extreme formation scenarios,[4] e.g., [53–56]). The extrinsic parameters comprise four parameters that specify the space-time location of the binary black-hole, namely the sky location (right ascension and declination), the luminosity distance, and the time of coalescence. In addition, there are three extrinsic parameters that determine the orientation of the binary with respect to Earth, namely the inclination angle of the orbit with respect to the observer, the polarization angle, and the orbital phase at coalescence.

We employ two waveform families that model binary black holes in GR: the effective-one-body based SEOBNRv4 [18] waveform family that assumes non-precessing spins for the black holes (we use the frequency domain reduced order model SEOBNRv4_ROM for reasons of computational efficiency), and the phenomenological waveform family IMRPhenomPv2 [19, 57, 58] that models the effects of precessing spins using two effective parameters by twisting up the underlying aligned-spin model. We use IMRPhenomPv2 to obtain all the main results given in this paper, and use SEOBNRv4 to check the robustness of these results, whenever possible. When we use IMRPhenomPv2, we impose a prior $m_1/m_2 \leq 18$ on the mass ratio, as the waveform family is not calibrated against numerical relativity simulations for $m_1/m_2 > 18$. We do not impose a similar prior when using SEOBNRv4, since it includes information about the extreme mass ratio limit. Neither of these waveform models includes the full spin dynamics (which requires 6 spin parameters). Fully precessing waveform models have been recently developed [21, 59–61] and will be used in future applications of these tests.

The waveform models used in this paper do not include the effects of subdominant (non-quadrupole) modes, which are expected to be small for comparable-mass binaries [62, 63]. The first generation of binary black hole waveform models including spin and higher order modes has recently been developed [61, 64–66]. Preliminary results in [14], using NR simulations supplemented by NR surrogate waveforms, indicate that the higher mode content of the GW signals detected by Advanced LIGO and Virgo is weak enough that models without the effect of subdominant modes do not introduce substantial biases in the intrinsic parameters of the binary. For

[4] These scenarios could occur often enough, compared to the expected rate of detections, that the inclusion of eccentricity in waveform models is a necessity for tests of GR in future observing runs; see, e.g., [47–52] for recent work on developing such waveform models.

unequal-mass binaries, the effect of the non-quadrupole modes is more pronounced [67], particularly when the binary's orientation is close to edge-on. In these cases, the presence of non-quadrupole modes can show up as a deviation from GR when using waveforms that only include the quadrupole modes, as was shown in [68]. Applications of tests of GR with the new waveform models that include non-quadrupole modes will be carried out in the future.

We believe that our simplifying assumptions on the waveform models (zero eccentricity, simplified treatment of spins, and neglect of subdominant modes) are justified by astrophysical considerations and previous studies. Indeed, as we show in the remainder of the paper, the observed signals are consistent with the waveform models. Of course, had our analyses resulted in evidence for violations of GR, we would have had to revisit these simplifications very carefully.

The tests described in this paper are performed within the framework of Bayesian inference, by means of the LALInference code [31] in the LIGO Scientific Collaboration Algorithm Library Suite (LALSuite) [69]. We estimate the PSD using the BayesWave code [70, 71], as described in Appendix B of [14]. Except for the residuals test described in Sec. V A, we use the waveform models described in this section to estimate from the data the posterior distributions of the parameters of the binary. These include not only the intrinsic and extrinsic parameters mentioned above, but also other parameters that describe possible departures from the GR predictions. Specifically, for the parameterized tests in Secs. VI and VII, we modify the phase $\Phi(f)$ of the frequency-domain waveform

$$\tilde{h}(f) = A(f)e^{i\Phi(f)}. \tag{1}$$

For the GR parameters, we use the same prior distributions as the main parameter estimation analysis described in [14], though for a number of the tests we need to extend the ranges of these priors to account for correlations with the non-GR parameters, or for the fact that only a portion of the signal is analyzed (as in Sec. V B). For the model agnostic residual test described in Sec. V A, we use the BayesWave code [70] which describes the GW signals in terms of a number of Morlet-Gabor wavelets.

## V. CONSISTENCY TESTS

### A. Residuals test

One way to evaluate the ability of GR to describe GW signals is to subtract the best-fit template from the data and make sure the residuals have the statistical properties expected of instrumental noise. This largely model-independent test is sensitive to a wide range of possible disagreements between the data and our waveform models, including those caused by deviations from GR and by modeling systematics. This analysis can look for GR violations without relying on specific parametrizations of the deviations, making it a versatile tool. Results from a similar study on our first detection were already presented in [4].

In order to establish whether the residuals agree with noise (Gaussian or otherwise), we proceed as follows. For each event in our set, we use LALInference and the IMRPhenomPv2 waveform family to obtain an estimate of the best-fit (i.e., maximum likelihood) binary black-hole waveform based on GR. This best-fit waveform is then subtracted from the data to obtain the residuals. If the GR-based model provides a good description of the signal, we expect the resulting residuals at each detector to lack any significant coherent SNR beyond what is expected from noise fluctuations. We compute the coherent SNR using BayesWave, which models the multi-detector residuals as a superposition of incoherent Gaussian noise and an elliptically-polarized coherent signal. The residual network SNR reported by BayesWave is the SNR that would correspond to such a coherent signal.

In particular, for each event, BayesWave produces a distribution of possible residual signals consistent with the data, together with corresponding *a posteriori* probabilities. This is trivially translated into a probability distribution over the coherent residual SNR. We summarize each of these distributions by computing the corresponding 90%-credible upper limit ($\mathrm{SNR}_{90}$). This produces one number per event that represents an upper bound on the coherent power that could be present in its residuals.

We may translate the $\mathrm{SNR}_{90}$ into a measure of how well the best-fit templates describe the signals in our data. We do this through the fitting factor [72], $\mathrm{FF} \coloneqq \mathrm{SNR}_{\mathrm{GR}}/(\mathrm{SNR}^2_{\mathrm{res}} + \mathrm{SNR}^2_{\mathrm{GR}})^{1/2}$, where $\mathrm{SNR}_{\mathrm{res}}$ is the coherent residual SNR and $\mathrm{SNR}_{\mathrm{GR}}$ is the network SNR of the best-fit template (see Table I for network SNRs). By setting $\mathrm{SNR}_{\mathrm{res}} = \mathrm{SNR}_{90}$, we produce a 90%-credible lower limit on the fitting factor ($\mathrm{FF}_{90}$). Because FF is itself a lower limit on the overlap between the true and best-fit templates, so is $\mathrm{FF}_{90}$. As in [4], we may then assert that the disagreement between the true waveform and our GR-based template is at most $(1-\mathrm{FF}_{90})\times 100\%$. This is interesting as a measure of the sensitivity of our test, but does not tell us about the consistency of the residuals with instrumental noise.

To assess whether the obtained residual $\mathrm{SNR}_{90}$ values are consistent with detector noise, we run an identical BayesWave analysis on 200 different sets of noise-only detector data near each event. This allows us to estimate the $p$-value for the null hypothesis that the residuals are consistent with noise. The $p$-value gives the probability of noise producing coherent power with $\mathrm{SNR}^{\mathrm{n}}_{90}$ greater than or equal to the residual value $\mathrm{SNR}_{90}$, i.e., $p \coloneqq P(\mathrm{SNR}^{\mathrm{n}}_{90} \geq \mathrm{SNR}_{90} \mid \mathrm{noise})$. In that sense, a smaller $p$-value indicates a smaller chance that the residual power arose from instrumental noise only. For each event, our estimate of $p$ is produced from the fraction of noise instantiations that yielded $\mathrm{SNR}^{\mathrm{n}}_{90} \geq \mathrm{SNR}_{90}$ (that is, from the empirical survival function).[5]

Our results are summarized in Table II. For each event, we present the values of the residual $\mathrm{SNR}_{90}$, the lower limit on the fitting factor ($\mathrm{FF}_{90}$), and the $\mathrm{SNR}_{90}$ $p$-value. The

[5] Computing $p$-values would not be necessary if the noise was perfectly Gaussian, in which case we could predict the noise-only distribution of $\mathrm{SNR}^{\mathrm{n}}_{90}$ from first principles.

TABLE II. Results of the residuals analysis. For each event, this table presents the 90%-credible upper limit on the reconstructed network SNR after subtraction of the best-fit GR waveform ($SNR_{90}$), a corresponding lower limit on the fitting factor ($FF_{90}$ in the text), and the $SNR_{90}$ $p$-value. $SNR_{90}$ is a measure of the maximum possible coherent signal power not captured by the best-fit GR template, while the $p$-value is an estimate of the probability that instrumental noise produced such $SNR_{90}$ or higher. We also indicate which interferometers (IFOs) were used in the analysis of a given event, either the two Advanced LIGO detectors (HL) or the two Advanced LIGO detectors plus Advanced Virgo (HLV). See Sec. V A in the main text for details.

| Event | IFOs | Residual $SNR_{90}$ | Fitting factor | $p$-value |
|---|---|---|---|---|
| GW150914 | HL | 6.4 | $\geq 0.97$ | 0.34 |
| GW151012 | HL | 6.9 | $\geq 0.81$ | 0.18 |
| GW151226 | HL | 5.7 | $\geq 0.91$ | 0.76 |
| GW170104 | HL | 5.2 | $\geq 0.94$ | 0.97 |
| GW170608 | HL | 7.8 | $\geq 0.90$ | 0.07 |
| GW170729 | HLV | 6.5 | $\geq 0.87$ | 0.72 |
| GW170809 | HLV | 6.7 | $\geq 0.91$ | 0.73 |
| GW170814 | HLV | 8.6 | $\geq 0.90$ | 0.19 |
| GW170818 | HLV | 10.1 | $\geq 0.78$ | 0.13 |
| GW170823 | HL | 5.4 | $\geq 0.92$ | 0.89 |

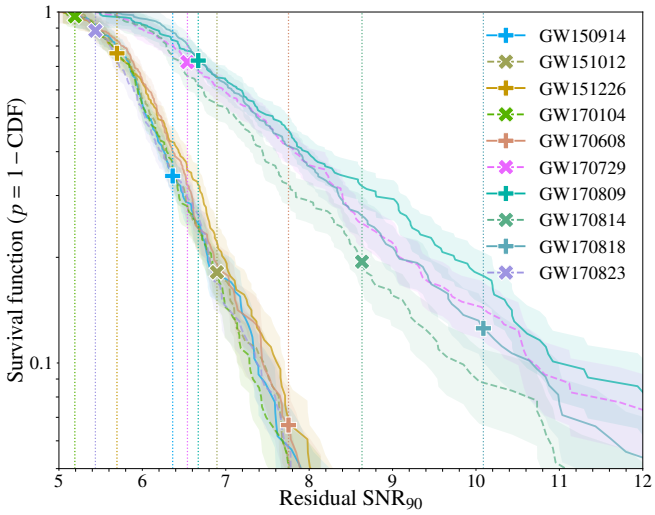


FIG. 1. Survival function ($p$ = 1 − CDF) of the 90%-credible upper limit on the network SNR ($SNR_{90}$) for each event (solid or dashed curves), compared to the measured residual values (vertical dotted lines). For each event, the value of the survival function at the measured $SNR_{90}$ gives the $p$-value reported in Table II (markers). The colored bands correspond to uncertainty regions for a Poisson process and have half width $\pm p/\sqrt{N}$, with $N$ the number of noise-only instantiations that yielded $SNR^{n}_{90}$ greater than the abscissa value.

background distributions that resulted in those $p$-values are shown in Fig. 1. In Fig. 1 we represent these distributions through the empirical estimate of their survival functions, i.e., $p(SNR_{90})$ = 1 − CDF($SNR_{90}$), with "CDF" the cumulative distribution function. Fig. 1 also displays the actual value of $SNR_{90}$ measured from the residuals of each event (dotted vertical line). In each case, the height of the curve evaluated at the $SNR_{90}$ measured for the corresponding detection yields the $p$-value reported in Table II (markers in Fig. 1).

TABLE III. Results from the inspiral-merger-ringdown consistency test for selected binary black hole events. $f_c$ denotes the cutoff frequency used to demarcate the division between the inspiral and post-inspiral regimes; $\rho_{IMR}$, $\rho_{insp}$, and $\rho_{post-insp}$ are the median values of the SNR in the full signal, the inspiral part, and the post-inspiral part, respectively; and the GR quantile denotes the fraction of the posterior enclosed by the isoprobability contour that passes through the GR value, with smaller values indicating better consistency with GR.

| Event | $f_c$ [Hz] | $\rho_{IMR}$ | $\rho_{insp}$ | $\rho_{post-insp}$ | GR quantile [%] |
|---|---|---|---|---|---|
| GW150914 | 132 | 25.3 | 19.4 | 16.1 | 55.5 |
| GW170104 | 143 | 13.7 | 10.9 | 8.5 | 24.4 |
| GW170729 | 91 | 10.7 | 8.6 | 6.9 | 10.4 |
| GW170809 | 136 | 12.7 | 10.6 | 7.1 | 14.7 |
| GW170814 | 161 | 16.8 | 15.3 | 7.2 | 7.8 |
| GW170818 | 128 | 12.0 | 9.3 | 7.2 | 25.5 |
| GW170823 | 102 | 11.9 | 7.9 | 8.5 | 80.4 |

The values of residual $SNR_{90}$ vary widely among events because they depend on the specific state of the instruments at the time of detection: segments of data with elevated noise levels are more likely to result in spurious coherent residual power, even if the signal agrees with GR. In particular, the background distributions for events seen by three detectors are qualitatively different from those seen by only two. This is both due to (i) the fact that BayesWave is configured to expect the SNR to increase with the number of detectors and (ii) the fact that Virgo data present a higher rate of non-Gaussianities than LIGO. We have confirmed both these factors play a role by studying the background $SNR_{90}$ distributions for real data from each possible pair of detectors, as well as distributions over fabricated Gaussian noise. Specifically, removing Virgo from the analysis results in a reduction in the coherent residual power for GW170729 ($SNR^{HL}_{90}$ = 6.0), GW170809 ($SNR^{HL}_{90}$ = 6.3), GW170814 ($SNR^{HL}_{90}$ = 5.9), and GW170818 ($SNR^{HL}_{90}$ = 6.6).

The event-by-event variation of $SNR_{90}$ is also reflected in the values of $FF_{90}$. GW150914 provides the strongest result with $FF_{90}$ = 0.97, which corresponds to an upper limit of 3% on the magnitude of potential deviations from our GR-based template,[6] in the specific sense defined in [4] and discussed above. On the other hand, GW170818 yields the weakest result with $FF_{90}$ = 0.78 and a corresponding upper limit on waveform mismatch of 22%. The average $FF_{90}$ over all events is 0.89.

The set of $p$-values shown in Table II is consistent with all coherent residual power being due to instrumental noise. Assuming that this is indeed the case, we expect the $p$-values to be uniformly distributed over [0, 1], which explains the variation in Table II. With only ten events, however, it is difficult to obtain strong quantitative evidence of the uniformity of this

[6] This value is better than the one quoted in [4] by 1 percentage point. The small difference is explained by several factors, including that paper's use of the maximum *a posteriori* waveform (instead of maximum likelihood) and 95% (instead of 90%) credible intervals, as well as improvements in data quality.

