

Testing general relativity using gravitational wave signals from the inspiral, merger and ringdown of binary black holes

Abhirup Ghosh¹, Nathan K. Johnson-McDaniel¹, Archisman Ghosh^{2,1}, Chandra Kant Mishra^{3,1}, Parameswaran Ajith¹, Walter Del Pozzo^{4,5}, Christopher P. L. Berry⁵, Alex B. Nielsen⁶, Lionel London⁷

¹ International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Bangalore 560089, India

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

³ Department of Physics, Indian Institute of Technology Madras, Chennai 600036, India

⁴ Dipartimento di Fisica “Enrico Fermi”, Università di Pisa, Pisa I-56127, Italy

⁵ School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

⁶ Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, Callinstraße 38, 30167 Hannover, Germany

⁷ School of Physics and Astronomy, Cardiff University, The Parade, Cardiff, CF24 3AA, United Kingdom

Abstract. Advanced LIGO’s recent observations of gravitational waves (GWs) from merging binary black holes have opened up a unique laboratory to test general relativity (GR) in the highly relativistic regime. One of the tests used to establish the consistency of the first LIGO event with a binary black hole merger predicted by GR was the *inspiral–merger–ringdown consistency test*. This involves inferring the mass and spin of the remnant black hole from the inspiral (low-frequency) part of the observed signal and checking for the consistency of the inferred parameters with the same estimated from the post-inspiral (high-frequency) part of the signal. Based on the observed rate of binary black hole mergers, we expect the advanced GW observatories to observe hundreds of binary black hole mergers every year when operating at their design sensitivities, most of them with modest signal to noise ratios (SNRs). Anticipating such observations, this paper shows how constraints from a large number of events with modest SNRs can be combined to produce strong constraints on deviations from GR. Using kludge modified GR waveforms, we demonstrate how this test could identify certain types of deviations from GR if such deviations are present in the signal waveforms. We also study the robustness of this test against reasonable variations of a variety of different analysis parameters.

1. Introduction

On September 14, 2015 the Advanced LIGO detectors [1, 2] made the first observation of gravitational waves (GWs) from a binary black hole system [3]. The signal, termed GW150914, was inferred to be produced by the merger of two black holes (with masses $\sim 36M_{\odot}$ and $29M_{\odot}$) at a distance of ~ 400 Mpc from the Earth [4, 5]. This observation was followed by a second one on December 26, 2015 [6]. The second signal, termed GW151226, was produced by the coalescence of two black holes less massive ($\sim 14M_{\odot}$ and $8M_{\odot}$) than those in GW150914 [5]. These observations herald the beginning of a new branch of observational astronomy. Apart from providing the first direct evidence of the existence of GWs, these observations confirmed the existence of stellar mass black holes that are much more massive than previously thought ($\gtrsim 25M_{\odot}$) [7]. They also provided the first evidence of binary black holes that inspiral under GW emission and subsequently merge. These observations also enabled us to perform the first tests of general relativity (GR) in the highly relativistic and nonlinear regime of gravity [8, 5] — a regime with binary orbital velocities $\sim 0.5c$, which is inaccessible by other astronomical observations and laboratory tests.

A handful of tests were performed to check the consistency of the observed signals with those predicted by GR. In particular, due to the large masses of the black holes in GW150914, the observed signal consists of signatures from the inspiral and merger of the two black holes and the subsequent ringdown of the final black hole. This allowed us to perform several consistency tests on GW150914 making use of one or more phases of the coalescence: The first involved testing the consistency between the mass and spin of the final remnant, determined from the low-frequency (inspiral) and high-frequency (post-inspiral) parts of the observed signal [9]. The second involved testing the consistency of the data after the peak of the observed signal (corresponding to the merger) with a quasi-normal mode (QNM) spectrum predicted by GR [10, 11, 12, 13]. The third involved bounding deviations from the GR predictions of the post-Newtonian (PN) coefficients describing the inspiral (and from phenomenological parameters describing the merger and ringdown) using parametrized waveform models [14, 15]. The fourth involved constraining the amount of dispersion in the observed GW signal and converting it to a bound on the mass of the graviton [16]. In addition, the residual after subtracting the best fit GR signal was found to be consistent with detector noise. Within the statistical uncertainties, these investigations provided no evidence for deviations from GR [8].

The Advanced LIGO detectors, which are currently operating at less than their design sensitivities, are expected to approach design sensitivity over the next few years [17]. Several observing runs will be conducted over this period. Based on the rate of GW events in the first observing run, and the anticipated improvement in the sensitivity of the instruments, we expect that GW observations will become routine. The binary black hole event rate is expected to approach several hundred detections per year when the LIGO detectors operate at their design sensitivity [18, 5]. The Advanced Virgo [19] detector in Italy and KAGRA [20] in Japan are expected to join the international network of advanced GW detectors in the next few months to years, and the LIGO-India [21] detector is planned to join this network by 2024. These additional detectors will dramatically improve the sky coverage, source localization accuracy and the polarization extraction capability of the international GW network.

