2016

Tests of General Relativity with GW150914

B. P. Abbott  
CALTECH, LIGO, Pasadena, CA 91125 USA;

R. Abbott  
CALTECH, LIGO, Pasadena, CA 91125 USA;

M. R. Abernathy  
CALTECH, LIGO, Pasadena, CA 91125 USA;

E. E. Mikhailov  
College of William & Mary, Williamsburg, VA 23187 USA

H. Rew  
College of William & Mary, Williamsburg, VA 23187 USA

See next page for additional authors

Follow this and additional works at: https://scholarworks.wm.edu/aspubs

Recommended Citation

This Article is brought to you for free and open access by the Arts and Sciences at W&M ScholarWorks. It has been accepted for inclusion in Arts & Sciences Articles by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.
Tests of General Relativity with GW150914

B. P. Abbott et al.*
(LIGO Scientific and Virgo Collaborations)
(Received 26 March 2016; revised manuscript received 9 May 2016; published 31 May 2016)

The LIGO detection of GW150914 provides an unprecedented opportunity to study the two-body motion of a compact-object binary in the large-velocity, highly nonlinear regime, and to witness the final merger of the binary and the excitation of uniquely relativistic modes of the gravitational field. We carry out several investigations to determine whether GW150914 is consistent with a binary black-hole merger in general relativity. We find that the final remnant’s mass and spin, as determined from the low-frequency (inspiral) and high-frequency (postinspiral) phases of the signal, are mutually consistent with the binary black-hole solution in general relativity. Furthermore, the data following the peak of GW150914 are consistent with the least-damped quasinormal mode inferred from the mass and spin of the remnant black hole. By using waveform models that allow for parametrized general-relativity violations during the inspiral and merger phases, we perform quantitative tests on the gravitational-wave phase in the dynamical regime and we determine the first empirical bounds on several high-order post-Newtonian coefficients. We constrain the graviton Compton wavelength, assuming that gravitons are dispersed in vacuum in the same way as particles with mass, obtaining a 90%-confidence lower bound of $10^{13}$ km. In conclusion, within our statistical uncertainties, we find no evidence for violations of general relativity in the genuinely strong-field regime of gravity.

DOI: 10.1103/PhysRevLett.116.221101

Introduction.—On September 14, 2015, at 09:50:45 Universal Time, the LIGO detectors at Hanford, Washington, and Livingston, Louisiana, detected a gravitational-wave (GW) signal, henceforth GW150914, with an observed signal-to-noise ratio (SNR) $\sim 24$. The probability that GW150914 was due to a random noise fluctuation was later established to be $<2 \times 10^{-7}$ [1,2]. GW150914 exhibited the expected signature of an inspiral, merger, and ringdown signal from a coalescing binary system [1]. Assuming that general relativity (GR) is the correct description for GW150914, detailed follow-up analyses determined the (detector-frame) component masses of the binary system to be $39^{+6}_{-4} M_\odot$ and $32^{+4}_{-5} M_\odot$ at 90% credible intervals [3], corroborating the hypothesis that GW150914 was emitted by a binary black hole.

In Newtonian gravity, binary systems move along circular or elliptical orbits with a constant orbital period [4,5]. In GR, binary systems emit GWs [6,7]: as a consequence, the binary’s orbital period decreases over time as energy and angular momentum are radiated away. Electromagnetic observations of binary pulsars over the four decades since their discovery [8,9] have made it possible to measure GW-induced orbital-period variations $\dot{P}_{\text{orb}} \sim -10^{-14} - 10^{-12}$, confirming the GW luminosity predicted at leading order in post-Newtonian (PN) theory [10] (i.e., Einstein’s quadrupole formula) with exquisite precision [11,12]. Nevertheless, even in the most relativistic binary pulsar known today, J0737-3039 [11], the orbital period changes at an effectively constant rate. The orbital velocity $v$ relative to the speed of light $c$ is $v/c \sim 2 \times 10^{-3}$, and the two neutron stars in the system will coalesce in $\sim 85 \times 10^6$ yr.

By contrast, GW150914 was emitted by a rapidly evolving, dynamical binary that swept through the detectors’ bandwidth and merged in a fraction of a second, with $\dot{P}_{\text{orb}}$ ranging from $\sim -0.1$ at $f_{GW} \sim 30$ Hz to $\sim -1$ at $f_{GW} \sim 132$ Hz (just before merger, where $v/c$ reached $\sim 0.5$). Thus, through GW150914 we observe the two-body motion in the large-velocity, highly dynamical, strong-field regime of gravity, leading to the formation of a new merged object and generating GWs. While Solar System experiments, binary-pulsar observations, and cosmological measurements are all in excellent agreement with GR (see Refs. [12–14] and the references therein), they test it in low-velocity, quasistatic, weak-field, or linear regimes. (While the orbits of binary pulsars are weakly relativistic, pulsars themselves are strongly self-gravitating bodies, so they do offer opportunities to test strong-field gravity [15,16].) Thus, GW150914 opens up the distinct opportunity of probing unexplored sectors of GR.

Here, we perform several studies of GW150914, aimed at detecting deviations from the predictions of GR. Within the limits set by LIGO’s sensitivity and by the nature of GW150914, we find no statistically significant evidence against the hypothesis that GW150914 was emitted by two black holes spiraling toward each other and merging to form a single, rotating black hole [17,18], and that the

---

*Full author list given at end of the article.
dynamics of the process as a whole was in accordance with the vacuum Einstein field equations. We begin by constraining the level of coherent (i.e., GW-like) residual strain left after removing the most-probable GR waveform from the GW150914 data, and we use this estimated level to bound GR violations which are not degenerate with changes in the parameters of the binary. We then verify that the mass and spin parameters of the final black hole, as predicted from the binary’s inspiral signal, are consistent with the final parameters inferred from the postinspiral (merger and ringdown) signal. We find that the data following the peak of GW150914 are consistent with the least-damped quasinormal mode (QNM) inferred from the vacuum Einstein field equations.

As we shall see, the constraints on the strong-field dynamics of gravity obtained from GW150914 are not yet very tight; for instance, some of the bounds on relative deviations in PN parameters are \( O(1) \). On the other hand, it is to be noted that the LIGO detectors are still a factor of a few away from their final design sensitivities [19], and even louder sources than GW150914 may be seen in the near future; moreover, as more detections are made, we will be able to combine information from all of the observed sources to obtain progressively sharper bounds on PN and other coefficients.