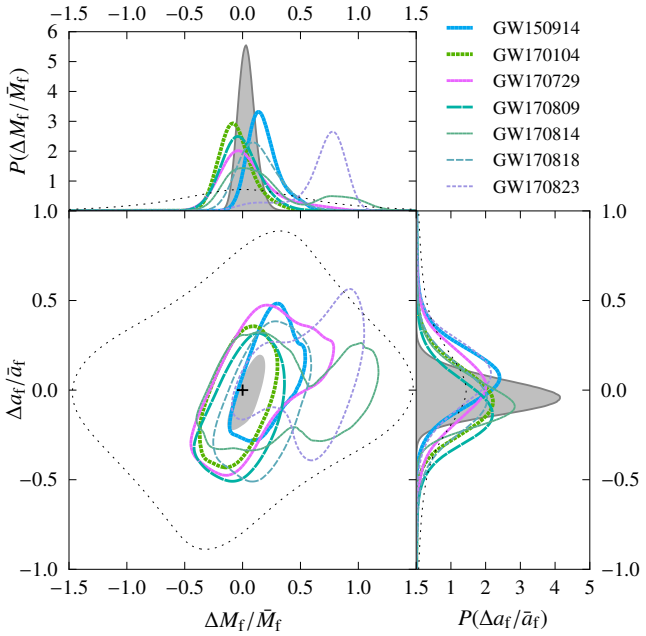


FIG. 2. Results of the inspiral-merger-ringdown consistency test for the selected BBH events (see Table I). The main panel shows 90% credible regions of the posterior distributions of $(\Delta M_f/\bar{M}_f, \Delta a_f/\bar{a}_f)$, with the cross marking the expected value for GR. The side panels show the marginalized posteriors for $\Delta M_f/\bar{M}_f$ and $\Delta a_f/\bar{a}_f$. The thin black dashed curve represents the prior distribution, and the grey shaded areas correspond to the combined posteriors from the five most significant events (as outlined in Sec. III and Table I).

distribution. Nevertheless, we follow Fisher's method [73] to compute a meta $p$-value for the null hypothesis that the individual $p$-values in Table II are uniformly distributed. We obtain a meta $p$-value of 0.4, implying that there is no evidence for coherent residual power that cannot be explained by noise alone. All in all, this means that there is no statistically significant evidence for deviations from GR.

### B. Inspiral-merger-ringdown consistency test

The inspiral-merger-ringdown consistency test for binary black holes [38, 74] checks the consistency of the low-frequency part of the observed signal (roughly corresponding to the inspiral of the black holes) with the high-frequency part (to a good approximation, produced by the post-inspiral stages). The cutoff frequency $f_c$ between the two regimes is chosen as the frequency of the innermost stable circular orbit of a Kerr black hole [75], with mass and dimensionless spin equal to the median values of the posterior distribution of the remnant's mass and spin. This determination of $f_c$ is performed separately for each event and based on parameter inference of the full signal (see Table III for values of $f_c$).[7] The binary's parameters are then estimated independently from the low (high) frequency parts of the data by restricting the noise-weighted integral in the likelihood calculation to frequencies below (above) this frequency cutoff $f_c$. For each of these independent estimates of the source parameters, we make use of fits to numerical-relativity simulations given in [76–78] to infer the mass $M_f$ and dimensionless spin magnitude $a_f = c|\vec{S}_f|/(GM_f^2)$ of the remnant black hole.[8] If the data are consistent with GR, these two independent estimates have to be consistent with each other [38, 74]. Because this consistency test ultimately compares between the inspiral and the post-inspiral results, posteriors of both parts must be informative. In the case of low-mass binaries, the SNR in the part $f > f_c$ is insufficient to perform this test, so that we only analyze seven events as indicated in Tables I and III.

In order to quantify the consistency of the two different estimates of the final black hole's mass and spin we define two dimensionless quantities that quantify the fractional difference between them: $\Delta M_f/\bar{M}_f := 2\,(M_f^{\rm insp} - M_f^{\rm post\text{-}insp})/(M_f^{\rm insp} + M_f^{\rm post\text{-}insp})$ and $\Delta a_f/\bar{a}_f := 2\,(a_f^{\rm insp} - a_f^{\rm post\text{-}insp})/(a_f^{\rm insp} + a_f^{\rm post\text{-}insp})$, where the superscripts indicate the estimates of the mass and spin from the inspiral and post-inspiral parts of the signal.[9] The posteriors of these dimensionless parameters, estimated from different events, are shown in Fig. 2. For all events, the posteriors are consistent with the GR value $(\Delta M_f/\bar{M}_f = 0, \Delta a_f/\bar{a}_f = 0)$. The fraction of the posterior enclosed by the isoprobability contour that passes through the GR value (i.e., the GR quantile) for each event is shown in Table III. Figure 2 also shows the posteriors obtained by combining all the events that pass the stronger significance threshold FAR $< (1000\,{\rm yr})^{-1}$, as outlined in Sec. III (see the same section for a discussion of caveats).

The parameter estimation is performed employing uniform priors in component masses and spin magnitudes and isotropic priors in spin directions [14]. This will introduce a non-flat prior in the deviation parameters $\Delta M_f/\bar{M}_f$ and $\Delta a_f/\bar{a}_f$, which is shown as a thin, dashed contour in Fig. 2. Posteriors are estimated employing the precessing spin phenomenological waveform family IMRPhenomPv2. To assess the systematic errors due to imperfect waveform modeling, we also estimate the posteriors using the effective-one-body based waveform family SEOBNRv4 that models binary black holes with non-precessing spins. There is no qualitative difference between

[7] The frequency $f_c$ was determined using preliminary parameter inference results, so the values in Table III are slightly different than those that would be obtained using the posterior samples in GWTC-1 [9]. However, the test is robust against small changes in the cutoff frequency [38].

[8] As in [6], we average the $M_f, a_f$ posteriors obtained by different fits [76–78] after augmenting the fitting formulae for aligned-spin binaries by adding the contribution from in-plane spins [79]. However, unlike in [6, 79], we do not evolve the spins before applying the fits, due to technical reasons.

[9] For black hole binaries with comparable masses and moderate spins, as we consider here, the remnant black hole is expected to have $a_f \gtrsim 0.5$; see, e.g., [76–78] for fitting formulae derived from numerical simulations, or Table I for values of the remnant's spins obtained from GW events. Hence, $\Delta a_f/\bar{a}_f$ is expected to yield finite values.

the results derived using the two different waveforms families (see Sec. 2 of the Appendix).

We see additional peaks in the posteriors estimated from GW170814 and GW170823. Detailed follow-up investigations did not show any evidence of the presence of a coherent signal in multiple detectors that differs from the GR prediction. The second peak in GW170814 is introduced by the posterior of $M_{\rm f}^{\rm post\text{-}insp}$, while the extra peak in GW170823 is introduced by the posterior of $M_{\rm f}^{\rm insp}$. Injection studies in real data around the time of these events, using simulated GR waveforms with parameters consistent with GW170814 and GW170823, suggest that such secondary peaks occur for $\sim 10\%$ of injections. Features in the posteriors of GW170814 and GW170823 are thus consistent with expected noise fluctuations.

## VI. PARAMETERIZED TESTS OF GRAVITATIONAL WAVE GENERATION

A deviation from GR could manifest itself as a modification of the dynamics of two orbiting compact objects, and in particular, the evolution of the orbital (and hence, GW) phase. In an analytical waveform model like IMRPhenomPv2, the details of the GW phase evolution are controlled by coefficients that are either analytically calculated or determined by fits to numerical-relativity (NR) simulations, always under the assumption that GR is the underlying theory. These coefficients have a functional dependence on the intrinsic parameters of the binary that is specific to GR (masses and spins for binary black holes). In this section we investigate deviations from the GR binary dynamics by introducing shifts in each of the individual GW phase coefficients of IMRPhenomPv2. We then treat those shifts as additional unconstrained GR-violating parameters, which we measure in addition to the standard parameters describing the binary.

The early inspiral of compact binaries is well modeled by the post-Newtonian (PN) approximation [80–83] to GR, which is based on the expansion of the orbital quantities in terms of a small velocity parameter $v/c$. For a given set of intrinsic parameters, coefficients for the different orders in $v/c$ in the PN series are uniquely determined. A consistency test of GR using measurements of the inspiral PN phase coefficients was first proposed in [84–86], while a generalized parametrization was motivated in [87]. Bayesian implementations based on such parametrized methods were presented and tested in [39, 88–90] and were also extended to the post-inspiral part of the gravitational-wave signal [91, 92]. These ideas were applied to the first GW observation, GW150914 [10], yielding the first bounds on higher-order PN coefficients [4]. Since then, the constraints have been revised with the binary black hole events that followed, GW151226 in O1 [5] and GW170104 in O2 [6]. More recently, the first such constraints from a binary neutron star merger were placed with the detection of GW170817 [8]. Bounds on parametrized violations of GR from GW detections have been mapped, to leading order, to constraints on specific alternative theories of gravity (see, e.g., [93]). In this paper, we present individual constraints on parametrized deviations from GR for each of the GW sources in O1 and O2 listed in Table I, as well as the tightest combined constraints obtained to date by combining information from all the significant binary black holes events observed so far, as described in Sec. III.

The frequency-domain GW phase evolution $\Phi(f)$ in the early-inspiral stage of IMRPhenomPv2 is described by a PN expansion, augmented with higher-order phenomenological coefficients. The PN phase evolution is analytically expressed in closed form by employing the stationary phase approximation. The late-inspiral and post-inspiral (intermediate and merger-ringdown) stages are described by phenomenological analytical expressions. In all cases, the phenomenological coefficients are calibrated with data from NR simulations of mass-ratios as asymmetric as 1 : 18 and of dimensionless spin-magnitudes up to 0.99, as well as the inspiral portion of (uncalibrated) EOB waveforms. The transition frequency[10] from inspiral to intermediate regime is given by the condition $GMf/c^3 = 0.018$, with $M$ the total mass of the binary in the detector frame, since this is the lowest frequency above which this model was calibrated with NR data [19]. Deviations from GR in all three stages are expressed by means of relative shifts $\delta\hat{p}_i$ in the corresponding waveform coefficients: $p_i \to (1 + \delta\hat{p}_i)\, p_i$, which are used as additional free parameters in our extended waveform models.

We denote the testing parameters corresponding to PN phase coefficients by $\delta\hat{\varphi}_i$, where $i$ indicates the power of $v/c$ beyond leading (Newtonian or 0PN) order in $\Phi(f)$. The frequency dependence of the corresponding phase term is $f^{(i-5)/3}$. In the parametrized model, $i$ varies from 0 to 7, including the terms with logarithmic dependence at 2.5PN and 3PN. The non-logarithmic term at 2.5PN (i.e., $i = 5$) cannot be constrained, because of its degeneracy with a constant reference phase (e.g., the phase at coalescence). These coefficients were introduced in their current form in Eq. (19) of [88]. In addition, we also test for $i = -2$, representing an effective $-1$PN term, which is motivated below. The full set of inspiral parameters are thus

$$\{\delta\hat{\varphi}_{-2}, \delta\hat{\varphi}_0, \delta\hat{\varphi}_1, \delta\hat{\varphi}_2, \delta\hat{\varphi}_3, \delta\hat{\varphi}_4, \delta\hat{\varphi}_{5l}, \delta\hat{\varphi}_6, \delta\hat{\varphi}_{6l}, \delta\hat{\varphi}_7\}.$$

Since the $-1$PN term and the 0.5PN term are absent in the GR phasing, we parametrize $\delta\hat{\varphi}_{-2}$ and $\delta\hat{\varphi}_1$ as *absolute* deviations, with a pre-factor equal to the 0PN coefficient.

The $-1$PN term of $\delta\hat{\varphi}_2$ can be interpreted as arising from the emission of dipolar radiation. For binary black holes, this could occur in, e.g., alternative theories of gravity where an additional scalar charge is sourced by terms related to curvature [94, 95]. At leading order, this introduces a deviation in the $-1$PN coefficient of the waveform [96, 97]. This effectively introduces a term in the inspiral GW phase, varying with frequency as $f^{-7/3}$, while the gravitational flux is modified as $\mathcal{F}_{\rm GR} \to \mathcal{F}_{\rm GR}(1 + Bc^2/v^2)$. The first bound on $\delta\hat{\varphi}_{-2}$ was published in [8]. The higher-order terms in the above expansion also lead to a modification in the higher-order PN coefficients. Unlike the case of GW170817 (which we study separately in [8]), where the higher-order terms in the expansion of the

[10] This frequency is different than the cutoff frequency used in the inspiral-merger-ringdown consistency test, as was briefly mentioned in Sec. III.

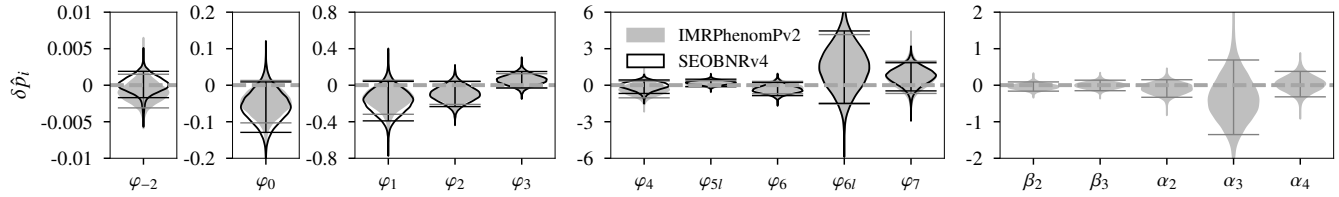


FIG. 3. Combined posteriors for parametrized violations of GR, obtained from all events in Table I with a significance of FAR $< (1000\ \mathrm{yr})^{-1}$ in both modeled searches. The combined posteriors on $\varphi_i$ in the inspiral regime are obtained from the events which in addition exceed the SNR threshold in the inspiral regime (GW150914, GW151226, GW170104, GW170608, and GW170814), analyzed with IMRPhenomPv2 (grey shaded region) and SEOBNRv4 (black outline). The combined posteriors on the intermediate and merger-ringdown parameters $\beta_i$ and $\alpha_i$ are obtained from events which exceed the SNR threshold in the post-inspiral regime (GW150914, GW170104, GW170608, GW170809, GW170814, and GW170823), analyzed with IMRPhenomPv2.

flux are negligible, the contribution of higher-order terms can be significant in the binary black-hole signals that we study here. This prohibits an exact interpretation of the −1PN term as the strength of dipolar radiation. Hence, this analysis only serves as a test of the presence of an effective −1PN term in the inspiral phasing, which is absent in GR.

To measure the above GR violations in the post-Newtonian inspiral, we employ two waveform models: (i) the analytical frequency-domain model IMRPhenomPv2 which also provided the natural parametrization for our tests and (ii) SEOBNRv4, which we use in the form of SEOBNRv4_ROM, a frequency-domain, reduced-order-model of the SEOBNRv4 model. The inspiral part of SEOBNRv4 is based on a numerical evolution of the aligned-spin effective-one-body dynamics of the binary, while its post-inspiral evolution is calibrated against NR simulations. Despite its non-analytical nature, SEOBNRv4_ROM can also be used to test the parametrized modifications of the early inspiral defined above. Using the method presented in [8], we add deviations to the waveform phase corresponding to a given $\delta\hat{\varphi}_i$ at low frequencies and then taper the corrections to zero at a frequency consistent with the transition frequency between early-inspiral and intermediate phases used by IMRPhenomPv2. The same procedure cannot be applied to the later stages of the waveform, thus the analysis performed with SEOBNRv4 is restricted to the post-Newtonian inspiral, cf. Fig. 3.