Upcoming GW observations will allow us to make significant improvements to the present constraints on deviations from GR, or potentially to detect such deviations — either from a small number of particularly loud events or by combining constraints from a large number of moderate events. Thus, a careful characterization of the analysis pipelines and various systematic errors in the analysis is of vital importance, so that systematic errors in our analysis will not be mistaken for a true deviation from GR. In anticipation of the binary black hole events expected in the upcoming observational runs of Advanced LIGO and sister projects, we present here a detailed follow up of the analysis method of the *inspiral–merger–ringdown* (IMR) *consistency test* for binary black holes presented in our previous paper [9], which we will henceforth refer to as Paper I.

The IMR consistency test was developed based on the original idea by Hughes and Menou [22] proposed in the context of the space-based LISA observatory.† The key idea is to infer the mass and spin of the remnant black hole, using two different parts of the observed signal, and then to compare these independent estimates. We first estimate the initial masses and spins from the inspiral (low-frequency) part of the signal, which allows us to infer the mass and spin of the final black hole making use of fitting formulas calibrated to numerical relativity (NR) simulations of binary black holes. Next, we estimate the same parameters independently from the merger–ringdown (high-frequency) part of the signal and then compare the two estimates. If the signal is correctly described by the merger of a (quasicircular) binary of Kerr black holes in GR, which is implicit in the waveform models that we use to estimate the parameters and the fits that went in, one should expect the two estimates to be consistent with each other. On the other hand, if there is a departure from GR, depending on the exact nature of the departure, it can manifest as a discrepancy between the two estimates. In particular, if the energy and angular momentum radiated during the merger regime

† See also, Nakano, Tanaka and Nakamura [23] for a recent study in the context of a ground-based detector.

(where the gravity is extremely strong and nonlinear) differ significantly from the GR predictions for these quantities, one could expect a discrepancy between the two estimates in the final mass and spin.

The original idea by Hughes and Menou [22] was to estimate the parameters purely from the early inspiral (well described by the PN approximation to GR) and the late ringdown (well described by a spectrum of QNMs). However, such a test will be possible only using a small number of *golden* events, where both the early inspiral and late ringdown are observed with high SNRs. While such a test might be possible in the future using LISA or third generation ground-based instruments, such tests are unlikely to be possible using the current generation of GW observatories. Meanwhile, Advanced LIGO is expecting to observe several hundred binary black hole events of moderate SNR in the coming years. Hence, tests of GR in the near future are going to progress through building evidence by combining multiple events. Our formulation of the test is geared in this direction. Indeed, the transition from inspiral regime to merger regime does not happen at a precise time or frequency. We use some reasonable choice of a cutoff frequency to separate the signal to the low-frequency (inspiral) and high-frequency (merger–ringdown) parts. As part of our studies, we show that the test is robust against variations of this cutoff frequency that still give large enough SNRs in for parameter estimation in both the low- and high-frequency portions independently.

Section 2 contains a detailed description of the analysis method. We then show how a large number of GW signals from binary black holes can be combined to produce strong constraints on the deviations from GR (section 3), and how a true deviation from GR might show up in our analysis (section 4). We also demonstrate the robustness of our analysis against reasonable choices of the frequency that is used to demarcate the inspiral and post-inspiral parts of the observed signal (section 5.1), the choice of the waveform template (section 5.2), fitting formulas for the mass and spin of the final black hole (section 5.3), and the effects of precession and higher modes (section 5.4). We conclude and consider future work in section 6.

2. Testing the consistency between the inspiral, merger and ringdown

The GW signal from a spinning binary black hole with negligible orbital eccentricity is described by 15 parameters (enumerated in, e.g., section II B of [24]). If the spins of the individual black holes are aligned (or antialigned) with the direction of the orbital angular momentum, then the number of parameters reduces to 11. These parameters are typically estimated by means of Bayesian inference. Bayes' theorem states that the posterior probability distribution of a parameter set $\lambda := \{\lambda_i\}$ of a model hypothesis \mathcal{H} , given data d and any other information I , is:

$$P(\lambda|d, \mathcal{H}, I) = \frac{P(\lambda|\mathcal{H}, I) \mathcal{L}(d|\lambda, \mathcal{H}, I)}{P(d|\mathcal{H}, I)}, \quad (1)$$

where $P(\lambda|\mathcal{H}, I)$ is the *prior* probability of the parameter set λ given \mathcal{H} and I , while $\mathcal{L}(d|\lambda, \mathcal{H}, I)$ is called the *likelihood* function, which is the probability of observing the data d , given λ , \mathcal{H} and I . What appears in the denominator is a normalization constant $P(d|\mathcal{H}, I) := \int P(\lambda|\mathcal{H}, I) \mathcal{L}(d|\lambda, \mathcal{H}, I) d\lambda$, called the *marginal likelihood*, or the *evidence* of the hypothesis \mathcal{H} .