In the rest of this Letter, when reporting physical quantities that are redshifted in the transformation between the source and detector frames, we refer to the detector frame unless we specify otherwise.

Waveform models, systematics, and statistical effects.— Tests of GR from GW observations build on the knowledge of the gravitational waveform in GR, and on the statistical properties of instrumental noise. Any uncontrolled systematic effect from waveform modeling and/or the detectors could, in principle, affect the outcome of our tests. Thus, we begin by checking to ensure that these uncertainties either are below our measurement precision or are accounted for.

The analytical inspiral-merger-ringdown (IMR) waveform models used in this Letter were developed within two frameworks: (i) the effective-one-body (EOB) formalism [20–24], which combines PN results [10] with NR [25–27] and perturbation theory [28–30], and (ii) a phenomenological approach [31–34] based on extending frequency-domain PN expressions and hybridizing PN and EOB with NR waveforms. Specifically, here we adopt the double-spin, nonprecessing waveform model developed in Ref. [35] using NR waveforms from Ref. [36], enhanced with reduced-order modeling [37] to speed up waveform generation [38,39] (henceforth, EOBNR), and the single effective spin, precessing waveform model of Refs. [40–42] (henceforth, IMRPHENOM). (The specific names of the two waveform models that we use in the LIGO ALGORITHM LIBRARY are SEOBNR V2_ROM_DOUBLESPIN and IMRPHENOMPv2.) Both models are calibrated against waveforms from direct numerical integration of the Einstein equations.

As shown in Refs. [3,35,41,43], in the region of parameter space relevant for GW150914, the error due to differences between the two analytical waveform models (and between the analytical and numerical-relativity waveforms) is smaller than the typical statistical uncertainty due to the finite SNR of GW150914. To assess potential modeling systematics, we collected existing NR waveforms and generated new, targeted simulations. The simulations were generated with multiple independent codes [44–49], and they sample the posterior region for the masses and spins inferred for GW150914 [3]. Since the posteriors for the magnitudes and the orientations of the component spins are not very constraining, the choices for these parameters covered wide ranges. To validate the studies below, we added the publicly available and new NR waveforms as mock signals to the data in the neighborhood of GW150914 [36,49,50]. A further possible cause for systematics involves uncertainties in the calibration of the gravitational-strain observable in the LIGO detectors. These uncertainties are modeled and included in the results presented here according to the treatment detailed in Ref. [3].

Residuals after subtracting the most-probable waveform model.—The burst analysis [51], which looks for unmodeled transients and hence does not rely on theoretical signal templates, can be used to test the consistency of GW150914 with waveform models derived from GR. Using the LALINFEREN CE [52] Bayesian-inference software library, we identify the most probable (i.e., maximum a posteriori, henceforth MAP) binary black-hole waveform [3], compute its effect in the Livingston and Hanford detectors, and then subtract it from the data. If the data are consistent with the theoretical signal, no detectable power should remain after subtraction other than what is consistent with instrumental noise. We analyze the residual with the BAYESWAVE [53] algorithm developed to characterize generic GW transients. BAYESWAVE uses the evidence ratio (the Bayes factor) to rank competing hypotheses given the observed data. We compare predictions from models in which (i) the data contain only Gaussian noise; (ii) the data contain Gaussian noise and uncorrelated noise transients, or glitches; and (iii) the data contain Gaussian noise and an elliptically polarized GW signal. We compute the signal-to-noise Bayes factor, which
is a measure of significance for the excess power in the data, and the signal-to-glitch Bayes factor, which measures the coherence of the excess power between the two detectors.

Our analysis reveals that the GW150914 residual favors the instrumental noise hypothesis over the presence of a coherent signal as well as the presence of glitches in either detectors; see the dashed lines in the top panel of Fig. 1. The positive Bayes factor for the signal-to-glitch hypotheses indicates that the data prefer the presence of a coherent signal over glitches; nevertheless, the signal remains below common significance thresholds, as indicated by the limit on the residual \( \text{SNR}_{\text{res}} \) given in the lower panel of Fig. 1 and further explained below. This is an indication of the stability of the LIGO detectors at the time of GW150914. We also apply the same analysis to 100 4-s long segments of data drawn within a few minutes of GW150914, and produce the cumulative distribution functions of Bayes factors shown in the upper panel of Fig. 1. We find that, according to the burst analysis, the GW150914 residual is not statistically distinguishable from the instrumental noise recorded in the vicinity of the detection, suggesting that all of the measured power is well represented by the GR prediction for the signal from a binary black-hole merger. The results of this analysis are very similar regardless of the MAP waveform used (i.e., EOBNR or IMRPHENOM).

We compute the 95% upper bound on the coherent network \( \text{SNR}_{\text{res}} \). This upper bound is \( \text{SNR}_{\text{res}} \leq 7.3 \) at 95% confidence, regardless of the MAP waveform used (i.e., EOBNR or IMRPHENOM). We note that this coherent-burst SNR has a different meaning compared to the (modeled) matched-filtering binary-coalescence SNR of 24 cited for GW150914. Indeed, the upper-limit \( \text{SNR}_{\text{res}} \) inferred for GW150914 lies in the typical range for the data segments around GW150914 (see the bottom panel of Fig. 1), so it can be attributed to instrument noise alone. If we assume that \( \text{SNR}_{\text{res}} \) is entirely due to the mismatch between the MAP waveform and the underlying true signal, and that the putative violation of GR cannot be reabsorbed in the waveform model by biasing the estimates of the physical parameters [54,55], we can constrain the minimum fitting factor (FF) [56] between the MAP model and GW150914. An imperfect fit to the data leaves \( \text{SNR}_{\text{res}}^2 = (1 - \text{FF}^2) \text{SNR}_{\text{det}}^2 \), where \( \text{SNR}_{\text{det}} = 25.3^{+0.1}_{-0.2} \) is the network SNR inferred by LALInference [3]. \( \text{SNR}_{\text{res}} \leq 7.3 \) then implies \( \text{FF} \geq 0.96 \). Considering that, for parameters similar to those inferred for GW150914, our waveform models have much higher FFs against numerical GR waveforms, we conclude that the noise-weighted correlation between the observed strain signal and the true GR waveform is \( \geq 96\% \). This statement can be read as implying that the GR prediction for GW150914 is verified to be better than 4%, in a precise sense related to noise-weighted signal correlation, and, conversely, that effects due to GR violations in GW150914 are limited to less than 4% (for effects that cannot be reabsorbed in a redefinition of physical parameters).