The analytical descriptions of the intermediate and merger-ringdown stages in the IMRPhenomPv2 model allow for a straightforward way of parametrizing deviations from GR, denoted by $\{\delta\hat{\beta}_2, \delta\hat{\beta}_3\}$ and $\{\delta\hat{\alpha}_2, \delta\hat{\alpha}_3, \delta\hat{\alpha}_4\}$ respectively, following [92]. Here the parameters $\delta\hat{\beta}_i$ correspond to deviations from the NR-calibrated phenomenological coefficients $\beta_i$ of the intermediate stage, while the parameters $\delta\hat{\alpha}_i$ refer to modifications of the merger-ringdown coefficients $\alpha_i$ obtained from a combination of phenomenological fits and analytical black-hole perturbation theory calculations [19].

Using LALInference, we calculate posterior distributions of the parameters characterizing the waveform (including those that describe the binary in GR). Our parametrization recovers GR at $\delta\hat{p}_i = 0$, so consistency with GR is verified if the posteriors of $\delta\hat{p}_i$ have support at zero. We perform the analyses by varying one $\delta\hat{p}_i$ at a time; as shown in Ref. [98], this is fully robust to detecting deviations present in multiple PN-orders. In addition, allowing for a larger parameter space by varying multiple coefficients simultaneously would not improve our efficiency in identifying violations of GR, as it would yield less informative posteriors. A specific alternative theory of gravity would likely yield correlated deviations in many parameters, including modifications that we have not considered here. This would be the target of an exact comparison of an alternative theory with GR, which would only be possible if a complete, accurate description of the inspiral-merger-ringdown signal in that theory was available.

We use priors uniform on $\delta\hat{p}_i$ and symmetric around zero. Figure 3 shows the combined posteriors of $\delta\hat{p}_i$ (marginalized over all other parameters) estimated from the combination of all the events that cross the significance threshold of FAR $< (1000\ \mathrm{yr})^{-1}$ in both modeled searches; see Table I. Events with SNR$< 6$ in the inspiral regime (parameters $\delta\hat{\varphi}_i$) or in the post-inspiral regime ($\delta\hat{\beta}_i$ and $\delta\hat{\alpha}_i$ for the intermediate and merger-ringdown parameters respectively) are not included in the results, since the data from those instances failed to provide useful constraints (see Sec. III for more details). This SNR threshold, however, is not equally effective in ensuring informative results for all cases; see Sec. 3 in the Appendix for a detailed discussion. In all cases considered, the posteriors are consistent with $\delta\hat{p}_i = 0$ within statistical fluctuations. Bounds on the inspiral coefficients obtained with the two different waveform models are found to be in good agreement with each other. Finally, we note that the event-combining analyses on $\delta\hat{p}_i$ incorporate the highly non-trivial assumption that these parametrized violations are constant across all events considered. While this is valid for our null hypothesis (GR), in an active search for signs of modified gravity this assumption should be dropped. Posterior distributions of $\delta\hat{p}_i$ for the individual-event analysis, also showing full consistency with GR, are provided in Sec. 3 of the Appendix.

Figure 4 shows the 90% upper bounds on $\delta\hat{\varphi}_i$ for all the individual events which cross the SNR threshold (SNR $> 6$) in the inspiral regime (the most massive of which is GW150914). The bounds from the combined posteriors are also shown; these include the events which exceed both the SNR threshold in the inspiral regime as well as the significance threshold, namely GW150914, GW151226, GW170104, GW170608, and GW170814. The bound from the likely lightest mass binary black hole event GW170608 at 1.5PN is currently the strongest

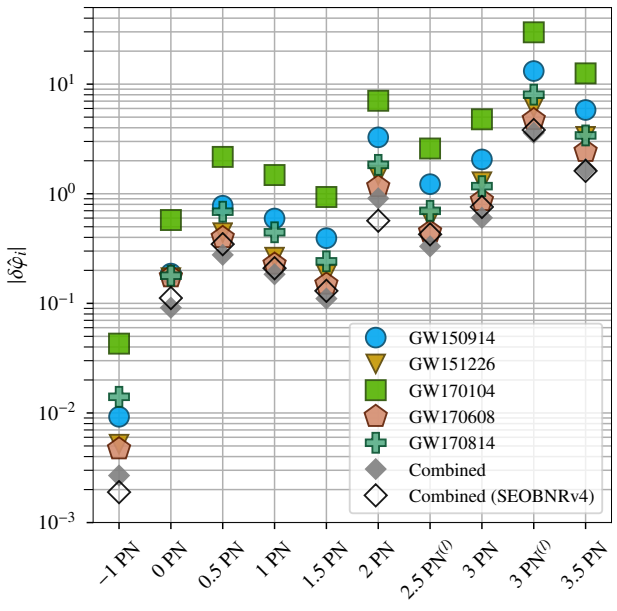


FIG. 4. 90% upper bounds on the absolute magnitude of the GR-violating parameters $\delta\hat{\varphi}_n$, from −1PN through 3.5PN in the inspiral phase. At each PN order, we show results obtained from each of the events listed in Table I that cross the SNR threshold in the inspiral regime, analyzed with IMRPhenomPv2. Bounds obtained from combining posteriors of events detected with a significance that exceeds a threshold of FAR $< (1000\ \mathrm{yr})^{-1}$ in both modelled searches are shown for both analyses, using IMRPhenomPv2 (filled diamonds) and SEOBNRv4 (empty diamonds).

constraint obtained on a positive PN coefficient from a single binary black hole event, as shown in Fig. 4. However, the constraint at this order is about five times worse than that obtained from the binary neutron star event GW170817 alone [8]. The −1PN bound is two orders of magnitude better for GW170817 than the best bound obtained from GW170608. For all other PN orders, GW170608 also provides the best bounds, which at high PN orders are of the same order of magnitude as the ones from GW170817. Our results can be compared statistically to those obtained by performing the same tests on simulated GR and non-GR waveforms given in [92]. The results presented here are consistent with those of GR waveforms injected into realistic detector data. The combined bounds are the tightest obtained so far, improving on the bounds obtained in [5] by factors between 1.1 and 1.8.

## VII. PARAMETERIZED TESTS OF GRAVITATIONAL WAVE PROPAGATION

We now place constraints on a phenomenological modification of the GW dispersion relation, i.e., on a possible frequency dependence of the speed of GWs. This modification, introduced in [99] and first applied to LIGO data in [6], is obtained by adding a power-law term in the momentum to the dispersion relation $E^2 = p^2c^2$ of GWs in GR, giving

$$E^2 = p^2c^2 + A_\alpha p^\alpha c^\alpha. \tag{2}$$

Here, $c$ is the speed of light, $E$ and $p$ are the energy and momentum of the GWs, and $A_\alpha$ and $\alpha$ are phenomenological parameters. We consider $\alpha$ values from 0 to 4 in steps of 0.5. However, we exclude $\alpha = 2$, where the speed of the GWs is modified in a frequency-independent manner, and therefore gives no observable dephasing.[11] Thus, in all cases except for $\alpha = 0$, we are considering a Lorentz-violating dispersion relation. The group velocity associated with this dispersion relation is $v_g/c = (dE/dp)/c = 1 + (\alpha-1)A_\alpha E^{\alpha-2}/2 + O(A_\alpha^2)$. The associated length scale is $\lambda_A := hc|A_\alpha|^{1/(\alpha-2)}$, where $h$ is Planck's constant. $\lambda_A$ gives the scale of modifications to the Newtonian potential (the Yukawa potential for $\alpha = 0$) associated with this dispersion relation.

While Eq. (2) is a purely phenomenological model, it encompasses a variety of more fundamental predictions (at least to leading order) [93, 99]. In particular, $A_0 > 0$ corresponds to a massive graviton, i.e., the same dispersion as for a massive particle in vacuo [101], with a graviton mass given by $m_g = A_0^{1/2}/c^2$.[12] Furthermore, $\alpha$ values of 2.5, 3, and 4 correspond to the leading predictions of multi-fractal spacetime [102]; doubly special relativity [103]; and Hořava-Lifshitz [104] and extra dimensional [105] theories, respectively. The standard model extension also gives a leading contribution with $\alpha = 4$ [106], only considering the non-birefringent terms; our analysis does not allow for birefringence.

In order to obtain a waveform model with which to constrain these propagation effects, we start by assuming that the waveform extracted in the binary's local wave zone (i.e., near to the binary compared to the distance from the binary to Earth, but far from the binary compared to its own size) is well-described by a waveform in GR.[13] Since we are able to bound these propagation effects to be very small, we can work to linear order in $A_\alpha$ when computing the effects of this dispersion on the frequency-domain GW phasing,[14] thus obtaining a correction [99] that is added to $\Phi(f)$ in Eq. (1):

$$\delta\Phi_\alpha(f) = \mathrm{sign}(A_\alpha) \begin{cases} \dfrac{\pi D_\mathrm{L}}{\alpha-1}\lambda_{A,\mathrm{eff}}^{\alpha-2}\left(\dfrac{f}{c}\right)^{\alpha-1}, & \alpha \neq 1 \\ \dfrac{\pi D_\mathrm{L}}{\lambda_{A,\mathrm{eff}}}\ln\left(\dfrac{\pi G\mathcal{M}^\mathrm{det} f}{c^3}\right), & \alpha = 1 \end{cases}. \tag{3}$$

[11] For a source with an electromagnetic counterpart, $A_2$ can be constrained by comparison with the arrival time of the photons, as was done with GW170817/GRB170817A [100].

[12] Thus, the Yukawa screening length is $\lambda_0 = h/(m_g c)$.

[13] This is likely to be a good assumption for $\alpha < 2$, where we constrain $\lambda_A$ to be much larger than the size of the binary. For $\alpha > 2$, where we constrain $\lambda_A$ to be much *smaller* than the size of the binary, one has to posit a screening mechanism in order to be able to assume that the waveform in the binary's local wave zone is well-described by GR, as well as for this model to evade Solar System constraints.

[14] The dimensionless parameter controlling the size of the linear correction is $A_\alpha f^{\alpha-2}$, which is $\lesssim 10^{-18}$ at the 90% credible level for the events we consider and frequencies up to 1 kHz.

Here, $D_{\rm L}$ is the binary's luminosity distance, $\mathcal{M}^{\rm det}$ is the binary's detector-frame (i.e., redshifted) chirp mass, and $\lambda_{A,\rm eff}$ is the effective wavelength parameter used in the sampling, defined as

$$\lambda_{A,\rm eff} := \left[\frac{(1+z)^{1-\alpha} D_{\rm L}}{D_\alpha}\right]^{1/(\alpha-2)} \lambda_A \,. \tag{4}$$

The parameter $z$ is the binary's redshift, and $D_\alpha$ is a distance parameter given by

$$D_\alpha = \frac{(1+z)^{1-\alpha}}{H_0} \int_0^z \frac{(1+\bar{z})^{\alpha-2}}{\sqrt{\Omega_{\rm m}(1+\bar{z})^3 + \Omega_\Lambda}} d\bar{z} \,, \tag{5}$$

where $H_0 = 67.90$ km s$^{-1}$ Mpc$^{-1}$ is the Hubble constant, and $\Omega_{\rm m} = 0.3065$ and $\Omega_\Lambda = 0.6935$ are the matter and dark energy density parameters; these are the TT+lowP+lensing+ext values from [107].[15]

The dephasing in Eq. (3) is obtained by treating the gravitational wave as a stream of particles (gravitons), which travel at the particle velocity $v_p/c = pc/E = 1 - A_\alpha E^{\alpha-2}/2 + O(A_\alpha^2)$. There are suggestions to use the particle velocity when considering doubly special relativity, though there are also suggestions to use the group velocity $v_g$ in that case (see, e.g., [109] and references therein for both arguments). However, the group velocity is appropriate for, e.g., multi-fractal spacetime theories (see, e.g., [110]). To convert the bounds presented here to the case where the particles travel at the group velocity, scale the $A_\alpha$ bounds for $\alpha \neq 1$ by factors of $1/(1-\alpha)$. The group velocity calculation gives an unobservable constant phase shift for $\alpha = 1$.

We consider the cases of positive and negative $A_\alpha$ separately, and obtain the results shown in Table IV and Fig. 5 when applying this analysis to the GW events under consideration. While we sample with a flat prior in $\log \lambda_{A,\rm eff}$, our bounds are given using priors flat in $A_\alpha$ for all results except for the mass of the graviton, where we use a prior flat in the graviton mass. We also show the results from combining together all the signals that satisfy our selection criterion. We are able to combine together the results from different signals with no ambiguity, since the known distance dependence is accounted for in the waveforms.

Figure 6 displays the full $A_\alpha$ posteriors obtained by combining all selected events (using IMRPhenomPv2 waveforms). To obtain the full $A_\alpha$ posteriors, we combine together the positive and negative $A_\alpha$ results for individual events by weighting by their Bayesian evidences; we then combine the posteriors from individual events. We give the analogous plots for the individual events in Sec. 4 of the Appendix. The combined positive and negative $A_\alpha$ posteriors are also used to compute the GR quantiles given in Table IV, which give the probability to have

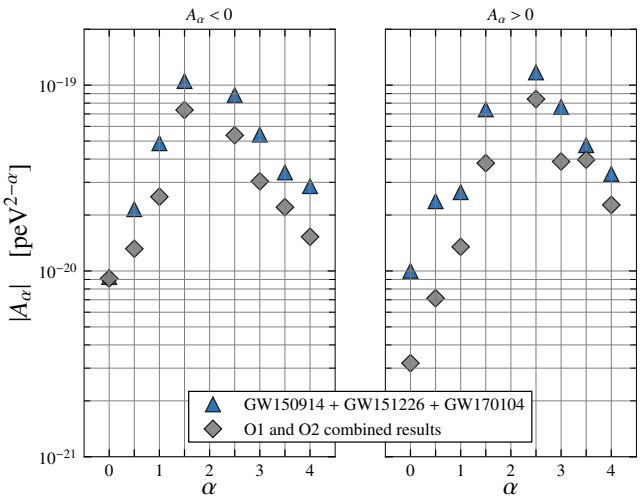


FIG. 5. 90% credible upper bounds on the absolute value of the modified dispersion relation parameter $A_\alpha$. We show results for positive and negative values of $A_\alpha$ separately. Specifically, we give the updated versions of the results from combining together GW150914, GW151226, and GW170104 (first given in [6]), as well as the results from combining together all the events meeting our significance threshold for combined results (see Table I). Picoelectronvolts (peV) provide a convenient scale, because 1 peV $\simeq h \times 250$ Hz, where 250 Hz is roughly around the most sensitive frequencies of the LIGO and Virgo instruments.

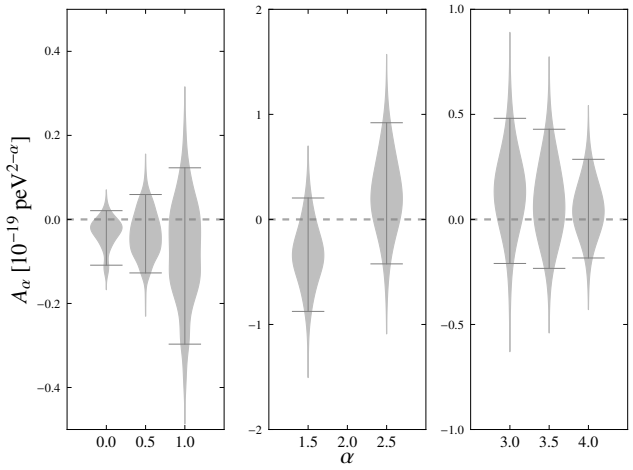


FIG. 6. Violin plots of the full posteriors on the modified dispersion relation parameter $A_\alpha$ calculated from the combined events, with the 90% credible interval around the median indicated.