If our hypothesis \mathcal{H} is that the data contains a GW signal described by a GR waveform model $h(\lambda)$ and stationary Gaussian noise described by the power spectral density $S_n(f)$, then, as described in appendix A of [25], the likelihood function can be defined as:

$$\mathcal{L}(d|\lambda, \mathcal{H}, I) \propto \exp \left[-\frac{1}{2} \langle d - h(\lambda) | d - h(\lambda) \rangle \right], \quad (2)$$

where $\langle \cdot | \cdot \rangle$ is the following noise-weighted inner product

$$\langle B|C \rangle := 2 \int_{f_{\text{low}}}^{f_{\text{cut}}} \frac{\tilde{B}^*(f) \tilde{C}(f) + \tilde{B}(f) \tilde{C}^*(f)}{S_n(f)} df. \quad (3)$$

Above, $\tilde{B}(f)$ denotes the Fourier transform of $B(t)$ and a $*$ denotes complex conjugation. The limits of integration f_{low} and f_{cut} are dictated by the bandwidth of the detector sensitivity, the bandwidth of the signal, as well as the cutoff frequencies in our calculations (described below). Owing to the large dimensionality of the parameter set λ , the posterior distribution $P(\lambda|d, \mathcal{H}, I)$ in equation (1) is computed by stochastically sampling the parameter space using techniques such as Markov-Chain Monte Carlo [26] or nested sampling [27]. For this paper, we use the `LALINFERENCE_NEST` [28, 24] code that provides an implementation of the nested sampling algorithm in the LSC Algorithms Library [29] for computing the posterior distributions.

Using the Bayesian framework described above, we start by computing the joint posterior probability distribution on the initial masses and dimensionless spins $P(m_1, m_2, \mathbf{a}_1, \mathbf{a}_2 | d)^\dagger$ of the binary black hole system (with $m_1 \geq m_2$ by convention). This is done by marginalizing the posterior over the remaining parameters describing the signal. Assuming quasicircular inspirals, fitting formulas calibrated to NR simulations then give us predictions of the mass M_f and dimensionless spin magnitude a_f of the final (remnant) black hole as functions of the initial masses and spins,

$$M_f = M_f(m_1, m_2, \mathbf{a}_1, \mathbf{a}_2), \quad a_f = a_f(m_1, m_2, \mathbf{a}_1, \mathbf{a}_2), \quad (4)$$

which allow us to obtain the posterior probability distribution $P(M_f, a_f | d)$ of the final mass and spin.

First, we estimate the posterior $P_{\text{IMR}}(M_f, a_f | d)$ using the full observed signal. We choose to demarcate the inspiral and post-inspiral parts of the signal using the $m = 2$ mode GW frequency f_{ISCO} of the innermost stable circular orbit (ISCO) of the remnant Kerr black hole [31], with mass and spin given by the median values of $P_{\text{IMR}}(M_f, a_f | d)$. We estimate the same parameters from the data containing only the inspiral (low-frequency) part of the observed signal; that is, by setting $f_{\text{cut}} = f_{\text{ISCO}}$ in equations (2) and (3). This allows us to compute the posterior distribution $P_{\text{I}}(M_f, a_f | d)$ of the mass and spin of the remnant purely from the inspiral part of the signal. Similarly, using only the merger–ringdown (high-frequency) part of the observed signal [by setting $f_{\text{low}} = f_{\text{ISCO}}$ in equations (2) and (3)], we can get yet another estimate of the posterior $P_{\text{MR}}(M_f, a_f | d)$ of the mass and spin of the remnant. If the observed signal is well described by GR, the two independent estimates $P_{\text{I}}(M_f, a_f | d)$ and $P_{\text{MR}}(M_f, a_f | d)$ have to be consistent with each other as well as with the estimate $P_{\text{IMR}}(M_f, a_f | d)$ using the full data (see, e.g., the left plot in figure 1).

To quantify the consistency of the observed signal with a binary black hole system predicted by GR, we define two parameters that describe the fractional difference between the two estimates of the remnant’s mass and spin

$$\epsilon := \frac{\Delta M_f}{\bar{M}_f}, \quad \sigma := \frac{\Delta a_f}{\bar{a}_f}, \quad (5)$$

where

$$\Delta M_f := M_f^{\text{I}} - M_f^{\text{MR}}, \quad \Delta a_f := a_f^{\text{I}} - a_f^{\text{MR}}, \quad (6)$$

and

$$\bar{M}_f := \frac{M_f^{\text{I}} + M_f^{\text{MR}}}{2}, \quad \bar{a}_f := \frac{a_f^{\text{I}} + a_f^{\text{MR}}}{2}. \quad (7)$$

In the absence of departures from GR, we expect the posterior $P(\epsilon, \sigma | d)$ of $\{\epsilon, \sigma\}$ to be consistent with zero (within the expected statistical fluctuations due to noise). \ddagger Appendix A describes the calculation of $P(\epsilon, \sigma | d)$ from the posteriors $P_{\text{I}}(M_f, a_f | d)$ and $P_{\text{MR}}(M_f, a_f | d)$. An example of the posterior distribution $P(\epsilon, \sigma | d)$ from a simulated GR signal is shown in figure 1 (right plot).