**Inspiral-merger-ringdown consistency test.**—We now perform a test to show that the entire GW150914 waveform does not deviate from the predictions of a binary black-hole coalescence in GR. One way to do that is to compare the estimates of the mass and spin of the remnant obtained from the low-frequency and high-frequency parts of the waveform, using the relations between the binary’s components and final masses and the spins provided by NR [59].

For the purpose of this test, we choose \( f_{\text{GW}}^{\text{end insp}} = 132 \) Hz as the frequency at which the late-inspiral phase ends. In Fig. 2 we plot the EOBNR MAP waveform [3] and its 90% credible intervals, as well as the corresponding instantaneous frequency; the vertical line marks \( f_{\text{GW}}^{\text{end insp}} \).

Figure 3 shows the frequency-domain MAP waveform amplitude; note that 132 Hz lies just before what is generally denoted as the merger-ringdown phase in the frequency domain.

To perform the test, we first truncate the frequency-domain representation of the waveforms to lie between...
20 Hz to $f_{\text{end insp}}^{\text{GW}}$, and we estimate the posterior distributions of the binary’s component masses and spins using this “inspiral” (low-frequency) part of the observed signal, using the nested-sampling algorithm in the LALINFERENCE software library [52]. We then use formulas obtained from NR simulations to compute posterior distributions of the remnant’s mass and spin. Next, we obtain the complementary “postinspiral” (high-frequency) signal, which is dominated by the contribution from the merger and ringdown stages, by restricting the frequency-domain representation of the waveforms to extend between $f_{\text{end insp}}^{\text{GW}}$ and 1024 Hz. Again, we derive the posterior distributions of the component masses and spins, and (by way of NR-derived formulas) of the mass and spin of the final compact object. We note that the MAP waveform has an expected SNR$_{\text{det}} \sim 19.5$ if we truncate its frequency-domain representation to have support between 20 and 132 Hz, and $\sim 16$ if we truncate it to have support between 132 and 1024 Hz. Finally, we compare these two estimates of the final $M_f$ and dimensionless spin $a_f$ and compare them against the estimate performed using full inspiral-merger-ringdown waveforms. In all cases, we average the posteriors obtained with the EOBNR and IMRPHENOM waveform models, following the procedure outlined in Ref. [3]. Technical details about the implementation of this test can be found in Ref. [60].

This test is similar in spirit to the $\chi^2$ GW search statistic [2,61], which divides the model waveform into frequency bands and checks to see that the SNR accumulates as expected across those bands. Large matched-filter SNR values which are accompanied by a large $\chi^2$ statistic are very likely due either to noise glitches or to a mismatch between the signal and the model matched-filter waveform. Conversely, reduced-$\chi^2$ values near unity indicate that the data are consistent with waveform plus the expected detector noise. Thus, large $\chi^2$ values are a warning that some parts of the waveform are a much worse fit than others, and thus the candidates may result from instrument glitches that are very loud, but they do not resemble binary-inspiral signals. However, $\chi^2$ tests are performed by comparing the data with a single theoretical waveform, while in this case we allow the inspiral and postinspiral partial waveforms to select different physical parameters. Thus, this test should be sensitive to subtler deviations from the predictions of GR.

In Fig. 4 we summarize our findings. The top panel shows the posterior distributions of $M_f$ and $a_f$ estimated from the inspiral and postinspiral signals, and from the entire inspiral-merger-ringdown waveform. The plot confirms the expected behavior: the inspiral and postinspiral 90% confidence regions (defined by the isoprobability contours that enclose 90% of the posterior) have a significant region of overlap. As a sanity check (which, strictly speaking, is not part of the test of GR that is being performed), we also produced the 90% confidence region computed with the full inspiral-merger-ringdown waveform; it lies comfortably within this overlap. We have verified that these conclusions are not affected by the specific formula [40,59,62] used to predict $M_f$ and $a_f$, or by the choice of $f_{\text{GW}}^{\text{end insp}}$ within $\pm 50$ Hz.
contour that encloses 68% of the posterior roughly 68% of the time, as expected from random noise fluctuations. By contrast, our test can rule out the null hypothesis (with high statistical significance) when analyzing a simulated signal that reflects a significant GR violation in the frequency dependence of the energy and angular momentum loss [60], even when we choose violations which would be too small to be noticeable in double-pulsar observations [12]; for an explicit example, we refer to Fig. 1 of Ref. [60]. This includes signals with a $\chi^2$ value close to unity, so that they would not have been missed by the modeled-signal searches. Thus, our inspiral-merger-ringdown test shows no evidence of discrepancies with the predictions of GR.

The component masses and spins estimated in Ref. [3], together with NR-derived relations, imply $M_f = 68^{+4}_{-4} M_{\odot}$ (62$^{+4}_{-4} M_{\odot}$ in the source frame) and $a_f = 0.67^{+0.05}_{-0.07}$ at 90% confidence. From the posterior distributions of the mass and spin of the final black hole, we can predict the frequency and decay time of the least-damped QNM (i.e., the $\ell = 2$, $m = 2$, $n = 0$ overtone) [63]. We find $f_{220}^{QNM} = 251^{+8}_{-8}$ Hz and $\tau_{220}^{QNM} = 4.0^{+0.3}_{-0.3}$ ms at 90% confidence.

**Testing for the least-damped QNM in the data.**—We perform a test to check the consistency of the data with the predicted least-damped QNM of the remnant black hole. For this purpose, we compute the Bayes factor between a damped-sinusoid waveform model and Gaussian noise, and we estimate the corresponding parameter posteriors. The signal model used is $h(t \geq t_0) = A e^{-(t-t_0)/\tau} \cos[2\pi f_0(t-t_0) + \phi_0]$, $h(t < t_0) = 0$, with a fixed starting time $t_0$, and uniform priors over the unknown frequency $f_0 \in [200, 300]$ Hz and damping time $\tau \in [0.5, 20]$ ms. The prior on amplitude $A$ and phase $\phi_0$ is chosen as a two-dimensional Gaussian isotropic prior in $\{A_r \equiv -A \sin \phi_0, A_\phi \equiv A \cos \phi_0\}$ with a characteristic scale $H$, which is in turn marginalized over the range $H \in [2, 10] \times 10^{-22}$ with a prior $\propto 1/H$. This is a practical choice that encodes relative ignorance about the detectable damped-sinusoid amplitude in this range. We use 8 s of data (centered on GW150914) from both detectors, bandpassed to [20, 1900] Hz. The data are analyzed coherently, assuming the signal arrived 7 ms earlier at Livingston compared to Hanford, and the amplitude received in the two detectors has an approximately equal magnitude and opposite sign (as seen in, e.g., Fig. 1 of Ref. [1]).