$A_\alpha < 0$, where $A_\alpha = 0$ is the GR value. Thus, large or small values of the GR quantile indicate that the distribution is not peaked close to the GR value. For a GR signal, the GR quantile will be distributed uniformly in $[0, 1]$ for different noise realizations. The GR quantiles we find are consistent with such a uniform distribution. In particular, the (two-tailed) meta $p$-value for all events and $\alpha$ values obtained using Fisher's method [73] (as in Sec. V A) is 0.9995.

We find that the combined bounds overall improve on those

[15] We use these values for consistency with the results presented in [14]. If we instead use the more recent results from [108], specifically the TT,TE,EE+lowE+lensing+BAO values used for comparison in [14], then there are very minor changes to the results presented in this section. For instance, the upper bounds in Table IV change by at most $\sim 0.1\%$.

TABLE IV. 90% credible level upper bounds on the graviton mass $m_g$ and the absolute value of the modified dispersion relation parameter $A_\alpha$, as well as the GR quantiles $Q_{\rm GR}$. The < and > labels denote the bounds for $A_\alpha < 0$ and $> 0$, respectively, and we have defined the dimensionless quantity $\bar{A}_\alpha := A_\alpha/\mathrm{eV}^{2-\alpha}$. Events with names in boldface are used to obtain the combined results.

| Event | $m_g$ | $\lvert\bar{A}_0\rvert$ | | | $\lvert\bar{A}_{0.5}\rvert$ | | | $\lvert\bar{A}_1\rvert$ | | | $\lvert\bar{A}_{1.5}\rvert$ | | | $\lvert\bar{A}_{2.5}\rvert$ | | | $\lvert\bar{A}_3\rvert$ | | | $\lvert\bar{A}_{3.5}\rvert$ | | | $\lvert\bar{A}_4\rvert$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $[10^{-23}$ eV$/c^2]$ | < $[10^{-44}]$ | > | $Q_{\rm GR}$ [%] | < $[10^{-38}]$ | > | $Q_{\rm GR}$ [%] | < $[10^{-32}]$ | > | $Q_{\rm GR}$ [%] | < $[10^{-25}]$ | > | $Q_{\rm GR}$ [%] | < $[10^{-13}]$ | > | $Q_{\rm GR}$ [%] | < $[10^{-8}]$ | > | $Q_{\rm GR}$ [%] | < $[10^{-2}]$ | > | $Q_{\rm GR}$ [%] | < $[10^{4}]$ | > | $Q_{\rm GR}$ [%] |
| **GW150914** | 10 | 1.4 | 1.1 | 71 | 6.1 | 5.1 | 57 | 5.6 | 3.6 | 74 | 3.3 | 2.2 | 74 | 2.4 | 2.2 | 50 | 17 | 19 | 36 | 11 | 20 | 42 | 8.0 | 10 | 52 |
| GW151012 | 17 | 3.9 | 3.5 | 26 | 3.7 | 11 | 27 | 6.7 | 9.8 | 36 | 2.0 | 2.5 | 58 | 2.9 | 1.6 | 58 | 22 | 10 | 68 | 19 | 15 | 58 | 15 | 6.9 | 61 |
| **GW151226** | 29 | 7.3 | 9.6 | 21 | 8.3 | 22 | 10 | 14 | 13 | 26 | 3.5 | 3.4 | 28 | 3.5 | 2.2 | 68 | 23 | 10 | 60 | 15 | 6.2 | 58 | 20 | 5.5 | 72 |
| **GW170104** | 9.4 | 2.6 | 1.0 | 63 | 4.4 | 2.5 | 72 | 7.4 | 3.1 | 76 | 2.0 | 0.85 | 83 | 1.2 | 2.9 | 23 | 11 | 14 | 35 | 6.8 | 11 | 37 | 6.9 | 9.3 | 43 |
| **GW170608** | 30 | 14 | 9.3 | 49 | 22 | 8.9 | 68 | 16 | 28 | 49 | 3.2 | 3.8 | 64 | 3.3 | 2.1 | 49 | 9.9 | 8.4 | 46 | 88 | 8.1 | 46 | 30 | 3.8 | 38 |
| GW170729 | 7.6 | 0.30 | 0.67 | 17 | 0.96 | 1.2 | 26 | 2.2 | 4.7 | 16 | 0.82 | 1.6 | 18 | 4.4 | 1.5 | 94 | 38 | 9.1 | 96 | 28 | 9.5 | 94 | 46 | 7.6 | 95 |
| **GW170809** | 9.6 | 1.4 | 1.3 | 64 | 2.6 | 2.7 | 49 | 11 | 6.1 | 49 | 1.5 | 2.4 | 37 | 4.0 | 1.4 | 79 | 24 | 7.3 | 81 | 19 | 6.7 | 78 | 14 | 7.0 | 82 |
| **GW170814** | 8.8 | 4.3 | 1.2 | 94 | 5.5 | 1.9 | 92 | 15 | 5.5 | 93 | 3.1 | 1.3 | 96 | 1.4 | 4.2 | 5.7 | 7.9 | 27 | 6.2 | 5.8 | 31 | 7.5 | 4.5 | 14 | 9.6 |
| GW170818 | 7.4 | 1.4 | 0.69 | 74 | 2.7 | 2.0 | 80 | 5.1 | 4.5 | 73 | 1.8 | 0.71 | 79 | 1.5 | 3.6 | 28 | 18 | 16 | 41 | 20 | 8.4 | 73 | 13 | 9.7 | 49 |
| **GW170823** | 6.4 | 1.3 | 0.51 | 61 | 1.2 | 1.7 | 51 | 2.8 | 2.4 | 49 | 1.0 | 1.3 | 46 | 2.6 | 1.4 | 52 | 11 | 17 | 41 | 9.4 | 12 | 37 | 15 | 12 | 46 |
| **Combined** | 5.0 | 0.91 | 0.32 | 80 | 1.3 | 0.71 | 71 | 2.5 | 1.3 | 72 | 0.73 | 0.38 | 85 | 0.54 | 0.84 | 27 | 3.0 | 3.9 | 25 | 2.2 | 4.0 | 33 | 1.5 | 2.3 | 38 |

quoted in [6] by roughly the factor of $\sqrt{7/3} \simeq 1.5$ expected from including more events, with the bounds for some quantities improving by up to a factor of 2.4, due to the inclusion of several more massive and distant systems in the sample. These massive and distant systems, notably GW170823 (and GW170729, which is not included in the combined results), generally give the best individual bounds, particularly for small values of $\alpha$, where the dephasing is largest at lower frequencies. Closer and less massive systems such as GW151226 and GW170608 provide weaker bounds, overall. However, their bounds can be comparable to those of the more massive, distant events for larger values of $\alpha$. The lighter systems have more power at higher frequencies where the dephasing from the modified dispersion is larger for larger values of $\alpha$.

The new combined bound on the mass of the graviton of $m_g \leq 5.0 \times 10^{-23}$ eV$/c^2$ is a factor of 1.5 improvement on the one presented in [6]. It is also a small improvement on the bound of $m_g \leq 6.76 \times 10^{-23}$ eV$/c^2$ (90% confidence level) obtained from Solar System ephemerides in [111].[16] However, these bounds are complementary, since the GW bound comes from the radiative sector, while the Solar System bound considers the static modification to the Newtonian potential. See, e.g., [113] for a review of bounds on the mass of the graviton.

We find that the posterior on $A_\alpha$ peaks away from 0 in some cases (illustrated in Sec. 4 of the Appendix), and the GR quantile is in one of the tails of the distribution. This feature is expected for a few out of 10 events, simply from Gaussian noise fluctuations. We have performed simulations of 100 GR sources with source-frame component masses lying between 25 and 45 $M_\odot$, isotropically distributed spins with dimensionless magnitudes up to 0.99, and at luminosity distances between 500 and 800 Mpc. These simulations used the waveform model IMRPhenomPv2 and considered the Advanced LIGO and Virgo network, using Gaussian noise with the detectors' design sensitivity power spectral densities. We found that in about 20 – 30% of cases, the GR quantile lies in the tails of the distribution (i.e., < 10% or > 90%), when the sources injected are analyzed using the same waveform model (IMRPhenomPv2).

In order to assess the impact of waveform systematics, we also analyze some events using the aligned-spin SEOBNRv4 model. We consider GW170729 and GW170814 in depth in this study because the GR quantiles of the IMRPhenomPv2 results lie in the tails of the distributions, and find that the 90% upper bounds and GR quantiles presented in Table IV differ by at most a factor of 2.3 for GW170729 and 1.5 for GW170814 when computed using the SEOBNRv4 model. These results are presented in Sec. 4 of the Appendix.

There are also uncertainties in the determination of the 90% bounds due to the finite number of samples and the long tails of the distributions. As in Ref. [6], we quantify this uncertainty using Bayesian bootstrapping [114]. We use 1000 bootstrap realizations for each value of $\alpha$ and sign of $A_\alpha$, obtaining a distribution of 90% bounds on $A_\alpha$. We consider the 90% credible interval of this distribution and find that its width is < 30% of the values for the 90% bounds on $A_\alpha$ given in Table IV for all but 10 of the 160 cases we consider (counting positive and negative $A_\alpha$ cases separately). For GW170608, $A_4 < 0$, the width of the 90% credible interval from bootstrapping is 91% of the value in Table IV. This ratio is ≤ 47% for all the remaining cases. Thus, there are a few cases where the bootstrapping uncertainty in the bound on $A_\alpha$ is large, but for most cases, this is not a substantial uncertainty.

## VIII. POLARIZATIONS

Generic metric theories of gravity may allow up to six polarizations of gravitational waves [115]: two tensor modes (helicity ±2), two vector modes (helicity ±1), and two scalar modes (helicity 0). Of these, only the two tensor modes (+ and ×) are permitted in GR. We may attempt to reconstruct the polarization content of a passing GW using a network of

[16] The much stronger bound in [112] is deduced from a post-fit analysis (i.e., using the residuals of a fit to Solar System ephemerides performed without including the effects of a massive graviton). It may therefore overestimate Solar System constraints, as is indeed seen to be the case in [111].

detectors [1, 116–119]. This is possible because instruments with different orientations will respond differently to signals from a given sky location depending on their polarization. In particular, the strain signal in detector $I$ can be written as $h_I(t) = \sum_A F^A{}_I h_A(t)$, with $F^A{}_I$ the detector's response function and $h_A(t)$ the $A$-polarized part of the signal [1, 120].

In order to fully disentangle the polarization content of a transient signal, at least 5 detectors are needed to break all degeneracies [116].[17] This limits the polarization measurements that are currently feasible. In spite of this, we may extract some polarization information from signals detected with both LIGO detectors and Virgo [119]. This was done previously with GW170814 and GW170817 to provide evidence that GWs are tensor polarized, instead of fully vector or fully scalar [7, 8]. Besides GW170814, there are three binary black hole events that were detected with the full network (GW170729, GW170809, and GW170818). Of these events, only GW170818 has enough SNR and is sufficiently well localized to provide any relevant information (cf. Fig. 8 in [14]). The Bayes factors (marginalized likelihood ratios) obtained in this case are $12 \pm 3$ for tensor vs vector and $407 \pm 100$ for tensor vs scalar, where the error corresponds to the uncertainty due to discrete sampling in the evidence computations. These values are comparable to those from GW170814, for which the latest recalibrated and cleaned data (cf. Sec. II) yield Bayes factors of $30 \pm 4$ and $220 \pm 27$ for tensor vs vector and scalar respectively.[18] Values from these binary black holes are many orders of magnitude weaker than those obtained from GW170817, where we benefited from the precise sky-localization provided by an electromagnetic counterpart [8].

## IX. CONCLUSIONS AND OUTLOOK

We have presented the results from various tests of GR performed using the binary black hole signals from the catalog GWTC-1 [9], i.e., those observed by Advanced LIGO and Advanced Virgo during the first two observing runs of the advanced detector era. These tests, which are among the first tests of GR in the highly relativistic, nonlinear regime of strong gravity, do not reveal any inconsistency of our data with the predictions of GR. We have presented full results on four tests of the consistency of the data with gravitational waveforms from binary black hole systems as predicted by GR. The first two of these tests check the self-consistency of our analysis. One checks that the residual remaining after subtracting the best-fit waveform is consistent with detector noise. The other checks that the final mass and spin inferred from the low- and high-frequency parts of the signal are consistent. The third and fourth tests introduce parameterized deviations in the waveform model and check that these deviations are consistent with their GR value of zero. In one test, these deviations are completely phenomenological modifications of the coefficients in a waveform model, including the post-Newtonian coefficients. In the other test, the deviations are those arising from the propagation of GWs with a modified dispersion relation, which includes the dispersion due to a massive graviton as a special case. In addition, we also check whether the observed polarizations are consistent with being purely tensor modes (as expected in GR) as opposed to purely scalar modes or vector modes.

We present results from all binary black hole events that are detected with a false alarm rate better than $(1\,\mathrm{yr})^{-1}$. This includes results from the re-analysis of some of the events which were published earlier [4–7], with better calibration and data quality. Assuming that the parameters that describe deviations from GR take values that are independent of source properties, we can combine results from multiple events. We choose to combine results only from highly significant events, detected with a false alarm rate better than $(1000\,\mathrm{yr})^{-1}$ in both modeled searches. Combining together these results has allowed us to significantly reduce the statistical errors on constraints on deviations from GR predictions, as compared to those from individual events. The combined constraints presented here improve our previously presented constraints by factors of 1.1 to 2.4, with the largest improvements obtained for certain cases of the modified dispersion test. Notable constraints include that on the graviton's mass $m_g \leq 5.0 \times 10^{-23}\ \mathrm{eV}/c^2$ (an improvement of a factor of 1.5 over previously presented constraints) and the first constraint on the $-1$PN coefficient obtained from binary black holes.

With the expected observations of additional binary black hole merger events in the upcoming LIGO/Virgo observing runs [14, 121], the statistical errors of the combined results will soon decrease significantly. A number of potential sources of systematic errors (due to imperfect modeling of GR waveforms, calibration uncertainties, noise artifacts, etc.) need to be understood for future high-precision tests of strong gravity using GW observations. However, work to improve the analysis on all these fronts is well underway, for instance the inclusion of full spin-precession dynamics [21, 59–61], non-quadrupolar modes [61, 64–66], and eccentricity [47–52] in waveform models, as well as analyses that compare directly with numerical relativity waveforms [122, 123]. Additionally, a new, more flexible parameter estimation infrastructure is currently being developed [124], and this will allow for improvements in, e.g., the treatment of calibration uncertainties or PSD estimation to be incorporated more easily. We thus expect that tests of general relativity using the data from upcoming observing runs will be able to take full advantage of the increased sensitivity of the detectors.

[17] Differential-arm detectors are only sensitive to the traceless scalar mode, meaning we can only hope to distinguish five, not six, polarizations.