\dagger Here onwards, we drop \mathcal{H}, I from the posteriors to simplifying the notation. Also, unless otherwise noted, the masses we consider are the ones in the detector frame, including the cosmological redshift [30].

\ddagger We used the full IMR posteriors for M_f and a_f to normalize the fractional parameters in Paper I. We have switched to using the average of the inspiral and merger–ringdown posteriors here so that we only are using posteriors from the two independent measurements, and not mixing in ones from the case where the parameters describing the inspiral and merger–ringdown parts of the waveform are not allowed to vary independently.

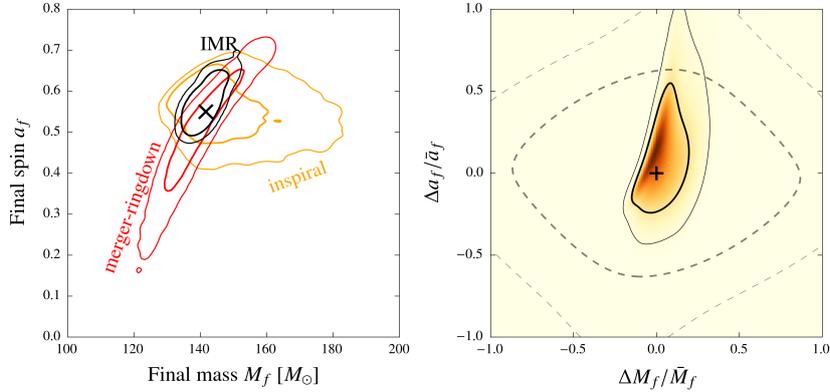


Figure 1. The left panel shows the 68% and 95% credible regions of the posterior distributions $P_I(M_f, a_f)$ and $P_{MR}(M_f, a_f)$ of the (redshifted) mass and spin of the final black hole estimated from the inspiral and merger–ringdown parts of a simulated GR signal, respectively. Also shown is the posterior $P_{IMR}(M_f, a_f)$ estimated from the full IMR signal. The simulated GR signal (using SEOBNRv2_ROM_DOUBLESPIN [32, 33]) is from a binary black hole with $(m_1, m_2) = (60.7, 86.7)M_\odot$, and aligned spins $(a_1, a_2) = (0.65, -0.77)$ at a distance of 860 Mpc, observed by the 3-detector Advanced LIGO–Virgo network. The corresponding value of the final mass and spin is indicated by a black cross. The right panel shows the posterior $P(\epsilon, \sigma)$ on the parameters $\epsilon := \Delta M_f / \bar{M}_f$ and $\sigma := \Delta a_f / \bar{a}_f$ that describe the deviation from GR for the same case as the left panel. Dashed lines show the 68% and 95% isoprobability contours of the prior used to compute the posterior. The posterior is consistent with the GR value, which is marked by a ‘+’ sign.

Finally, the likelihoods from multiple observations of binary black holes can be combined (multiplicatively) to construct a combined posterior on $\{\epsilon, \sigma\}$ that can better constrain deviations from GR. Here, we use the posterior from previous observations as the prior for the current observation. In practice, the posteriors $P_j(\epsilon, \sigma | d_j)$ from N such observations can be combined to produce an overall posterior in the following way:

$$P(\epsilon, \sigma | \{d_j\}) = P(\epsilon, \sigma) \prod_{j=1}^N \frac{P_j(\epsilon, \sigma | d_j)}{P_j(\epsilon, \sigma)}, \quad (8)$$

where $P(\epsilon, \sigma)$ is the overall prior distribution on ϵ and σ and $P_j(\epsilon, \sigma)$ is the prior distribution used to compute the posterior $P_j(\epsilon, \sigma | d_j)$.

3. Simulations of GR signals in Gaussian noise

First, we estimate the expected constraints that we can place on parameters describing deviations from GR, when the actual signals are well described by GR. We inject simulated GW signals modeling inspiral, merger and ringdown of binary black holes (based on GR, modeled by the SEOBNRv2_ROM_DOUBLESPIN approximant [32, 33]) into colored Gaussian noise with the design power spectrum of the Advanced LIGO detectors in the high-power, zero-detuning configuration [34]. Although we perform the analysis assuming the 3-detector Advanced LIGO–Virgo network, for simplicity, we assume that Advanced Virgo has the same noise power spectrum as the Advanced LIGO detectors. All detectors are assumed to have a low frequency cutoff $f_{\text{low}} = 10$ Hz. We simulate a population of binary black holes that is uniformly distributed in comoving four-volume in the redshift interval $z = [0, 2]$, with isotropic orientations (see Section II C of [35] for more details). Binaries have (source frame) component masses uniformly distributed in the range $m_{1,2}^s = [10, 80]M_\odot$ and component spin magnitudes uniformly distributed in the range $[0, 0.98]$. The GW signals are redshifted so that the effective masses describing the signal in