We compute the Bayes factor and posterior estimates of $\{f_0, \tau\}$ as a function of the unknown QNM start time $t_0$, which we parametrize as an offset from a fiducial GPS merger time $t_M = 1126259462.423$ s (at the LIGO Hanford site). (The merger time is obtained by taking the EOBNR MAP waveform and lining this waveform up with the data such that the largest SNR is obtained. The merger time is then defined as the point at which the quadrature sum of the $h_+$ and $h_\times$ polarizations is maximum.) Figure 5 shows the 90% credible contours in the

---

**FIG. 4.** (Top panel) 90% credible regions in the joint posterior distributions for the mass $M_f$ and dimensionless spin $a_f$ of the final compact object as determined from the inspiral (dark violet, dashed curve) and postinspiral (violet, dot-dashed curve) signals, and from a full inspiral-merger-ringdown analysis (black curve). (Bottom panel) Posterior distributions for the parameters $\Delta M_f/M_f$ and $\Delta a_f/a_f$ that describe the fractional difference in the estimates of the final mass and spin from inspiral and postinspiral signals. The contour shows the 90% confidence region. The plus symbol indicates the expected GR value (0, 0).
The con...ing time of the MAP waveform for GW150914. The black solid line shows the 90\% credible region for the frequency and decay time of the $\ell = 2$, $m = 2$, $n = 0$ (i.e., the least-damped) QNM, as derived from the posterior distributions of the remnant mass and spin parameters.

\[
\{f_0, \tau\} \text{ plane as a function of the merger-to-start time offset } t_0 - t_M, \text{ as well as the corresponding contour for the least-damped QNM, as predicted in GR for the remnant mass and spin parameters estimated for GW150914.}
\]

The 90\% posterior contour starts to overlap with the GR prediction from the IMR waveform for $t_0 = t_M + 3$ ms, or $\sim 10M$ after the merger. The corresponding log Bayes factor at this point is $\log_{10} B \sim 14$ and the MAP waveform SNR is $\sim 8.5$. For $t_0 = t_M + 5$ ms, the MAP parameters fall within the contour predicted in GR for the least-damped QNM, with $\log_{10} B \sim 6.5$ and SNR $\sim 6.3$. At $t_0 = t_M + 6.5$ ms, or about $20M$ after merger, the Bayes factor is $\log_{10} B \sim 3.5$ with SNR $\sim 4.8$. The signal becomes undetectable shortly thereafter, for $t_0 \gtrsim t_M + 9$ ms, where $B \lesssim 1$.

Measuring the frequency and decay time of one damped sinusoid in the data does not by itself allow us to conclude that we have observed the least-damped QNM of the final black hole since the measured quality factor could be biased by the presence of the other QNMs in the ringdown signal (see, e.g., Refs. [63,64] and the references therein). However, based on the numerical simulations discussed in Refs. [65–67], one should expect the GW frequency to level off at $10M \rightarrow 20M$ after the merger, which is where the description of ringdown in terms of QNMs becomes valid. For a mass $M \sim 68M_{\odot}$, the corresponding range is $\sim 3 \rightarrow 7$ ms after merger. Since this is where we observe the 90\% posterior contours of the damped-sinusoid waveform model and the 90\%-confidence region estimated from the IMR waveform to be consistent with each other, we may conclude that the data are compatible with the presence of the least-damped QNM, as predicted by GR.

In the future, we will extend the analysis to two damped sinusoids and will explore the possibility of independently extracting the final black hole’s mass and spin. A test of the general-relativistic no-hair theorem [68,69] requires the identification of at least two QNM frequencies in the ringdown waveform [64,70,71]. Such a test would benefit from the observation of a system with a total mass similar to the one of GW150914, but with a larger asymmetry between component masses, which would increase the amplitudes of the subdominant modes; a stronger misalignment of the orbital angular momentum with the line of sight would further improve their visibility [70]. Finally, the determination of the remnant mass and spin independent of binary component parameters will allow us to test the second law of black-hole dynamics [72,73].

Constraining parametrized deviations from general-relativistic inspiral-merger-ringdown waveforms.— Because GW150914 was emitted by a binary black hole in its final phase of rapid orbital evolution, its gravitational phasing (or phase evolution) encodes nonlinear conservative and dissipative effects that are not observable in binary pulsars, whose orbital period changes at an approximately constant rate. (Current binary-pulsar observations do constrain conservative dynamics at 1 PN order and they partially constrain spin-orbit effects at 1.5 PN order through geodetic spin precession [12]) Those effects include tails of radiation due to backscattering of GWs by the curved background around the coalescing black holes [74], nonlinear tails (i.e., tails of tails) [75], couplings between black-hole spins and the binary’s orbital angular momentum, interactions between the spins of the two bodies [76–78], and excitations of QNMs [28–30] as the remnant black hole settles in the stationary configuration.

Whether all of these subtle effects can actually be identified in GW150914 and tested against GR predictions depends, of course, on their strength with respect to instrument noise and on whether the available waveform models are parametrized in terms of those physical effects. GW150914 is moderately loud, with SNR $\sim 24$, certainly much smaller than what can be achieved in binary-pulsar observations. Our ability to analyze the fine structure of the GW150914 waveform is correspondingly limited. Our approach is to adopt a parametrized analytical family of inspiral-merger-ringdown waveforms, then treat the waveform coefficients as free variables that can be estimated (either individually or in groups) from the GW150914 data [79–85]. We can then verify that the posterior probability distributions for the coefficients include their GR values.