[18] These values are less stringent than those previously published in [7]. This is solely due to the change in data, which impacted the sky locations inferred under the non-GR hypotheses.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the United States National Science Foundation (NSF) for the construction and operation of the LIGO Laboratory and Advanced LIGO as well as the Science and Technology Facilities Coun-

cil (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. The authors gratefully acknowledge the Italian Istituto Nazionale di Fisica Nucleare (INFN), the French Centre National de la Recherche Scientifique (CNRS) and the Foundation for Fundamental Research on Matter supported by the Netherlands Organisation for Scientific Research, for the construction and operation of the Virgo detector and the creation and support of the EGO consortium. The authors also gratefully acknowledge research support from these agencies as well as by the Council of Scientific and Industrial Research of India, the Department of Science and Technology, India, the Science & Engineering Research Board (SERB), India, the Ministry of Human Resource Development, India, the Spanish Agencia Estatal de Investigación, the Vicepresidència i Conselleria d'Innovació, Recerca i Turisme and the Conselleria d'Educació i Universitat del Govern de les Illes Balears, the Conselleria d'Educació, Investigació, Cultura i Esport de la Generalitat Valenciana, the National Science Centre of Poland, the Swiss National Science Foundation (SNSF), the Russian Foundation for Basic Research, the Russian Science Foundation, the European Commission, the European Regional Development Funds (ERDF), the Royal Society, the Scottish Funding Council, the Scottish Universities Physics Alliance, the Hungarian Scientific Research Fund (OTKA), the Lyon Institute of Origins (LIO), the Paris Île-de-France Region, the National Research, Development and Innovation Office Hungary (NKFIH), the National Research Foundation of Korea, Industry Canada and the Province of Ontario through the Ministry of Economic Development and Innovation, the Natural Science and Engineering Research Council Canada, the Canadian Institute for Advanced Research, the Brazilian Ministry of Science, Technology, Innovations, and Communications, the International Center for Theoretical Physics South American Institute for Fundamental Research (ICTP-SAIFR), the Research Grants Council of Hong Kong, the National Natural Science Foundation of China (NSFC), the Leverhulme Trust, the Research Corporation, the Ministry of Science and Technology (MOST), Taiwan and the Kavli Foundation. The authors gratefully acknowledge the support of the NSF, STFC, MPS, INFN, CNRS and the State of Niedersachsen/Germany for provision of computational resources. The authors would like to thank Clifford Will and Nicolás Yunes for useful discussions.

## Appendix: Individual results and systematics studies

In the main body of the paper, for most analyses, we present only the combined results from all events. Here we present the posteriors from various tests obtained from individual events. In addition, we offer a limited discussion on systematic errors in the analysis, due to the specific choice of a GR waveform approximant.

### 1. Residuals test

As mentioned in Sec. V A, the residuals test is sensitive to all kinds of mismatch between the best-fit GR-based waveform and the data. This is true whether the mismatch is due to actual deviations from GR or more mundane reasons, like physics missing from our waveform models (e.g., higher-order modes). Had we found compelling evidence of coherent power in the residuals that could not be explained by instrumental noise, further investigations would be required to determine its origin. However, given our null result, we can simply state that we find no evidence for shortcomings in the best-fit waveform, neither from deviations from GR nor modeling systematics.

As the sensitivity of the detectors improves, the issue of systematics will become increasingly more important. To address this, future versions of this test will be carried out by subtracting a best-fit waveform produced with more accurate GR-based models, including numerical relativity.

### 2. Inspiral-merger-ringdown consistency test

In order to gauge the systematic errors in the IMR consistency test results due to imperfect waveform modeling, we have also estimated the posteriors of the deviation parameters $\Delta M_{\rm f}/\bar{M}_{\rm f}$ and $\Delta a_{\rm f}/\bar{a}_{\rm f}$ using the effective-one-body based waveform family SEOBNRv4 that models binary black holes with non-precessing spins. This analysis uses the same priors as used in the main analysis presented in Sec.V B, except that spins are assumed to be aligned/antialigned with the orbital angular momentum of the binary. The resulting posteriors are presented in Fig. 7 and are broadly consistent with the posteriors using IMRPhenomPv2 presented in Fig. 2. The differences in the posteriors of some of the individual events are not surprising, due to the different assumptions on the spins. For all events, the GR value is recovered in the 90% credible region of the posteriors.

### 3. Parametrized tests of gravitational wave generation

Figures 8 and 9 report the parameterized tests of waveform deviations for the individual events, augmenting the results shown in Fig. 3. A statistical summary of the posterior PDFs, showing median and symmetric 90% credible level bounds for the measured parameters is given in Table V. Sources with low SNR in the inspiral regime yield uninformative posterior distributions on $\delta\hat{\varphi}_i$. These sources are the ones further away and with higher mass, which merge at lower frequencies. For instance, although GW170823 has a total mass close to that of GW150914, being much further away (and redshifted to lower frequencies) makes it a low-SNR event, leaving very little information content in the inspiral regime. The same holds true for GW170729, which has a larger mass. Conversely, low-mass events like GW170608, having a significantly larger SNR in the inspiral regime and many more cycles in the frequency band provide very strong constraints in the $\delta\hat{\varphi}_i$ parameters

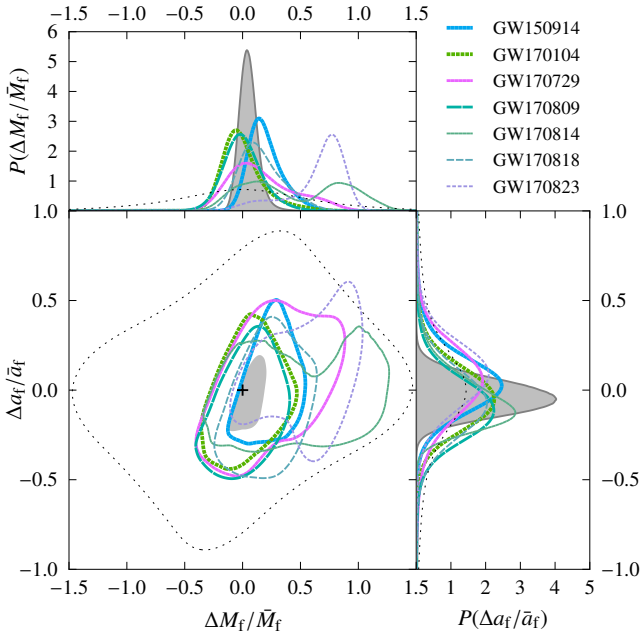


FIG. 7. Same as Fig. 2, except that the posteriors are computed using the nonprecessing-spin SEOBNRv4 waveforms.

(especially the low-order ones) while providing no useful constraints in the merger-ringdown parameters $\delta\hat{\alpha}_i$.

The choice of the SNR $> 6$ threshold explained in Sec. III ensures that most analyses are informative. However, this is not true in all cases, as not all parameters are as easily determined from the data (cf. the good constraints one obtains on the chirp mass with the much weaker constraints on the mass ratio). The two events for which the SNR threshold is insufficient are GW151012 and GW170608, where some post-inspiral parameters are largely unconstrained. The post-inspiral regime is itself divided into the intermediate and merger-ringdown regimes, and for both these events we find the intermediate regime parameters ($\delta\hat{\beta}_i$) to be informative; however, the merger–ringdown $\delta\hat{\alpha}_3$ for GW170608 and all $\delta\hat{\alpha}_i$ for GW151012 extend across the entire prior range considered in the analyses, far beyond the range constrained by other events (as can be seen in Fig. 9). Although we use the results from GW170608 $\delta\hat{\alpha}_i$ for combining posteriors in Fig. 3, the combined bounds remain unaffected by adding these results. For future tests, a more discerning threshold than a simple SNR cut, for example including information like the number of cycles of the signal in band, may be used to select which events will provide useful constraints.

Both here and in Sec. VI we report results on the parametrized deviations in the PN regime using two waveform models, IMRPhenomPv2 and SEOBNRv4. There is a subtle difference between the ways deviations from GR are introduced and parametrized in the two models. With IMRPhenomPv2, we directly constrain $\delta\hat{\varphi}_i$, which represent fractional deviations in the non-spinning portion of the $(i/2)$PN phase coefficients. The SEOBNRv4 analysis instead uses a parameterization that also applies the fractional deviations to spin contributions, as described in [8]. The results are then mapped post-hoc from this native parameterization to posteriors on $\delta\hat{\varphi}_i$, shown in Figs. 3 and 8 (black solid lines).

In the SEOBNRv4 analysis at 3.5PN, the native (spin-inclusive) posteriors contain tails that extend to the edge of the prior range. This is due to a zero-crossing of the 3.5PN term in the $(\eta, a_1, a_2)$ parameter space, which makes the corresponding relative deviation ill-defined. After the post-hoc mapping to posteriors on $\delta\hat{\varphi}_7$, no tails appear and we find good agreement with the IMRPhenomPv2 analysis, as expected. By varying the prior range, we estimate a systematic uncertainty of at most a few percent on the quoted 90% bounds due to the truncation of tails.

## 4. Parameterized tests of gravitational wave propagation

Posteriors on $A_\alpha$ for individual events are shown in Fig. 10, with data for positive and negative $A_\alpha$ combined into one violin plot. We provide results for all events with the IMRPhenomPv2 waveform model and also show results of the analysis with the SEOBNRv4 waveform model for GW170729 and GW170814. In Table VI we compare the 90% bounds on $A_\alpha$ and GR quantiles obtained with IMRPhenomPv2 and SEOBNRv4 for GW170729 and GW170814. We focus on these two events because the GR quantiles obtained with IMRPhenomPv2 lie in the tails of the distributions, and we find that this remains true for most $\alpha$ values in the analysis with SEOBNRv4. For GW170729 and $\alpha \in \{0, 0.5, 1\}$, the GR quantiles obtained using the two waveforms differ by factors of $\sim 2$; the two waveforms give values that are in much closer agreement for the other cases.

Additionally, for the GW151012 event and certain $\alpha$ values, a technical issue with our computation of the likelihood meant that specific points with relatively large values of $A_\alpha$ had to be manually removed from the posterior distribution. In particular, for computational efficiency, the likelihood is calculated on as short a segment of data as practical, with duration set by the longest waveform to be sampled. Large values of $A_\alpha$ yield highly dispersed waveforms that are pushed beyond the confines of the segment we use, causing the waveform templates to wrap around the boundaries. This invalidates the assumptions underlying our likelihood computation and causes an artificial enhancement of the SNR as reported by the analysis. As expected, recomputing the SNRs for these points on a segment that properly fits the waveform results in smaller values that are consistent with noise. Therefore, we exclude from our analysis parameter values yielding waveforms that would not be contained by the data segment used, which is equivalent to using a stricter prior on $A_\alpha$. Failure to do this may result in the appearance of outliers with spuriously high likelihood for large values of $A_\alpha$, as we have seen in our own analysis.

TABLE V. Median value and symmetric 90% credible level bounds of the waveform parameters $\delta\hat{p}_i$, as well as the GR quantiles $Q_{\rm GR}$. For the inspiral parameters, we show results in pairs of rows for when the data from individual events are analyzed using IMRPhenomPv2 (P) and SEOBNRv4 (S), while for the post-inspiral phenomenological parameters, results are obtained only for IMRPhenomPv2.