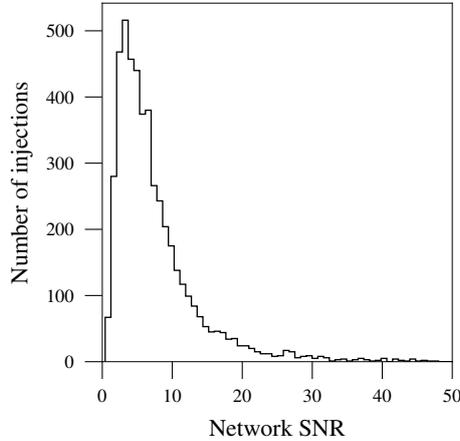


Figure 2. Distribution of SNR of Gaussian noise simulations of GW signals from binary black hole coalescences uniformly distributed in a comoving four-volume between redshifts $[0, 2]$, initial source frame component masses and component spin magnitudes uniformly distributed between $[10, 80]M_{\odot}$, and $[0, 0.98]$.

the detector frame become $m_{1,2} = m_{1,2}^s(1+z)$ [30]. Black hole spins are always assumed to be aligned/antialigned with the orbital angular momentum, as is required by the waveform model used: This is not an overly restrictive assumption, because the spin components along the orbital angular momentum have the dominant effect on the binary’s dynamics and GW emission [36]. Figure 2 shows the distribution of the network SNR of the injected population of binaries. For our analysis, we consider signals with network SNR greater than 8 for both the inspiral and merger–ringdown parts, using the Kerr ISCO frequency to split the signal †. In addition, we only consider binaries with total redshifted mass less than $150M_{\odot}$, so that the observed signals contain imprints of the inspiral, merger and ringdown.

On the injections that pass the SNR and mass thresholds, we perform the analysis described in section 2, using SEOBNRv2_ROM_DOUBLESPIN as our template for parameter estimation and the final mass and spin fits from Healy *et al.* [37]. Parameter estimation using LALINFERENCE_NEST is performed using priors that are uniform in component masses in the interval $m_{1,2} \in [1, 300]M_{\odot}$ and component (aligned/antialigned) spins in the interval $[0, 0.98]$. This corresponds to a non-uniform prior in the deviation parameters $\{\epsilon, \sigma\}$ that is centered around the GR value of $\{0, 0\}$. See figure 1 for an example result. The left panel shows the estimated posterior distributions of the final mass and spin from the inspiral and post-inspiral parts of the data, along with that estimated from the full data. The right panel shows the posterior distribution on the parameters describing deviations from GR, along with the prior used for computing this posterior.

Constraints on the deviations expected from single observations of binary black holes by second-generation ground-based detectors are expected to be modest, due to the large statistical errors from the relatively small SNRs expected for most systems (see, e.g., figure 2). However, the constraints can be improved by combining the likelihoods of $\{\epsilon, \sigma\}$ from multiple events. Figure 3 shows the combined posteriors $P(\epsilon, \sigma)$ as a function of the number of simulated events. The constraints on the deviation parameters $\{\epsilon, \sigma\}$ become narrower when multiple events are combined. These constraints are within the reach of Advanced LIGO–Virgo observations in the next few years [18, 5].

† Here, for simplicity, we use the Kerr ISCO frequency computed from the injected values of masses and spins to split the signal. In an actual observation, we use the median estimate of the Kerr ISCO frequency estimated from the observed IMR signal.

From the posterior distribution $P(\epsilon, \sigma)$, one can obtain the credible level corresponding to the GR value of the deviation parameters, i.e., $\{\epsilon, \sigma\} = \{0, 0\}$. The credible level for a certain set of parameters $\{\epsilon_0, \sigma_0\}$ is the total probability enclosed within an isoprobability contour passing through $\{\epsilon_0, \sigma_0\}$:

$$\text{credible level} = \int_{\{\epsilon, \sigma\} \text{ where } P(\epsilon, \sigma) > P(\epsilon_0, \sigma_0)} P(\epsilon, \sigma) d\epsilon d\sigma. \quad (9)$$

See, e.g., equation (42) in [24] for a general definition. The area enclosed within the isoprobability contour is the credible area (or credible interval in 1 dimension). A lower credible level for the GR parameters indicates a greater agreement with GR. However, this GR credible level is not expected to be zero even when the data is correctly described by GR, because the peak of posterior on the deviation parameters $\{\epsilon, \sigma\}$ is randomly shifted away from $\{0, 0\}$ in the presence of noise.

If the parameter values $\{\epsilon, \sigma\}$ were sampled from the prior distribution used to compute their posterior, we would expect the p credible interval of the posterior to include the true value approximately p of the time, for the case of an appropriately chosen prior [38, 39]. However, for the GR injections that we performed, we do not expect the true value of $\{\epsilon = 0, \sigma = 0\}$ to be found at credible level p for a fraction p of events. This is because of our prior distribution for $\{\epsilon, \sigma\}$ does not match the intrinsic distribution (which would be a delta function): we are allowing for the possibility of mismatches between inspiral and merger–ringdown even though this is not the case for the injected GR signals. To demonstrate this for our set of GR injections, in figure 4 we show the fraction of events for which the (true) GR point is found within the p credible region. From figure 4, we see the true GR value is often near the peak of the distribution, meaning that it is found within the p credible region for more than a fraction p of events.† In order to make a frequentist statement about the significance of the measurement of non-zero ϵ or σ (how common such a deviation is if GR were correct), a similar study would need to be done for the region of mass and spin space of interest in order to calibrate expectations.