The simplest and fastest parametrized waveform model that is currently available [41] can be used to bound physical effects only for the coefficients that enter the early-inspiral phase because, for the late-inspiral, merger, and ringdown phases, it uses phenomenological coefficients fitted to NR waveforms. Louder GW events, to be collected as detector
sensitivity improves, and more sophisticated parametrized waveform models, will allow us to do much more stringent and physical tests targeted at specific relativistic effects. We work within a subset of the TIGER framework [85,86] and perform a null-hypothesis test by comparing GW150914 with a generalized, analytical inspiral-merger-ringdown waveform model (henceforth, gIMR) that includes parameterized deformations with respect to GR. In this framework, deviations from GR are modeled as fractional changes \( \{\delta \hat{p}_i\} \) in any of the parameters \( \{p_i\} \) that parametrize the GW phase expression in the baseline waveform model. Similar to Refs. [85,86], we only consider deviations from GR in the GW phase, while we leave the GW amplitude unperturbed. Indeed, at the SNR of GW150914 (i.e., SNR ~ 24), we expect to have much higher sensitivity to the GW phase than to its amplitude. Also, amplitude deviations could be reabsorbed in the calibration error model used to analyze GW150914 [3].

We construct gIMR starting from the frequency-domain IMRPHENOM waveform model. The dynamical stages that characterize the coalescence process can be represented in the frequency domain by plotting the absolute value of the waveform’s amplitude. We review those stages in Fig. 3 to guide the reader toward the interpretation of the results that are summarized in Table I and Figs. 6 and 7. We refer to the early-inspiral stage as the PN part of the GW phase. This stage of the phase evolution is known analytically up to \( (v/c)^7 \) and it is parametrized in terms of the PN coefficients \( \varphi_j, j = 0, \ldots, 7 \) and the logarithmic terms \( \varphi_j, j = 5, 6 \). The late-inspiral stage, parametrized in terms of \( \sigma_j, j = 1, \ldots, 4 \), is defined as the phenomenological extension of the PN series to \( (v/c)^11 \). The early- and late-inspiral stages are denoted simply as inspiral both in Ref. [41] and in Fig. 3. The intermediate stage that models the transition between the inspiral and the merger-ringdown phase is parametrized in terms of the phenomenological coefficients \( \beta_j, j = 1, 2, 3 \). Finally, the merger-ringdown phase is parametrized in terms of the phenomenological coefficients \( \alpha_j, j = 1, 2, 3 \). The \( \beta_j \)'s and \( \alpha_j \)'s aim to capture the frequency dependencies of the phase of the corresponding regimes; see the column labeled “\( f \) dependence” in Table I. Because of the procedure through which the model is constructed, which involves fitting a waveform phasing ansatz to a calibration set of EOB waveforms joined to NR waveforms [41], there is an intrinsic uncertainty in the values of the phenomenological parameters of the IMRPHENOM model. For the intermediate and merger-ringdown regime, we verified that these intrinsic uncertainties are much smaller than the corresponding statistical uncertainties for GW150914 and thus do not affect our conclusions. In the late-inspiral case, the uncertainties associated with the calibration of the \( \sigma_j \) parameters are large and almost comparable to the statistical measurement uncertainties. For this reason, we do not report results for the \( \sigma_j \) parameters.

As mentioned, we construct the gIMR model by introducing (fractional) deformations \( \delta \hat{p}_i \) for each of the IMRPHENOM phase parameters \( p_i \), which dominate the evolution of the phase at the different stages in the coalescence explained above. At each point in parameter

<table>
<thead>
<tr>
<th>Waveform regime</th>
<th>Parameter</th>
<th>( f ) dependence</th>
<th>Median</th>
<th>GR quantile</th>
<th>log_{10}B_{GR_{model}}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Single</td>
<td>Multiple</td>
<td>Single</td>
</tr>
<tr>
<td>Early-inspiral</td>
<td>( \delta \hat{p}_0 )</td>
<td>( f^{-3/3} )</td>
<td>( -0.1^{+0.1}_{-0.1} )</td>
<td>( 1.4^{+2.3}_{-3.0} )</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>( \delta \hat{p}_1 )</td>
<td>( f^{-4/3} )</td>
<td>( 0.3^{+0.4}_{-0.4} )</td>
<td>( -0.4^{+0.7}_{-0.7} )</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>( \delta \hat{p}_2 )</td>
<td>( f^{-1} )</td>
<td>( -0.35^{+0.35}_{-0.35} )</td>
<td>( -3.2^{+19.4}_{-10.3} )</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>( \delta \hat{p}_3 )</td>
<td>( f^{-2/3} )</td>
<td>( 0.2^{+0.2}_{-0.2} )</td>
<td>( 2.6^{+15.7}_{-17.3} )</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>( \delta \hat{p}_4 )</td>
<td>( f^{-1/3} )</td>
<td>( -2.0^{+1.6}_{-1.8} )</td>
<td>( 0.5^{+18.2}_{-18.2} )</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>( \delta \hat{p}_{51} )</td>
<td>log ( f )</td>
<td>( 8.6^{+0.5}_{-0.5} )</td>
<td>( -1.5^{+19.1}_{-16.3} )</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>( \delta \hat{p}_{61} )</td>
<td>( f^{1/3} )</td>
<td>( -1.5^{+1.1}_{-1.1} )</td>
<td>( -0.6^{+17.2}_{-17.2} )</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>( \delta \hat{p}_{62} )</td>
<td>( f^{1/3} )</td>
<td>( 8.9^{+6.8}_{-6.8} )</td>
<td>( -2.4^{+18.7}_{-15.2} )</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>( \delta \hat{p}_{7} )</td>
<td>( f^{-3/3} )</td>
<td>( 3.7^{+7.6}_{-7.5} )</td>
<td>( -3.4^{+19.3}_{-14.8} )</td>
<td>0.02</td>
</tr>
<tr>
<td>Intermediate</td>
<td>( \delta \hat{p}_{2} )</td>
<td>log ( f )</td>
<td>( 0.1^{+0.4}_{-0.4} )</td>
<td>( 0.1^{+0.5}_{-0.5} )</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>( \delta \hat{p}_{3} )</td>
<td>( f^{-3} )</td>
<td>( 0.1^{+0.5}_{-0.5} )</td>
<td>( -0.0^{+0.8}_{-0.6} )</td>
<td>0.38</td>
</tr>
<tr>
<td>Merger-ringdown</td>
<td>( \delta \alpha_2 )</td>
<td>( f^{-1} )</td>
<td>( -0.1^{+0.4}_{-0.4} )</td>
<td>( -0.0^{+1.0}_{-1.5} )</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>( \delta \alpha_3 )</td>
<td>( f^{-3/4} )</td>
<td>( -0.5^{+2.0}_{-2.0} )</td>
<td>( -0.0^{+4.4}_{-4.4} )</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>( \delta \alpha_4 )</td>
<td>tan(^{-1}(af+b))</td>
<td>( -0.1^{+0.5}_{-0.5} )</td>
<td>( -0.0^{+12.3}_{-11.1} )</td>
<td>0.61</td>
</tr>
</tbody>
</table>
parameters is allowed to vary freely (in addition to masses, spins, etc.), while the remaining ones are fixed to their GR value, that is zero, and a *multiple-parameter* analysis in which all of the parameters in one of the three sets enumerated above are allowed to vary simultaneously.