| Parameter | | Model | GW150914 | | GW151226 | | GW170104 | | GW170608 | | GW170814 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\bar{X}^+_-$ | $Q_{\rm GR}$ [%] | $\bar{X}^+_-$ | $Q_{\rm GR}$ [%] | $\bar{X}^+_-$ | $Q_{\rm GR}$ [%] | $\bar{X}^+_-$ | $Q_{\rm GR}$ [%] | $\bar{X}^+_-$ | $Q_{\rm GR}$ [%] |
| $\delta\hat{\phi}_{-2}$ | $[10^{-2}]$ | P | $-0.46^{+0.68}_{-0.59}$ | 88 | $0.14^{+0.52}_{-0.35}$ | 27 | $2.0^{+3.6}_{-1.9}$ | 3 | $-0.06^{+0.61}_{-0.35}$ | 60 | $-0.76^{+0.74}_{-0.81}$ | 96 |
| | | S | $-0.63^{+0.62}_{-0.61}$ | 95 | $0.10^{+0.40}_{-0.31}$ | 30 | $1.4^{+2.6}_{-1.4}$ | 5.9 | $0.10^{+0.27}_{-0.34}$ | 26 | $-0.99^{+0.89}_{-0.79}$ | 97 |
| $\delta\hat{\phi}_0$ | $[10^{-1}]$ | P | $-1.0^{+1.0}_{-1.1}$ | 95 | $-0.1^{+1.7}_{-1.7}$ | 53 | $2.5^{+4.3}_{-2.3}$ | 3.1 | $-0.5^{+1.4}_{-1.6}$ | 73 | $-0.8^{+1.2}_{-1.2}$ | 88 |
| | | S | $-1.1^{+1.0}_{-1.1}$ | 97 | $0.3^{+2.4}_{-2.8}$ | 41 | $1.8^{+3.5}_{-2.0}$ | 7.3 | $-0.9^{+1.8}_{-2.2}$ | 77 | $-1.1^{+1.5}_{-2.0}$ | 89 |
| $\delta\hat{\phi}_1$ | $[10^{0}]$ | P | $-0.39^{+0.40}_{-0.50}$ | 95 | $0.03^{+0.43}_{-0.44}$ | 45 | $1.0^{+1.7}_{-0.9}$ | 4.1 | $-0.14^{+0.33}_{-0.33}$ | 76 | $-0.31^{+0.47}_{-0.50}$ | 86 |
| | | S | $-0.51^{+0.41}_{-0.41}$ | 98 | $0.02^{+0.55}_{-0.56}$ | 48 | $0.7^{+1.2}_{-0.8}$ | 6.6 | $-0.16^{+0.38}_{-0.40}$ | 77 | $-0.27^{+0.57}_{-0.61}$ | 79 |
| $\delta\hat{\phi}_2$ | $[10^{0}]$ | P | $-0.35^{+0.32}_{-0.31}$ | 97 | $-0.01^{+0.29}_{-0.24}$ | 52 | $0.7^{+1.1}_{-0.6}$ | 3.5 | $-0.07^{+0.21}_{-0.20}$ | 72 | $-0.17^{+0.34}_{-0.36}$ | 80 |
| | | S | $-0.34^{+0.28}_{-0.30}$ | 97 | $0.02^{+0.33}_{-0.32}$ | 47 | $0.47^{+0.83}_{-0.56}$ | 8.6 | $-0.09^{+0.24}_{-0.25}$ | 75 | $-0.10^{+0.39}_{-0.38}$ | 67 |
| $\delta\hat{\phi}_3$ | $[10^{-1}]$ | P | $2.2^{+2.0}_{-1.9}$ | 3.4 | $-0.1^{+1.5}_{-2.0}$ | 54 | $-4.8^{+4.2}_{-6.5}$ | 97 | $0.5^{+1.2}_{-1.2}$ | 26 | $0.7^{+2.1}_{-2.2}$ | 30 |
| | | S | $2.2^{+1.8}_{-1.7}$ | 1.8 | $-0.2^{+2.7}_{-2.1}$ | 56 | $-3.1^{+3.7}_{-5.9}$ | 92 | $0.7^{+2.0}_{-1.6}$ | 25 | $0.2^{+2.4}_{-2.7}$ | 46 |
| $\delta\hat{\phi}_4$ | $[10^{0}]$ | P | $-1.9^{+1.7}_{-1.6}$ | 97 | $0.1^{+1.6}_{-1.3}$ | 47 | $3.7^{+4.6}_{-3.6}$ | 4.2 | $-0.3^{+1.1}_{-1.1}$ | 67 | $-0.5^{+1.7}_{-1.6}$ | 67 |
| | | S | $-0.2^{+0.7}_{-1.4}$ | 69 | $0.2^{+1.8}_{-1.9}$ | 41 | $2.2^{+4.1}_{-2.9}$ | 11 | $-0.4^{+1.3}_{-1.6}$ | 70 | $0.1^{+1.9}_{-1.7}$ | 45 |
| $\delta\hat{\phi}^{(l)}_5$ | $[10^{0}]$ | P | $0.70^{+0.56}_{-0.58}$ | 2.2 | $-0.03^{+0.49}_{-0.67}$ | 54 | $-1.4^{+1.3}_{-1.6}$ | 97 | $0.09^{+0.42}_{-0.42}$ | 36 | $0.10^{+0.69}_{-0.62}$ | 40 |
| | | S | $0.58^{+0.49}_{-0.44}$ | 1.7 | $0.0^{+2.5}_{-0.9}$ | 51 | $-0.9^{+1.1}_{-2.6}$ | 91 | $0.3^{+2.1}_{-0.6}$ | 25 | $-0.14^{+0.70}_{-0.90}$ | 63 |
| $\delta\hat{\phi}_6$ | $[10^{0}]$ | P | $-1.2^{+1.0}_{-1.1}$ | 97 | $0.2^{+1.3}_{-1.1}$ | 40 | $2.6^{+2.9}_{-2.5}$ | 4.7 | $-0.11^{+0.84}_{-0.84}$ | 59 | $-0.1^{+1.2}_{-1.1}$ | 56 |
| | | S | $-1.03^{+0.81}_{-0.94}$ | 98 | $0^{+14}_{-9}$ | 46 | $1.4^{+5.0}_{-1.9}$ | 12 | $-0.4^{+8.7}_{-8.9}$ | 64 | $0.4^{+1.6}_{-1.2}$ | 30 |
| $\delta\hat{\phi}^{(l)}_6$ | $[10^{1}]$ | P | $0.78^{+0.63}_{-0.70}$ | 3.5 | $-0.05^{+0.56}_{-0.66}$ | 55 | $-1.6^{+1.5}_{-1.9}$ | 96 | $0.09^{+0.44}_{-0.47}$ | 36 | $0.20^{+0.73}_{-0.78}$ | 32 |
| | | S | $0.72^{+0.57}_{-0.55}$ | 1.8 | $-0.13^{+0.70}_{-0.76}$ | 61 | $-1.0^{+1.1}_{-1.6}$ | 92 | $0.15^{+0.57}_{-0.52}$ | 31 | $-0.07^{+0.69}_{-0.78}$ | 56 |
| $\delta\hat{\phi}_7$ | $[10^{0}]$ | P | $3.2^{+2.7}_{-2.6}$ | 2.2 | $-0.4^{+2.8}_{-3.5}$ | 58 | $-7.0^{+6.9}_{-7.4}$ | 95 | $0.3^{+2.5}_{-2.2}$ | 42 | $0.2^{+3.2}_{-3.5}$ | 47 |
| | | S | $2.7^{+3.3}_{-2.2}$ | 1.8 | $-2^{+38}_{-33}$ | 62 | $-4^{+65}_{-106}$ | 81 | $0^{+98}_{-113}$ | 48 | $-1^{+7}_{-14}$ | 70 |
| $\delta\hat{\beta}_2$ | $[10^{0}]$ | P | $0.03^{+0.35}_{-0.28}$ | 43 | – | – | $-0.23^{+0.38}_{-0.35}$ | 85 | $0.13^{+0.40}_{-0.30}$ | 25 | $-0.04^{+0.28}_{-0.23}$ | 61 |
| $\delta\hat{\beta}_3$ | $[10^{0}]$ | P | $0.00^{+0.36}_{-0.30}$ | 50 | – | – | $-0.13^{+0.44}_{-0.41}$ | 71 | $0.12^{+0.28}_{-0.26}$ | 21 | $0.07^{+0.35}_{-0.29}$ | 35 |
| $\delta\hat{\alpha}_2$ | $[10^{0}]$ | P | $-0.02^{+0.34}_{-0.42}$ | 54 | – | – | $-0.13^{+0.67}_{-0.58}$ | 64 | $3^{+62}_{-55}$ | 40 | $-0.64^{+0.57}_{-0.59}$ | 97 |
| $\delta\hat{\alpha}_3$ | $[10^{0}]$ | P | $-0.1^{+1.7}_{-1.4}$ | 55 | – | – | $-0.3^{+3.8}_{-3.8}$ | 55 | $-1^{+95}_{-89}$ | 52 | $-1.9^{+2.1}_{-1.8}$ | 93 |
| $\delta\hat{\alpha}_4$ | $[10^{0}]$ | P | $-0.01^{+0.54}_{-0.51}$ | 52 | – | – | $0.5^{+2.8}_{-1.3}$ | 28 | $-1^{+45}_{-41}$ | 51 | $-0.6^{+0.9}_{-1.5}$ | 85 |

| Parameter | | Model | GW151012 | | GW170729 | | GW170809 | | GW170818 | | GW170823 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\bar{X}^+_-$ | $Q_{\rm GR}$ [%] | $\bar{X}^+_-$ | $Q_{\rm GR}$ [%] | $\bar{X}^+_-$ | $Q_{\rm GR}$ [%] | $\bar{X}^+_-$ | $Q_{\rm GR}$ [%] | $\bar{X}^+_-$ | $Q_{\rm GR}$ [%] |
| $\delta\hat{\beta}_2$ | $[10^{0}]$ | P | $0.3^{+3.8}_{-0.6}$ | 24 | $0.01^{+0.56}_{-0.51}$ | 49 | $-0.13^{+0.37}_{-0.33}$ | 74 | $-0.32^{+0.44}_{-0.29}$ | 89 | $0.00^{+0.49}_{-0.38}$ | 50 |
| $\delta\hat{\beta}_3$ | $[10^{0}]$ | P | $3.3^{+3.6}_{-3.4}$ | 5.6 | $0.31^{+0.81}_{-0.63}$ | 21 | $-0.20^{+0.42}_{-0.36}$ | 80 | $-0.41^{+0.58}_{-0.36}$ | 90 | $-0.02^{+0.74}_{-0.57}$ | 52 |
| $\delta\hat{\alpha}_2$ | $[10^{0}]$ | P | $0^{+14}_{-3}$ | 37 | $0.78^{+0.85}_{-0.73}$ | 4 | $0.42^{+0.69}_{-0.64}$ | 14 | $0.01^{+0.71}_{-0.65}$ | 49 | $-0.04^{+0.70}_{-0.66}$ | 55 |
| $\delta\hat{\alpha}_3$ | $[10^{0}]$ | P | $5^{+81}_{-33}$ | 29 | $4.4^{+4.6}_{-3.8}$ | 2.1 | $2.2^{+4.0}_{-3.0}$ | 12 | $0.6^{+3.5}_{-2.9}$ | 38 | $0.1^{+3.5}_{-2.3}$ | 49 |
| $\delta\hat{\alpha}_4$ | $[10^{0}]$ | P | $6^{+24}_{-30}$ | 33 | $1.3^{+1.2}_{-1.1}$ | 2.2 | $0.9^{+1.1}_{-1.0}$ | 7.6 | $0.4^{+1.0}_{-1.1}$ | 26 | $-0.09^{+0.97}_{-0.92}$ | 56 |

TABLE VI. 90% credible level upper bounds on the absolute value of the modified dispersion relation parameter $A_\alpha$, as well as the GR quantiles $Q_{\rm GR}$ for IMRPhenomPv2 (P) and SEOBNRv4 (S) runs. The $<$ and $>$ labels denote the bounds for $A_\alpha < 0$ and $> 0$, respectively, with the given scalings and $\bar{A}_\alpha := A_\alpha/\mathrm{eV}^{2-\alpha}$.

| Quantity | | GW170729 | | | | | | GW170814 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $<$ | | $>$ | | $Q_{\rm GR}$ [%] | | $<$ | | $>$ | | $Q_{\rm GR}$ [%] | |
| | | P | S | P | S | P | S | P | S | P | S | P | S |
| $\bar{A}_0$ | $[10^{-44}]$ | 0.30 | 0.69 | 0.67 | 0.69 | 17 | 39 | 4.3 | 2.8 | 1.2 | 1.5 | 94 | 87 |
| $\bar{A}_{0.5}$ | $[10^{-38}]$ | 0.96 | 0.88 | 1.2 | 1.9 | 26 | 16 | 5.5 | 7.9 | 1.9 | 1.9 | 92 | 92 |
| $\bar{A}_1$ | $[10^{-32}]$ | 2.2 | 4.8 | 4.7 | 4.4 | 16 | 26 | 15 | 13 | 5.5 | 4.1 | 93 | 95 |
| $\bar{A}_{1.5}$ | $[10^{-25}]$ | 0.82 | 0.77 | 1.6 | 1.2 | 18 | 17 | 3.1 | 2.9 | 1.3 | 1.1 | 96 | 94 |
| $\bar{A}_{2.5}$ | $[10^{-13}]$ | 4.4 | 4.9 | 1.5 | 1.7 | 94 | 93 | 1.4 | 1.3 | 4.2 | 5.0 | 5.7 | 5.1 |
| $\bar{A}_3$ | $[10^{-8}]$ | 38 | 32 | 9.1 | 17 | 96 | 92 | 7.9 | 8.8 | 27 | 38 | 6.2 | 6.4 |
| $\bar{A}_{3.5}$ | $[10^{-2}]$ | 28 | 33 | 9.5 | 13 | 94 | 94 | 5.8 | 6.2 | 31 | 18 | 7.5 | 8.5 |
| $\bar{A}_4$ | $[10^{4}]$ | 46 | 50 | 7.6 | 9.7 | 95 | 95 | 4.5 | 4.9 | 14 | 21 | 9.6 | 11 |

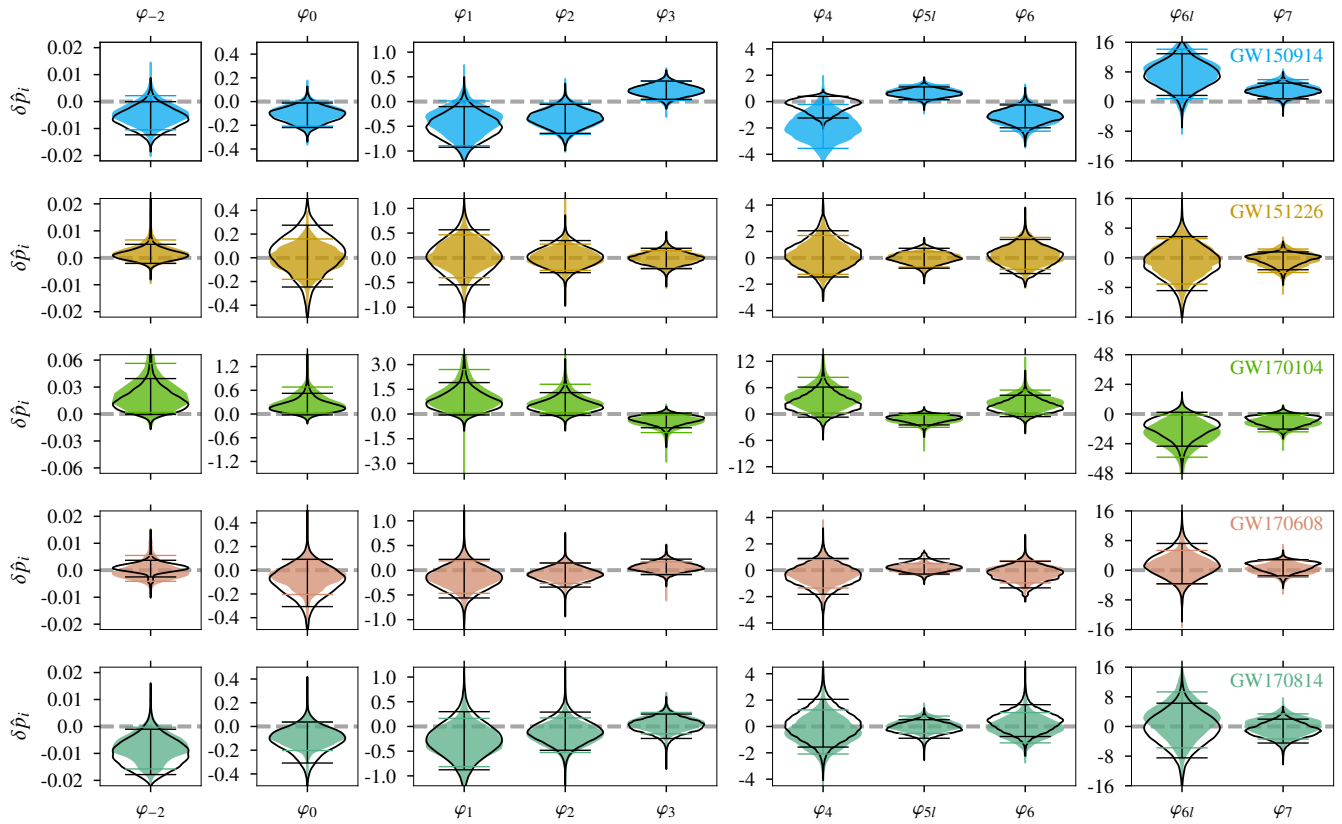


FIG. 8. Violin plots showing inspiral $\delta\hat{p}_i$ posteriors for the individual binary black-hole events of GWTC-1 [14] outlined in Sec. III (see "PI" column of Table I), using IMRPhenomPv2 (shaded regions) and SEOBNRv4 (black solid lines). Thin horizontal lines indicate the 90% credible intervals, which show an overall statistical consistency with GR (dashed grey line).


[1] C. M. Will, "The Confrontation between General Relativity and Experiment," Living Rev. Relativity, **17**, 4 (2014), arXiv:1403.7377 [gr-qc].

[2] J. Aasi *et al.* (LIGO Scientific Collaboration), "Advanced LIGO," Classical Quantum Gravity, **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].

[3] F. Acernese *et al.* (Virgo Collaboration), "Advanced Virgo: a second-generation interferometric gravitational wave detector," Classical Quantum Gravity, **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].

[4] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "Tests of general relativity with GW150914," Phys. Rev. Lett., **116**, 221101 (2016), **121**, 129902(E) (2018), arXiv:1602.03841 [gr-qc].

[5] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "Binary Black Hole Mergers in the first Advanced LIGO Observing Run," Phys. Rev. X, **6**, 041015 (2016), **8**, 039903(E) (2018), arXiv:1606.04856 [gr-qc].

[6] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2," Phys. Rev. Lett., **118**, 221101 (2017), **121**, 129901(E) (2018), arXiv:1706.01812 [gr-qc].

[7] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence," Phys. Rev. Lett., **119**, 141101 (2017), arXiv:1709.09660 [gr-qc].

[8] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "Tests of General Relativity with GW170817," (2018), arXiv:1811.00364 [gr-qc].

[9] LIGO Scientific Collaboration and Virgo Collaboration, "GWTC-1," https://doi.org/10.7935/82H3-HH23 (2018).

[10] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "Observation of Gravitational Waves from a Binary Black Hole Merger," Phys. Rev. Lett., **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].

[11] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "GW150914: First results from the search for binary black hole coalescence with Advanced LIGO," Phys. Rev. D, **93**, 122003 (2016), arXiv:1602.03839 [gr-qc].

[12] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "GW151226: Observation of gravitational waves from a 22-solar-mass binary black hole coalescence," Phys. Rev. Lett., **116**, 241103 (2016), arXiv:1606.04855 [gr-qc].