4. Simulations of modified GR signals in Gaussian noise

4.1. Generation of modified GR waveforms

We also test the sensitivity of our pipeline towards certain kinds of deviations from GR, using signals that differ slightly from the predictions of GR. Specifically, we generate kludge modified GR waveforms that are qualitatively similar to binary black hole waveforms in GR, but differ in their energy and angular momentum loss as a function of frequency. These waveforms are similar enough to GR binary black hole waveforms that they would likely be detected with a standard detection pipeline, as we discuss below. However, we demonstrate that the consistency test described in this paper is able to identify such deviations from GR by combining the posteriors from multiple observations.

We start from the nonspinning effective one-body (EOB) waveform model given in [40], which is available as Matlab code at IHES [41], and modify the GW flux to yield our kludge modified GR waveform. Specifically, we multiply all the modes of the waveform that first enter the flux at second PN order [i.e. $O(v^4/c^4)$, where v is the binary’s orbital velocity], viz the $(\ell, m) = (3, \pm 2)$, $(4, \pm 4)$, and $(4, \pm 2)$ modes, by a constant factor $\sqrt{\alpha_2}$, so that their contributions to the flux are multiplied by α_2 . This also modifies the flux at 3PN and all higher PN orders. We start with the modes that first enter at 2PN rather than the ones that first enter at 1PN since the latter modes vanish for equal-mass nonspinning binaries, and we still want a nonvanishing modification in this limit.

We retain the termination condition for the inspiral–plunge phase of the waveform used the original code. This uses the maximum of the orbital frequency (calculated from the EOB

† For the marginalised ϵ distribution, we see that the true GR value is approximately found within the p credible interval a fraction p of the time. In this case, the scatter due to noise is sufficient to balance the mismatch in the prior used.