The rationale behind our choices of single- and multiple-parameter analyses comes from the following considerations. In most known alternative theories of gravity [13,14,88], the corrections to GR extend to all PN orders even if, in most cases, they have been computed only at leading PN order. Considering that GW150914 is an inspiral-merger-ringdown signal sweeping through the detector between 20 and 300 Hz, we expect to see signal deviations from GR at all PN orders. The single-parameter analysis corresponds to minimally extended models that can capture deviations from GR that occur predominantly, but not only, at a specific PN order. Nevertheless, should a deviation be measurably present at multiple PN orders, we expect the single-parameter analyses to also capture these. In the multiple-parameter analysis, the correlations among the parameters are very significant. In other words, a shift in one of the testing parameters can always be compensated for by a change of the opposite sign in another parameter and still return the same overall GW phase. Thus, it is not surprising that the multiple-parameter case provides a much more conservative statement on the agreement between GW150914 and GR. We defer to future studies the identification of optimally determined directions in the \( \delta \hat{p}_i \) space by performing a singular value decomposition along the lines suggested in Ref. [89].

For each set of testing parameters, we perform a separate LALINFERECE analysis, where, in concert with the full set of GR parameters [3], we also explore the posterior distributions for the specified set of testing parameters. Since our testing parameters are purely phenomenological (except the parameters that govern the PN early-inspiral stage), we choose their prior probability distributions to be uniform and wide enough to encompass the full posterior probability density function in the single-parameter case. Specifically, we set \( \delta \hat{p}_i \in [-20,20]; \delta \hat{\alpha}_i \in [-3,3]; \delta \hat{\delta}_i \in [-5,5] \). In all cases, we obtain estimates of the physical parameters—e.g., masses and spins—that are in agreement with those reported in Ref. [3].

In Fig. 6 we show the 90% upper bounds on deviations in the (known) PN parameters, \( \delta \hat{p}_i \) with \( i = 0, \ldots , 7 \) (except for \( i = 5 \), which is degenerate with the reference phase), when varying the testing parameters one at the time, keeping the other parameters fixed to the GR value. As an illustration, following Ref. [87], we also show in Fig. 6 the bounds obtained from the measured orbital-period derivative \( \dot{P}_{\text{orb}} \) of the double pulsar J0737-3039 [12]. Also, for the latter, bounds are computed by allowing for possible violations of GR at different powers of frequency, one at a time. Not surprisingly, since in binary pulsars the orbital period changes at essentially a constant rate, the corresponding

---

**FIG. 6.** 90% upper bounds on the fractional variations of the known PN coefficients with respect to their GR values. The orange squares are the 90% upper bounds obtained from the single-parameter analysis of GW150914. As a comparison, the blue triangles show the 90% upper bounds extrapolated exclusively from the measured orbital-period derivative \( \dot{P}_{\text{orb}} \) of the double pulsar J0737-3039 [12,87], here, too, allowing for possible GR violations at different powers of frequency, one at a time. The GW phase deduced from an almost constant \( P \_\text{orb} \) of GW150914 is increased, so we show the bounds for the latter only up to 1PN order. We do not report on the deviation of the 2.5 PN coefficient, which is unmeasurable because it is degenerate with the reference phase. We also do not report on the deviations of the logarithmic terms in the PN series at 2.5 PN and 3 PN order, which can be found in Table I and in Fig. 7.

space, the coefficients \( p_i \) are evaluated for the local physical parameters (masses, spins) and multiplied by factors \( (1 + \delta \hat{p}_i) \). When using such waveforms as templates, the parameters that are allowed to vary freely are then the ones that are also present in the GR waveforms (masses, spins, sky position, orientation, distance, and a reference time and phase), together with one or more of the \( \delta \hat{p}_i \)'s; the \( p_i \)'s themselves are calculated using their GR expressions in terms of masses and spins. In this parameterization, GR is uniquely defined as the locus in the parameter space where all of the testing parameters \( \delta \hat{p}_i \) are zero. In summary, our battery of testing parameters consists of (i) the early-inspiral stage: \( \{ \delta \hat{\delta}_0, \delta \hat{\delta}_1, \delta \hat{\delta}_2, \delta \hat{\delta}_3, \delta \hat{\delta}_6, \delta \hat{\delta}_4, \delta \hat{\delta}_7 \} \), (ii) the intermediate regime: \( \{ \delta \hat{\delta}_2, \delta \hat{\delta}_4 \} \), and (iii) the merger-ringdown regime: \( \{ \delta \hat{\delta}_2, \delta \hat{\alpha}_3, \delta \hat{\delta}_4 \} \). (Unlike Ref. [41], we explicitly include the logarithmic terms \( \delta \hat{\phi}_{5i} \) and \( \delta \hat{\phi}_{6i} \). We also include the 0.5 PN parameter \( \delta \hat{\phi}_1 \); since \( \phi_1 \) is zero in GR, we define \( \delta \hat{\phi}_1 \) to be an absolute shift rather than a fractional deformation.) We do not consider parameters that are degenerate with either the reference time or the reference phase. For our analysis, we explore two scenarios: a *single-parameter* analysis in which only one of the testing
bounds quickly become rather loose as the PN order is increased. As a consequence, the double-pulsar bounds are significantly less informative than GW150914, except at 0 PN order, where the double-pulsar bound is better thanks to the long observation time (∼10 yr against ∼0.4 s for GW150914). (We note that when computing the upper bounds with the binary-pulsar observations, we include the effect of eccentricity only in the 0 PN parameter. For the higher PN parameters, the effect is not essential considering that the bounds are not very tight.) Thus, GW150914 allows us for the first time to constrain the coefficients in the PN series of the phasing up to 3.5 PN order.