[13] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "GW170608: Observation of a 19 solar-mass

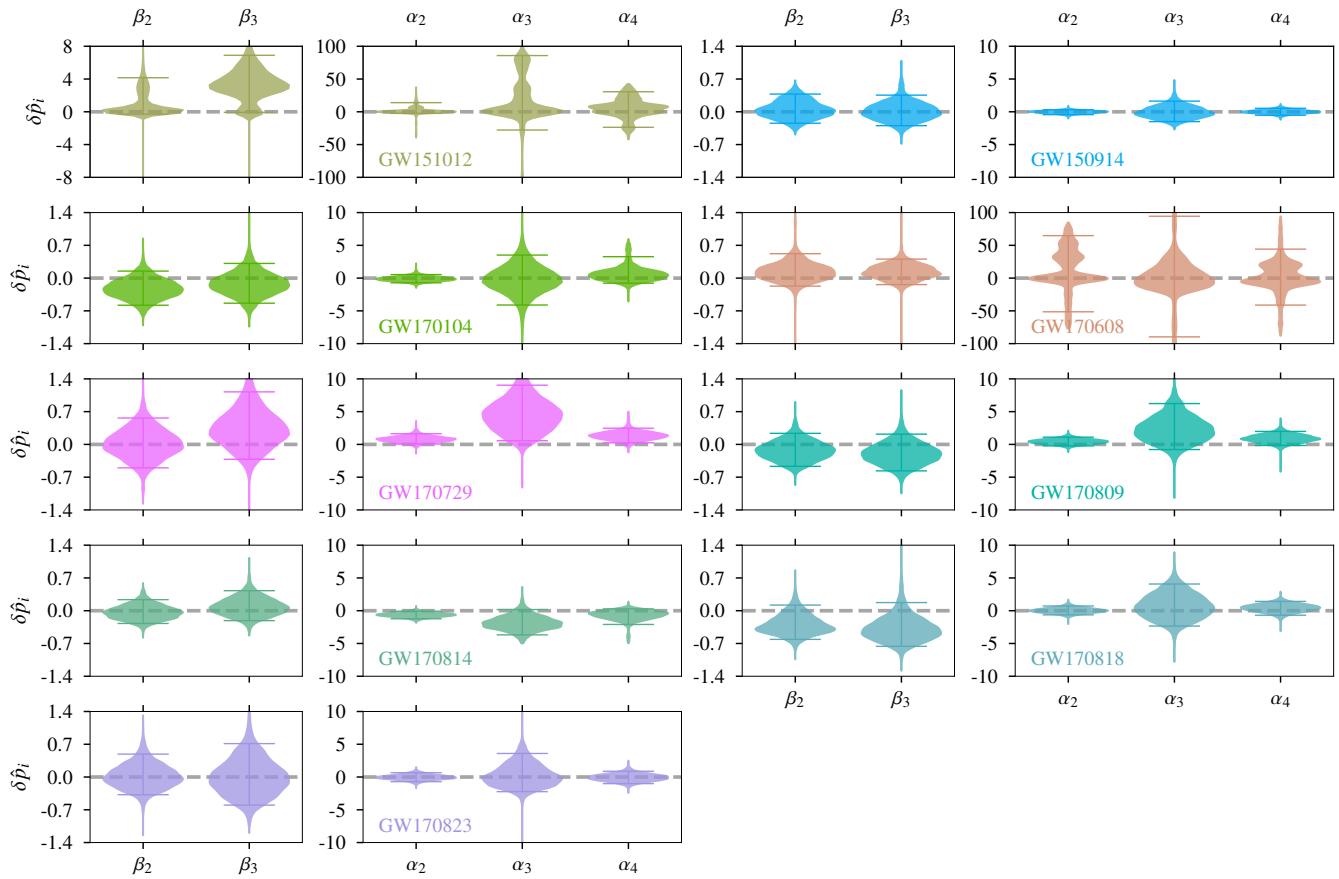


FIG. 9. Violin plots showing post-inspiral $\delta\hat{p}_i$ posteriors for the individual binary black-hole events of GWTC-1 [14] outlined in Sec. III (see "PPI" column of Table I), using IMRPhenomPv2. Thin horizontal lines indicate the 90% credible intervals, which show an overall statistical consistency with GR (dashed grey line).


binary black hole coalescence," Astrophys. J. Lett., **851**, L35 (2017), arXiv:1711.05578 [astro-ph.HE].

[14] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs," (2018), arXiv:1811.12907 [astro-ph.HE].

[15] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "Observing gravitational-wave transient GW150914 with minimal assumptions," Phys. Rev. D, **93**, 122004 (2016), arXiv:1602.03843 [gr-qc].

[16] M. Vallisneri and N. Yunes, "Stealth Bias in Gravitational-Wave Parameter Estimation," Phys. Rev. D, **87**, 102002 (2013), arXiv:1301.2627 [gr-qc].

[17] S. Vitale and W. Del Pozzo, "How serious can the stealth bias be in gravitational wave parameter estimation?" Phys. Rev. D, **89**, 022002 (2014), arXiv:1311.2057 [gr-qc].

[18] 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), arXiv:1611.03703 [gr-qc].

[19] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Jiménez Forteza, and A. Bohé, "Frequency-domain gravitational waves from non-precessing black-hole binaries. II. A phenomenological model for the advanced detector era," Phys. Rev. D, **93**, 044007 (2016), arXiv:1508.07253 [gr-qc].

[20] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, "A Surrogate Model of Gravitational Waveforms from Numerical Relativity Simulations of Precessing Binary Black Hole Mergers," Phys. Rev. D, **95**, 104023 (2017), arXiv:1701.00550 [gr-qc].

[21] S. Khan, K. Chatziioannou, M. Hannam, and F. Ohme, "Phenomenological model for the gravitational-wave signal from precessing binary black holes with two-spin effects," (2018), arXiv:1809.10113 [gr-qc].

[22] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "Effects of waveform model systematics on the interpretation of GW150914," Classical Quantum Gravity, **34**, 104002 (2017), arXiv:1611.07531 [gr-qc].

[23] LIGO Scientific Collaboration and Virgo Collaboration, "Data release for testing GR with GWTC-1," https://dcc.ligo.org/LIGO-P1900087/public (2019).

[24] LIGO Scientific Collaboration, Virgo Collaboration, "Gravitational Wave Open Science Center," https://www.gw-openscience.org (2018).

[25] S. Karki *et al.*, "The Advanced LIGO Photon Calibrators," Rev. Sci. Instrum., **87**, 114503 (2016), arXiv:1608.05055 [astro-ph.IM].

[26] C. Cahillane *et al.*, "Calibration uncertainty for Advanced LIGO's first and second observing runs," Phys. Rev. D, **96**,

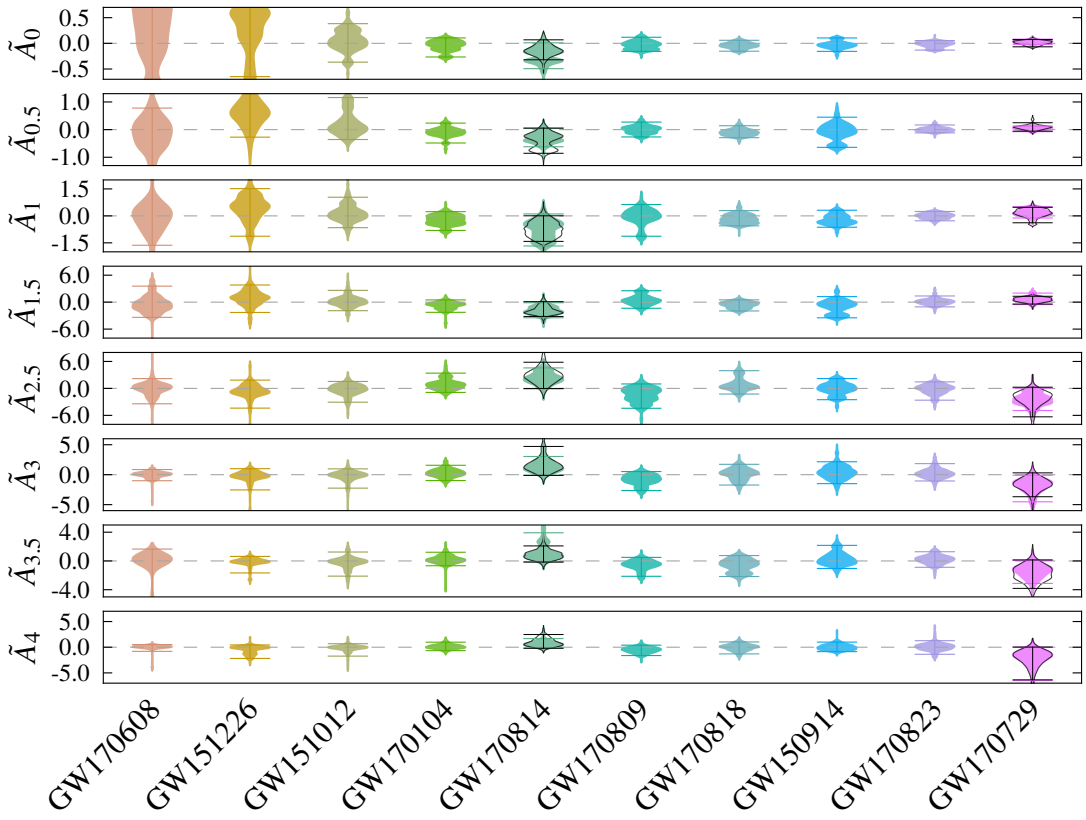


FIG. 10. Violin plots of the full posteriors for the modified dispersion relation parameter $A_\alpha$ for the individual binary black-hole events of GWTC-1 [14], with the 90% credible interval around the median indicated. The events are ordered left-to-right by increasing median total mass. The filled violins are the IMRPhenomPv2 results, while the unfilled violins give the SEOBNRv4 results for GW170729 and GW170814. Here $\tilde{A}_\alpha := A_\alpha/(10^{-19}\ \mathrm{peV}^{2-\alpha})$. We have let the much less constraining posteriors extend off the edges of the plots in order to show the more constraining posteriors in detail. The violin plots are scaled so that the maximum value of the posterior always has the same width, and these maximum values never occur off the plot.


102001 (2017), arXiv:1708.03023 [astro-ph.IM].

[27] A. Viets *et al.*, "Reconstructing the calibrated strain signal in the Advanced LIGO detectors," Classical Quantum Gravity, **35**, 095015 (2018), arXiv:1710.09973 [astro-ph.IM].

[28] D. Estevez *et al.*, *V1O2Repro2A h(t) reprocessing for Virgo O2 data*, Tech. Rep. VIR-0362A-18 (Virgo project, 2018).

[29] F. Acernese *et al.* (Virgo Collaboration), "Calibration of Advanced Virgo and Reconstruction of the Gravitational Wave Signal $h(t)$ during the Observing Run O2," Classical Quantum Gravity, **35**, 205004 (2018), arXiv:1807.03275 [gr-qc].

[30] W. M. Farr, B. Farr, and T. Littenberg, *Modelling Calibration Errors In CBC Waveforms*, Tech. Rep. LIGO-T1400682 (LIGO Project, 2015).

[31] J. Veitch *et al.*, "Robust parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library," Phys. Rev. D, **91**, 042003 (2015), arXiv:1409.7215 [gr-qc].

[32] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "Properties of the binary neutron star merger GW170817," Phys. Rev. X, **9**, 011001 (2019), arXiv:1805.11579 [gr-qc].

[33] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral," Phys. Rev. Lett., **119**, 161101 (2017).

[34] J. C. Driggers *et al.* (LIGO Scientific Collaboration Instrument Science Authors), "Improving astrophysical parameter estimation via offline noise subtraction for Advanced LIGO," Phys. Rev. D, **99**, 042001 (2019), arXiv:1806.00532 [astro-ph.IM].

[35] D. Davis, T. J. Massinger, A. P. Lundgren, J. C. Driggers, A. L. Urban, and L. K. Nuttall, "Improving the Sensitivity of Advanced LIGO Using Noise Subtraction," (2018), arXiv:1809.05348 [astro-ph.IM].

[36] S. Klimenko, I. Yakushin, A. Mercer, and G. Mitselmakher, "Coherent method for detection of gravitational wave bursts," Classical Quantum Gravity, **25**, 114029 (2008), arXiv:0802.3232 [gr-qc].

[37] S. Klimenko *et al.*, "Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors," Phys. Rev. D, **93**, 042004 (2016), arXiv:1511.05999