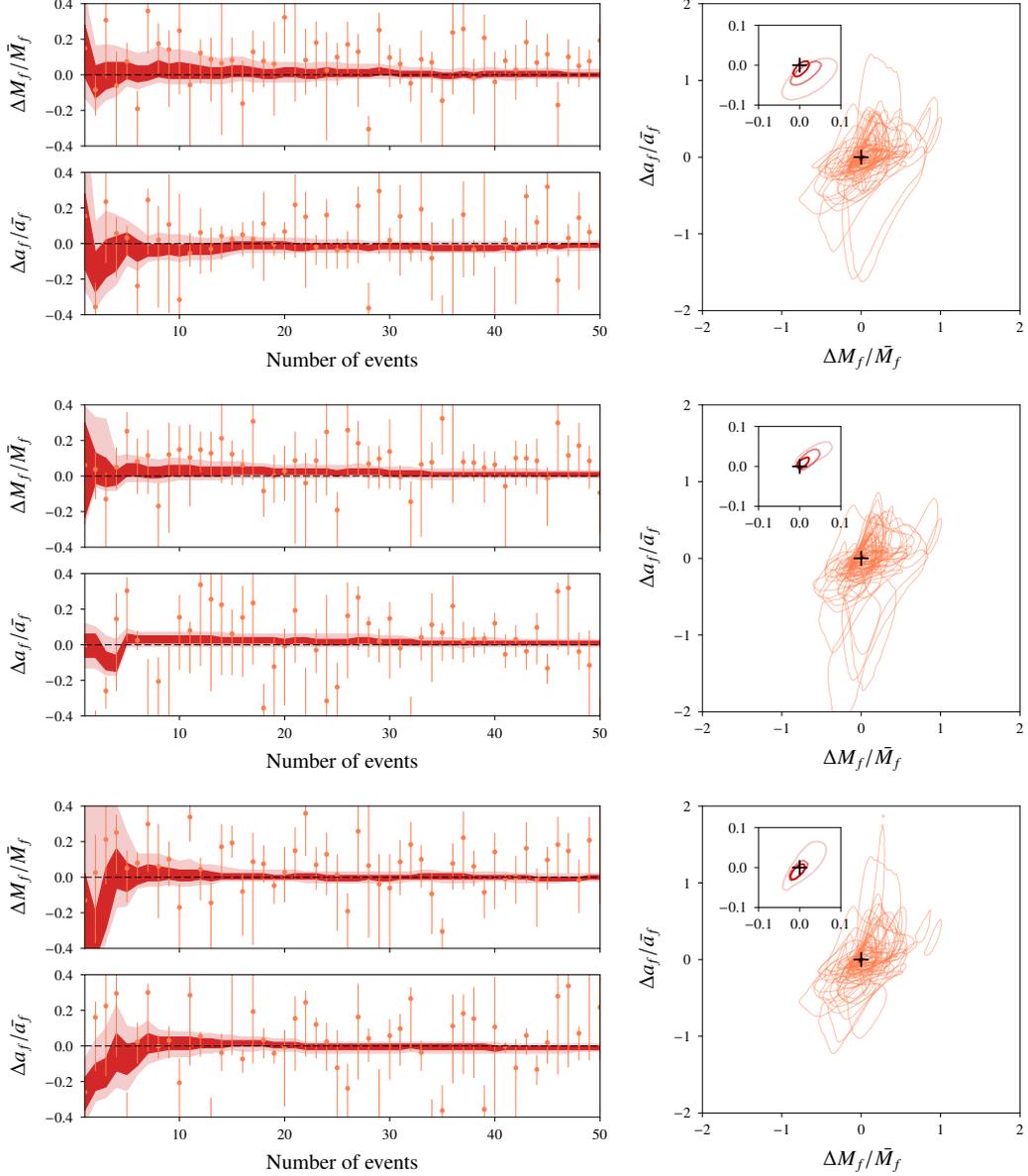


Figure 3. *Left:* In each plot shaded regions show the 68% and 95% credible intervals on the combined posteriors on $\epsilon := \Delta M_f / \bar{M}_f$, $\sigma := \Delta a_f / \bar{a}_f$ from multiple observations of GR signals, plotted against the number of binary black hole observations. The GR value ($\epsilon = \sigma = 0$) is indicated by horizontal dashed lines. The mean value of the posterior from each event is shown as an orange dot along with the corresponding 68% credible interval as an orange vertical line. Posteriors on ϵ are marginalized over σ , and vice versa. *Right:* The thin orange contours show the 68% credible regions of the individual posteriors on the ϵ, σ computed from the same events. The GR value ($\epsilon = \sigma = 0$) is indicated by the black + sign. *Right inset:* The red contours show the 68% credible regions on the combined posterior from 5, 10 and 25 events (with increasing shades of darkness). The GR value ($\epsilon = \sigma = 0$) is indicated by the black + sign. Different rows correspond to different catalogs of 50 randomly chosen events from a total of ~ 100 simulated events.

Hamiltonian) to mark the termination of the calculation of the inspiral–plunge waveform (by evolving the EOB Hamilton equations) and the start of the matching to the QNMs used to model

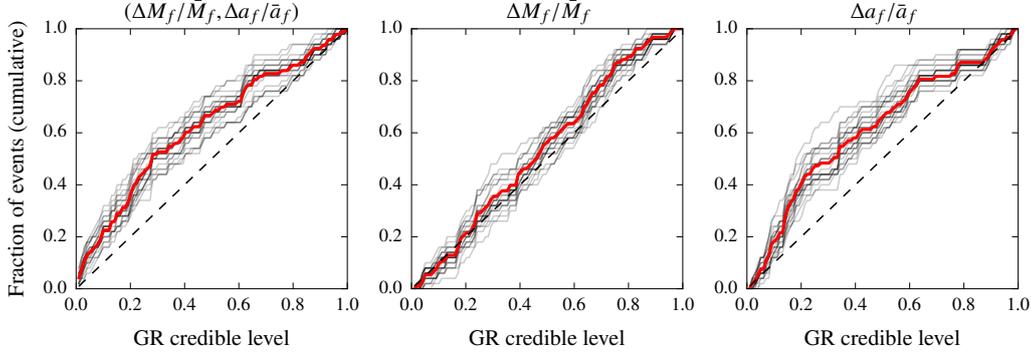


Figure 4. A p vs p plot for the deviation parameters ($\epsilon := \Delta M_f / \bar{M}_f$, $\sigma := \Delta a_f / \bar{a}_f$) computed from the set of GR injections described in section 3. In each plot, the horizontal axis indicates a credible level, and the vertical axis indicates the fraction of events with the deviation parameters below the given credible level. In each plot, the gray lines are computed for 50 different subsets of the full population of ~ 100 events, and the red line correspond to the same computed using all the ~ 100 events. The left plot corresponds to credible intervals computed from the 2-dimensional posteriors $P(\epsilon, \sigma | d)$, while the middle and right plots correspond to marginalized posteriors on ϵ and σ , respectively. The diagonal lines indicate the case where the Bayesian credible levels match the frequentist confidence levels.