Furthermore, in Table I and Fig. 7 we summarize the constraints on each testing parameter δΦi for the single- and multiple-parameter analyses. In particular, in the sixth and seventh columns of Table I, we list the quantile at which the GR value of zero is found within the marginalized one-dimensional posterior (i.e., the integral of the posterior from the lower bound of the prior up to zero). We note that in the single-parameter analysis, for several parameters, the GR value is found at quantiles close to an equivalent of 2σ − 2.5σ, i.e., close to the tails of their posterior probability functions. It is not surprising that this should happen for the majority of the early-inspiral parameters since we find that these parameters have a substantial degree of correlation. Thus, if a particular noise realization causes the posterior distribution of one parameter to be off centered with respect to zero, we expect that the posteriors of all of the other parameters will also be off centered. This is indeed what we observe. The medians of the early-inspiral single-parameter posteriors reported in Table I show opposite sign shifts that follow closely the sign pattern found in the PN series.

We repeated our single-parameter analysis on 20 data sets obtained by adding the same NR waveform with GW150914-like parameters to different noise-only data segments close to GW150914. In one instance, we observed δΦi posterior distributions very similar to those of Table I and Fig. 7, both in terms of their displacements from zero and of their widths, whereas for the others the displacements tended to be much smaller (though the widths were still comparable). Thus, it is not unlikely that instrumental noise fluctuations would cause the degree of apparent deviation from GR found to occur in the single-parameter quantiles for GW150914, even in the absence of an actual deviation from GR. However, we cannot fully exclude a systematic origin from inaccuracies or even missing physics in our waveform models. Future observations will shed light on this aspect.

In the multiple-parameter analysis, which accounts for correlations between parameters, the GR value is usually found to be very close to the median of the marginalized distributions. This is partly due to the fact that we are not sensitive to most of the early-inspiral parameters, with the exception of the 0PN and 0.5PN coefficients. As for the intermediate and merger-ringdown parameters, since most of the SNR for GW150914 comes from the high-frequency portion of the observed signal, we find that the constraints on those coefficients are very robust and essentially independent of the analysis configuration chosen, single or multiple.

Finally, the last two columns of Table I report the logarithm of the ratio of the marginal likelihoods (the logarithm of the Bayes factor \( \log_{10}(B_{\text{model}}) \)) as a measure of the relative goodness of fit between the IMRPHENOM and cIMR models (see Ref. [3] and the references therein). If
log\_10 \beta_{model}^{GR} < 0(>0), then GR fits the data worse (better) than the competing model. The uncertainty over log\_10 \beta_{model}^{GR} is estimated by running several independent instances of LALINFERENCe. The log\_10 \beta_{model}^{GR} values shown in Table I corroborate our finding that GW150914 provides no evidence in support of the hypothesis that GR is violated [90].

As an aside, we note that GW150914 was detected with the LIGO detectors at about one third of their final design sensitivity, which is expected to be achieved around 2019 [19]. Hence, future detections are expected to occur with larger SNRs, leading to tighter bounds on the phase coefficients. It is also worth noting that the posterior density functions for the δ\_p from all future detections can be combined, leading to a progressive improvement of the bounds on these parameters.

Constraining the graviton Compton wavelength.—Since the 1970s, there have been attempts to construct theories of gravity mediated by a graviton with a nonzero mass. Those attempts have led to conceptual difficulties; some of these have been addressed, circumvented, or overcome, but others remain open (see Ref. [91] and the references therein). Here, we take a phenomenological approach and consider a hypothetical massive-graviton theory in which, due to a modification of the dispersion relation, GWs travel at a speed different than the speed of light.

In GR, gravitons are massless and travel at the speed of light, v\_g = c. In a massive-graviton theory, the dispersion relation can be modified to E\^2 = p\^2 c\^2 + m^2\_g c\^4, where E is the graviton energy, p the momentum, and m\_g the graviton rest mass, related to the graviton’s Compton wavelength by λ\_g = h/(m\_g c), with h the Planck constant. Thus, we have v\^2\_g/c\^2 = c\^2 p\^2/E\^2 = 1 - h\^2 c\^2/(λ\^2\_g c\^4), and the massive graviton propagates at an energy (or frequency) dependent speed. Another effect one expects on general grounds is that the Newtonian potential gets altered by a Yukawa-type correction whose characteristic length scale is λ\_y; \( \varphi(r) = (GM/r)[1 - \exp(-r/\lambda_y)] \).

Existing bounds on λ\_y that do not probe the propagation of gravitational interactions (i.e., the so-called static bounds), come from Solar System observations [92,93] (which probe the above Yukawa-corrected Newtonian potential), the nonobservation of superradiant instabilities in supermassive black holes [94], model-dependent studies of the large-scale dynamics of galactic clusters [95], and weak lensing observations [96]; these bounds are 2.8 × 10\^12, 2.5 × 10\^13, 6.2 × 10\^19, and 1.8 × 10\^22 km, respectively. We note that the bound from superradiance relies on the assumption that the very massive, compact objects in the centers of galaxies are indeed supermassive Kerr black holes, as opposed to other, more exotic objects. As was also stressed in Ref. [93], the model-dependent bounds from clusters and weak lensing should be taken with caution, in view of the uncertainties on the amount of dark matter in the Universe and its spatial distribution. The only dynamical bound to date comes from binary-pulsar observations [97] and it is \( \lambda_y > 1.6 \times 10^{10} \text{ km} \). If the Compton wavelength of gravitons is finite, then lower frequencies propagate slower compared to higher frequencies, and this dispersion of the waves can be incorporated into the gravitational phase from a coalescing binary. Specifically, neglecting all possible effects on the binary dynamics that could be introduced by the massive-graviton theory, Ref. [93] found that the phase term \( \Phi_{MG}(f) = -\pi Dc/(\lambda^2_y [1 + z] f) \) (formally a 1PN-order term) should be added to the overall GW phase. In this expression, z is the cosmological redshift and D is a cosmological distance defined in Eq. (2.5) of Ref. [93].

GW150914 allows us to search for evidence of dispersion as the signal propagated toward Earth. We perform the analysis by explicitly including the formally 1PN-order term above [93,98] in the EOBNR and IMRPHENOM GW phases and treating λ\_y as an additional, independent parameter [99]. We assume a standard Λ cold dark matter cosmology [100] and a uniform prior probability on the graviton mass m\_g \( \in [10^{-26}, 10^{-16}] \text{ eV}/c^2 \), thus the prior on λ\_y is \( \propto 1/\lambda^2_y \). In Fig. 8 we show the cumulative posterior probability distribution for λ\_y obtained from combining the results of the two waveforms (EOBNR and IMRPHENOM) following the procedure outlined in Ref. [3]. We find no evidence for a finite value of λ\_y, and we derive a dynamical lower bound, λ\_y > 10\^13 km at 90% confidence, which corresponds to a graviton mass m\_g > 1.2 × 10^{-22} eV/c\^2. This bound is approximately a factor of 3 better than the current Solar System bound [92,93], and ~3 orders of magnitude better than the bound from binary-pulsar observations [97], but it is less constraining than model-dependent bounds coming from the large-scale dynamics of galactic clusters [95], weak gravitational-lensing observations [96], and the nonobservation of superradiant instability in supermassive black holes [94].