[gr-qc].
[38] A. Ghosh, N. K. Johnson-McDaniel, A. Ghosh, C. K. Mishra, P. Ajith, W. Del Pozzo, C. P. L. Berry, A. B. Nielsen, and L. London, "Testing general relativity using gravitational wave signals from the inspiral, merger and ringdown of binary black holes," Classical Quantum Gravity, **35**, 014002 (2018), arXiv:1704.06784 [gr-qc].
[39] 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), arXiv:1311.0420 [gr-qc].
[40] A. H. Nitz *et al.*, "PyCBC software," `https://github.com/ligo-cbc/pycbc` (2018).
[41] T. Dal Canton *et al.*, "Implementing a search for aligned-spin neutron star-black hole systems with advanced ground based gravitational wave detectors," Phys. Rev. D, **90**, 082004 (2014), arXiv:1405.6731 [gr-qc].
[42] S. A. Usman *et al.*, "The PyCBC search for gravitational waves from compact binary coalescence," Classical Quantum Gravity, **33**, 215004 (2016), arXiv:1508.02357 [gr-qc].
[43] S. Sachdev *et al.*, "The GstLAL Search Analysis Methods for Compact Binary Mergers in Advanced LIGO's Second and Advanced Virgo's First Observing Runs," (2019), arXiv:1901.08580 [gr-qc].
[44] C. Messick *et al.*, "Analysis Framework for the Prompt Discovery of Compact Binary Mergers in Gravitational-wave Data," Phys. Rev. D, **95**, 042001 (2017), arXiv:1604.04324 [astro-ph.IM].
[45] P. C. Peters and J. Mathews, "Gravitational radiation from point masses in a Keplerian orbit," Phys. Rev., **131**, 435 (1963).
[46] P. C. Peters, "Gravitational Radiation and the Motion of Two Point Masses," Phys. Rev., **136**, B1224 (1964).
[47] T. Hinderer and S. Babak, "Foundations of an effective-one-body model for coalescing binaries on eccentric orbits," Phys. Rev. D, **96**, 104048 (2017), arXiv:1707.08426 [gr-qc].
[48] Z. Cao and W.-B. Han, "Waveform model for an eccentric binary black hole based on the effective-one-body-numerical-relativity formalism," Phys. Rev. D, **96**, 044028 (2017), arXiv:1708.00166 [gr-qc].
[49] I. Hinder, L. E. Kidder, and H. P. Pfeiffer, "Eccentric binary black hole inspiral-merger-ringdown gravitational waveform model from numerical relativity and post-Newtonian theory," Phys. Rev. D, **98**, 044015 (2018), arXiv:1709.02007 [gr-qc].
[50] E. A. Huerta *et al.*, "Eccentric, nonspinning, inspiral, Gaussian-process merger approximant for the detection and characterization of eccentric binary black hole mergers," Phys. Rev. D, **97**, 024031 (2018), arXiv:1711.06276 [gr-qc].
[51] A. Klein, Y. Boetzel, A. Gopakumar, P. Jetzer, and L. de Vittori, "Fourier domain gravitational waveforms for precessing eccentric binaries," Phys. Rev. D, **98**, 104043 (2018), arXiv:1801.08542 [gr-qc].
[52] B. Moore, T. Robson, N. Loutrel, and N. Yunes, "Towards a Fourier domain waveform for non-spinning binaries with arbitrary eccentricity," Classical Quantum Gravity, **35**, 235006 (2018), arXiv:1807.07163 [gr-qc].
[53] J. Samsing, "Eccentric Black Hole Mergers Forming in Globular Clusters," Phys. Rev. D, **97**, 103014 (2018), arXiv:1711.07452 [astro-ph.HE].
[54] C. L. Rodriguez, P. Amaro-Seoane, S. Chatterjee, and F. A. Rasio, "Post-Newtonian Dynamics in Dense Star Clusters: Highly-Eccentric, Highly-Spinning, and Repeated Binary Black Hole Mergers," Phys. Rev. Lett., **120**, 151101 (2018), arXiv:1712.04937 [astro-ph.HE].
[55] M. Zevin, J. Samsing, C. Rodriguez, C.-J. Haster, and E. Ramirez-Ruiz, "Eccentric Black Hole Mergers in Dense Star Clusters: The Role of Binary-Binary Encounters," Astrophys. J., **871**, 1 (2019), arXiv:1810.00901 [astro-ph.HE].
[56] C. L. Rodriguez, P. Amaro-Seoane, S. Chatterjee, K. Kremer, F. A. Rasio, J. Samsing, C. S. Ye, and M. Zevin, "Post-Newtonian Dynamics in Dense Star Clusters: Formation, Masses, and Merger Rates of Highly-Eccentric Black Hole Binaries," Phys. Rev. D, **98**, 123005 (2018), arXiv:1811.04926 [astro-ph.HE].
[57] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, "Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal," Phys. Rev. D, **93**, 044006 (2016), arXiv:1508.07250 [gr-qc].
[58] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, "Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms," Phys. Rev. Lett., **113**, 151101 (2014), arXiv:1308.3271 [gr-qc].
[59] A. Taracchini *et al.*, "Effective-one-body model for black-hole binaries with generic mass ratios and spins," Phys. Rev. D, **89**, 061502 (2014), arXiv:1311.2544 [gr-qc].
[60] S. Babak, A. Taracchini, and A. Buonanno, "Validating the effective-one-body model of spinning, precessing binary black holes against numerical relativity," Phys. Rev. D, **95**, 024010 (2017), arXiv:1607.05661 [gr-qc].
[61] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, C. D. Ott, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and B. Szilágyi, "Numerical relativity waveform surrogate model for generically precessing binary black hole mergers," Phys. Rev. D, **96**, 024058 (2017), arXiv:1705.07089 [gr-qc].
[62] I. Kamaretsos, M. Hannam, S. Husa, and B. S. Sathyaprakash, "Black-hole hair loss: Learning about binary progenitors from ringdown signals," Phys. Rev. D, **85**, 024018 (2012), arXiv:1107.0854 [gr-qc].
[63] L. London, D. Shoemaker, and J. Healy, "Modeling ringdown: Beyond the fundamental quasinormal modes," Phys. Rev. D, **90**, 124032 (2014), **94**, 069902(E) (2016), arXiv:1404.3197 [gr-qc].
[64] L. London, S. Khan, E. Fauchon-Jones, C. García, M. Hannam, S. Husa, X. Jiménez-Forteza, C. Kalaghatgi, F. Ohme, and F. Pannarale, "First higher-multipole model of gravitational waves from spinning and coalescing black-hole binaries," Phys. Rev. Lett., **120**, 161102 (2018), arXiv:1708.00404 [gr-qc].
[65] 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), arXiv:1803.10701 [gr-qc].
[66] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, L. E. Kidder, and H. P. Pfeiffer, "Surrogate model of hybridized numerical relativity binary black hole waveforms," (2018), arXiv:1812.07865 [gr-qc].
[67] V. Varma and P. Ajith, "Effects of nonquadrupole modes in the detection and parameter estimation of black hole binaries with nonprecessing spins," Phys. Rev. D, **96**, 124024 (2017), arXiv:1612.05608 [gr-qc].
[68] P. T. H. Pang, J. Calderón Bustillo, Y. Wang, and T. G. F. Li, "Potential observations of false deviations from general relativity in gravitational wave signals from binary black holes," Phys. Rev. D, **98**, 024019 (2018), arXiv:1802.03306 [gr-qc].
[69] LIGO Scientific Collaboration and Virgo Collaboration, "LALSuite software," (2018).
[70] N. J. Cornish and T. B. Littenberg, "BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument

Glitches," Classical Quantum Gravity, **32**, 135012 (2015), arXiv:1410.3835 [gr-qc].
[71] T. B. Littenberg and N. J. Cornish, "Bayesian inference for spectral estimation of gravitational wave detector noise," Phys. Rev. D, **91**, 084034 (2015), arXiv:1410.3852 [gr-qc].
[72] T. A. Apostolatos, "Search templates for gravitational waves from precessing, inspiraling binaries," Phys. Rev. D, **52**, 605 (1995).
[73] R. A. Fisher, "Questions and answers," Am. Statistician, **2**, 30 (1948).
[74] A. Ghosh *et al.*, "Testing general relativity using golden black-hole binaries," Phys. Rev. D, **94**, 021101(R) (2016), arXiv:1602.02453 [gr-qc].
[75] J. M. Bardeen, W. H. Press, and S. A. Teukolsky, "Rotating black holes: Locally nonrotating frames, energy extraction, and scalar synchrotron radiation," Astrophys. J., **178**, 347 (1972).
[76] J. Healy and C. O. Lousto, "Remnant of binary black-hole mergers: New simulations and peak luminosity studies," Phys. Rev. D, **95**, 024037 (2017), arXiv:1610.09713 [gr-qc].
[77] F. Hofmann, E. Barausse, and L. Rezzolla, "The final spin from binary black holes in quasi-circular orbits," Astrophys. J. Lett., **825**, L19 (2016), arXiv:1605.01938 [gr-qc].
[78] X. Jiménez-Forteza, D. Keitel, S. Husa, M. Hannam, S. Khan, and M. Pürrer, "Hierarchical data-driven approach to fitting numerical relativity data for nonprecessing binary black holes with an application to final spin and radiated energy," Phys. Rev. D, **95**, 064024 (2017).
[79] N. K. Johnson-McDaniel *et al.*, *Determining the final spin of a binary black hole system including in-plane spins: Method and checks of accuracy*, Tech. Rep. LIGO-T1600168 (LIGO Project, 2016) https://dcc.ligo.org/LIGO-T1600168/public/main.
[80] L. Blanchet, T. Damour, B. R. Iyer, C. M. Will, and A. G. Wiseman, "Gravitational-radiation damping of compact binary systems to second post-Newtonian order," Phys. Rev. Lett., **74**, 3515 (1995), arXiv:gr-qc/9501027.
[81] L. Blanchet, T. Damour, G. Esposito-Farèse, and B. R. Iyer, "Gravitational radiation from inspiralling compact binaries completed at the third post-Newtonian order," Phys. Rev. Lett., **93**, 091101 (2004), arXiv:gr-qc/0406012.
[82] L. Blanchet, T. Damour, G. Esposito-Farèse, and B. R. Iyer, "Dimensional regularization of the third post-Newtonian gravitational wave generation from two point masses," Phys. Rev. D, **71**, 124004 (2005), arXiv:gr-qc/0503044.
[83] L. Blanchet, "Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries," Living Rev. Relativity, **17**, 2 (2014), arXiv:1310.1528 [gr-qc].
[84] 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), arXiv:gr-qc/0604067.
[85] 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), arXiv:gr-qc/0604018.
[86] 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), arXiv:1005.0304 [gr-qc].
[87] N. Yunes and F. Pretorius, "Fundamental Theoretical Bias in Gravitational Wave Astrophysics and the Parameterized Post-Einsteinian Framework," Phys. Rev. D, **80**, 122003 (2009), arXiv:0909.3328 [gr-qc].
[88] 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), arXiv:1110.0530 [gr-qc].
[89] 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: Further investigations," J. Phys. Conf. Ser., **363**, 012028 (2012), arXiv:1111.5274 [gr-qc].
[90] N. Cornish, L. Sampson, N. Yunes, and F. Pretorius, "Gravitational Wave Tests of General Relativity with the Parameterized Post-Einsteinian Framework," Phys. Rev. D, **84**, 062003 (2011), arXiv:1105.2088 [gr-qc].
[91] L. Sampson, N. Cornish, and N. Yunes, "Mismodeling in gravitational-wave astronomy: The trouble with templates," Phys. Rev. D, **89**, 064037 (2014), arXiv:1311.4898 [gr-qc].
[92] 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), arXiv:1712.08772 [gr-qc].
[93] 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), arXiv:1603.08955 [gr-qc].
[94] 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), arXiv:1510.02152 [gr-qc].
[95] K. Yagi and L. C. Stein, "Black Hole Based Tests of General Relativity," Classical Quantum Gravity, **33**, 054001 (2016), arXiv:1602.02413 [gr-qc].
[96] E. Barausse, N. Yunes, and K. Chamberlain, "Theory-Agnostic Constraints on Black-Hole Dipole Radiation with Multiband Gravitational-Wave Astrophysics," Phys. Rev. Lett., **116**, 241104 (2016), arXiv:1603.04075 [gr-qc].
[97] K. G. Arun, "Generic bounds on dipolar gravitational radiation from inspiralling compact binaries," Classical Quantum Gravity, **29**, 075011 (2012), arXiv:1202.5911 [gr-qc].
[98] L. Sampson, N. Cornish, and N. Yunes, "Gravitational Wave Tests of Strong Field General Relativity with Binary Inspirals: Realistic Injections and Optimal Model Selection," Phys. Rev. D, **87**, 102001 (2013), arXiv:1303.1185 [gr-qc].
[99] S. Mirshekari, N. Yunes, and C. M. Will, "Constraining Generic Lorentz Violation and the Speed of the Graviton with Gravitational Waves," Phys. Rev. D, **85**, 024041 (2012), arXiv:1110.2720 [gr-qc].
[100] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, Fermi-GBM Collaboration, and INTEGRAL Collaboration), "Gravitational Waves and Gamma-rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A," Astrophys. J. Lett., **848**, L13 (2017), arXiv:1710.05834 [astro-ph.HE].
[101] C. M. Will, "Bounding the mass of the graviton using gravitational wave observations of inspiralling compact binaries," Phys. Rev. D, **57**, 2061 (1998), arXiv:gr-qc/9709011 [gr-qc].
[102] G. Calcagni, "Fractal universe and quantum gravity," Phys. Rev. Lett., **104**, 251301 (2010), arXiv:0912.3142 [hep-th].
[103] G. Amelino-Camelia, "Doubly special relativity," Nature (London), **418**, 34 (2002), arXiv:gr-qc/0207049 [gr-qc].
[104] P. Hořava, "Quantum Gravity at a Lifshitz Point," Phys. Rev. D, **79**, 084008 (2009), arXiv:0901.3775 [hep-th].
[105] A. S. Sefiedgar, K. Nozari, and H. R. Sepangi, "Modified dispersion relations in extra dimensions," Phys. Lett., **B696**,

119 (2011), arXiv:1012.1406 [gr-qc].
[106] V. A. Kostelecký and M. Mewes, "Testing local Lorentz invariance with gravitational waves," Phys. Lett., **B757**, 510 (2016), arXiv:1602.04782 [gr-qc].
[107] P. A. R. Ade *et al.* (Planck Collaboration), "Planck 2015 results. XIII. Cosmological parameters," Astron. Astrophys., **594**, A13 (2016), arXiv:1502.01589 [astro-ph.CO].
[108] N. Aghanim *et al.* (Planck Collaboration), "Planck 2018 results. VI. Cosmological parameters," (2018), arXiv:1807.06209 [astro-ph.CO].
[109] G. Amelino-Camelia, M. Arzano, Y. Ling, and G. Mandanici, "Black-hole thermodynamics with modified dispersion relations and generalized uncertainty principles," Classical Quantum Gravity, **23**, 2585 (2006), arXiv:gr-qc/0506110 [gr-qc].
[110] G. Calcagni, "Lorentz violations in multifractal spacetimes," Eur. Phys. J., **C77**, 291 (2017), arXiv:1603.03046 [gr-qc].
[111] L. Bernus, O. Minazzoli, A. Fienga, M. Gastineau, and J. Laskar, "Constraining the mass of the graviton with the planetary ephemeris INPOP," (2019), arXiv:1901.04307 [gr-qc].
[112] C. M. Will, "Solar system vs. gravitational-wave bounds on the graviton mass," Classical Quantum Gravity, **35**, 17LT01 (2018), arXiv:1805.10523 [gr-qc].
[113] C. de Rham, J. T. Deskins, A. J. Tolley, and S.-Y. Zhou, "Graviton Mass Bounds," Rev. Mod. Phys., **89**, 025004 (2017), arXiv:1606.08462 [astro-ph.CO].
[114] D. B. Rubin, "The Bayesian Bootstrap," Ann. Statist., **9**, 130 (1981).
[115] D. M. Eardley, D. L. Lee, and A. P. Lightman, "Gravitational-wave observations as a tool for testing relativistic gravity," Phys. Rev. D, **8**, 3308 (1973).
[116] K. Chatziioannou, N. Yunes, and N. Cornish, "Model-Independent Test of General Relativity: An Extended post-Einsteinian Framework with Complete Polarization Content," Phys. Rev. D, **86**, 022004 (2012), **95**, 129901(E) (2017), arXiv:1204.2585 [gr-qc].
[117] M. Isi, M. Pitkin, and A. J. Weinstein, "Probing Dynamical Gravity with the Polarization of Continuous Gravitational Waves," Phys. Rev. D, **96**, 042001 (2017), arXiv:1703.07530 [gr-qc].
[118] T. Callister, A. S. Biscoveanu, N. Christensen, M. Isi, A. Matas, O. Minazzoli, T. Regimbau, M. Sakellariadou, J. Tasson, and E. Thrane, "Polarization-based Tests of Gravity with the Stochastic Gravitational-Wave Background," Phys. Rev. X, **7**, 041058 (2017), arXiv:1704.08373 [gr-qc].
[119] M. Isi and A. J. Weinstein, "Probing gravitational wave polarizations with signals from compact binary coalescences," Tech. Note, LIGO-P1700276 (2017), arXiv:1710.03794 [gr-qc].
[120] A. Błaut, "Angular and frequency response of the gravitational wave interferometers in the metric theories of gravity," Phys. Rev. D, **85**, 043005 (2012).
[121] B. P. Abbott *et al.* (KAGRA Collaboration, LIGO Scientific Collaboration, and Virgo Collaboration), "Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO, Advanced Virgo and KAGRA," Living Rev. Relativity, **21**, 3 (2018), arXiv:1304.0670 [gr-qc].
[122] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), "Directly comparing GW150914 with numerical solutions of Einstein's equations for binary black hole coalescence," Phys. Rev. D, **94**, 064035 (2016), arXiv:1606.01262 [gr-qc].
[123] J. Lange *et al.*, "Parameter estimation method that directly compares gravitational wave observations to numerical relativity," Phys. Rev. D, **96**, 104041 (2017), arXiv:1705.09833 [gr-qc].
[124] G. Ashton *et al.*, "Bilby: A user-friendly Bayesian inference library for gravitational-wave astronomy," (2018), arXiv:1811.02042 [astro-ph.IM].