No constraint on non-GR polarization states.—GR predicts the existence of two transverse-traceless tensor polarizations for GWs. More general metric theories of gravitation allow for up to four additional polarization states: a transverse scalar mode and three longitudinal modes [13,101]. Because the Hanford and Livingston LIGO instruments have similar orientations, they are sensitive to a very similar linear combination of the GW polarizations, so it is difficult to distinguish between the GR and non-GR states.

As an illustration, we use the BAYESWAVE GW-transient analysis algorithm [53] to reconstruct the GW150914 waveform, assuming the simplest case in which the signal consists entirely of the transverse scalar (breathing) mode. We compare the reconstructed waveforms and power spectral densities (PSDs) for the pure scalar mode and GR models, and we find the log Bayes factor between the
two hypotheses to be \( \log B_{\text{GR}}^{\text{scalar}} = 1.3 \pm 0.5 \) when using the PSD from the breathing mode analysis and \( \log B_{\text{GR}}^{\text{scalar}} = -0.2 \pm 0.5 \) when using the PSD from the GR analysis. In both cases, the log Bayes factors do not significantly favor one model over the other. The only notable difference is in the reconstructed sky locations; the latter reflects the different response of the detector network to the tensor components compared to the purely scalar mode.

We reiterate that this test is only meant to illustrate the difficulty in distinguishing between GR and non-GR polarization states on the basis of GW150914 data alone. Furthermore, the results are not in contradiction with the comprehensive parameter estimation studies of GW150914 [3], which model only the transverse-traceless GR polarizations. Finally, we note that in the weakly dynamical regime, binary pulsars [12] do provide evidence in favor of GR, in that they would have a different decay rate if scalar radiation were to dominate. To directly study the polarization content of gravitational radiation from the strong-field dynamics, a larger network including detectors with different orientations, such as Advanced Virgo [102], KAGRA [103], and LIGO-India [104], will be required, at least in the context of unmodeled GW-signal reconstruction.

**Outlook.**—The observation of GW150914 has given us the opportunity to perform quantitative tests of the genuinely strong-field dynamics of GR. We investigated the nature of GW150914 by performing a series of tests devised to detect inconsistencies with the predictions of GR. With the exception of the graviton Compton wavelength and the test for the presence of a non-GR polarization, we did not perform any studies aimed at constraining parameters that might arise from specific alternative theories [13,14,88], such as Einstein-æther theory [105] and dynamical Chern-Simons theory [106], or from compact-object binaries composed of exotic objects such as boson stars [107] and gravastars [108]. Studies of this kind are not yet possible since we lack predictions for what the inspiral-merger-ringdown GW signal should look like in those cases. We hope that the observation of GW150914 will boost the development of such models in the near future.

In future work we will also attempt to measure more than one damped sinusoid from the data after GW150914’s peak, thus extracting the QNMs and inferring the final black hole’s mass and spin. We will thus be able to test the no-hair theorem [68,69] and the second law of black-hole dynamics [72,73]. However, signals louder than GW150914 might be needed to achieve these goals. GR predicts the existence of only two transverse polarizations for GWs. We plan to investigate whether an extended detector network will allow the measurement of non-transverse components [13] in further GW signals.

The constraints provided by GW150914 on deviations from GR are unprecedented due to the nature of the source, but they do not reach high precision for some types of deviation, particularly those affecting the inspiral regime. A much higher SNR and longer signals are necessary for more stringent tests. However, it is not clear up to which SNR our parametrized waveform models are still a faithful representation of solutions of Einstein’s equations. Furthermore, to extract specific physical effects we need waveform models that are expressed in terms of relevant parameters. We hope that others, encouraged by GW150914, will make further efforts to develop reliable, physically relevant, and computationally fast waveform models. More stringent bounds can be obtained by combining results from multiple GW observations [60,85,86,99]. Given the rate of coalescence of binary black holes as inferred in Ref. [109], we are looking forward to the upcoming joint observing runs of LIGO and Virgo.

The detection of GW150914 ushers in a new era in the field of experimental tests of GR. The first result of this era is that, within the limits set by our sensitivity, all of the tests performed on GW150914 provided no evidence of disagreement with the predictions of GR.

The authors gratefully acknowledge the support of the U.S. National Science Foundation (NSF) for the construction and operation of the LIGO Laboratory and Advanced LIGO as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. The authors gratefully appreciate...
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, Department of Science and Technology, India, Science and Engineering Research Board (SERB), India, Ministry of Human Resource Development, India, the Spanish Ministerio de Economía y Competitividad, the Consellera d’Economia i Competitivitat and Conselleria d’Educaçó, Cultura i Universitats of the Govern de les Illes Balears, the National Science Centre of Poland, the European Commission, 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 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, and Innovation, the Russian Foundation for Basic Research, the Leverhulme Trust, the Research Corporation, 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 the provision of computational resources.

Finally, we thank the anonymous referees, whose comments helped improve the clarity of the Letter.

[90] Because of the normalization of the prior probability distributions, the Bayes factors include a penalty factor—the so-called Occam factor—for models that have more parameters. The wider the prior range for the additional parameters, the more severe the penalization. Therefore, different choices for $\delta p_1$ would lead to different numerical values of $\log B_{\text{GR}}^{\text{max}}$. To fully establish the significance of the Bayes factors, validation studies [85,86] would be necessary and will be presented in forthcoming studies.
130 Caltech JPL, Pasadena, California 91109, USA
131 Università di Siena, I-53100 Siena, Italy
132 Trinity University, San Antonio, Texas 78212, USA
133 University of Washington, Seattle, Washington 98195, USA
134 Kenyon College, Gambier, Ohio 43022, USA
135 Abilene Christian University, Abilene, Texas 79699, USA

^Deceased, May 2015.
^Deceased, March 2015.