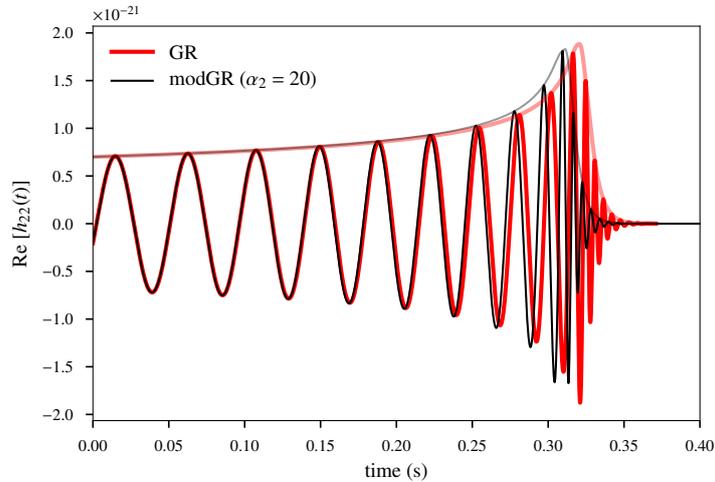


Figure 5. The real part (darker lines) and amplitude (lighter lines) of the $(2, 2)$ mode gravitational waveform from an equal-mass nonspinning binary black hole computed using the IHES EOB code with no modification (GR) and with our modification to the flux with $\alpha_2 = 20$ (modGR). For this illustration, we have taken the total mass to be $100M_\odot$, the distance to be 1 Gpc, and have aligned the waveforms at $t = 0$, which we have taken to correspond to a $(2, 2)$ mode frequency of 20 Hz.

the merger and ringdown. This termination condition coincides with crossing the light ring in the extreme mass-ratio limit. We also keep the next-to-quasicircular parameters set to the values determined from NR simulations that are already given in the code, for simplicity. Additionally, we keep all the other parameters set to their default values, except for a parameter that keeps one from probing unphysical regions of the EOB potential while the EOB equations are solved numerically. This was necessary to change in order to compute waveforms for higher mass ratios, even without other modifications to the code.

With these choices, we find that the waveforms look similar to a GR waveform up to the

values of $\alpha_2 = 20$ we consider here. This is illustrated in figure 5 for an equal-mass binary; the unequal-mass waveforms we consider in the next subsection also look similar to GR waveforms. Additionally, we have checked that this modification to the waveform does not induce any excess eccentricity: The eccentricity of the modified waveforms is small, and the same size as the eccentricities of the unmodified waveforms ($\lesssim 10^{-5}$ for the initial frequencies we are considering).[†]

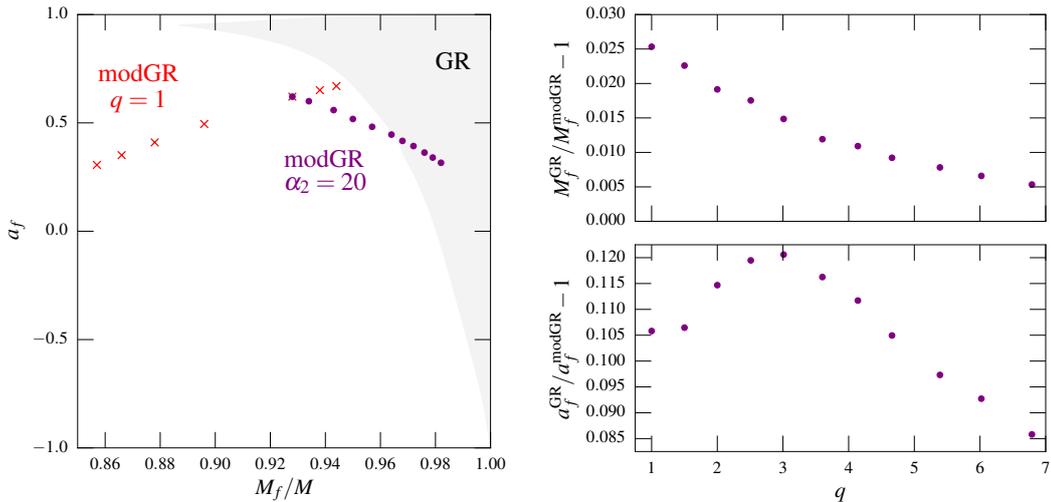


Figure 6. *Left:* The final mass and spin allowed by GR (shaded region), as calculated from the fits from [37], as well as the values for the kludge waveforms with a mass ratio of $q = 1$ and $\alpha_2 \in \{5, 10, 20, 100, 200, 300, 400\}$ (crosses, with α_2 increasing from right to left), as well as $\alpha_2 = 20$ and q from 1 to 6.8 in roughly steps of 0.5 (circles, with q increasing from left to right). The $\alpha_2 = 20$ points are taken from the injection set. *Right:* The fractional differences between the final mass and spin for the kludge waveforms with $\alpha_2 = 20$, as a function of the mass ratio, q .

Since the final mass and spin in the original EOB waveform are set by a fit to NR results, we replace this determination by demanding self-consistency of the radiated energy and angular momentum. That is, we choose the final mass and spin by minimizing the difference between the final mass and spin used to obtain the merger–ringdown part of the model and the final mass and spin obtained from energy and angular momentum balance using the initial energy and angular momentum and the radiated energy and angular momentum (calculated from the waveform). Here we are able to obtain fractional disagreements of less than $\sim 10^{-4}$, which is considerably better than the agreement of the original IHES EOB waveform for some mass ratios, and much smaller than the differences in radiated energy and angular momentum between the different waveforms we consider. We compute the energy and angular momentum loss using multipoles up to $\ell = 7$, since we are only able to fit for the ringdown up to $\ell = 7$, the highest ℓ for which the tabulated QNM results used in the code are available [13, 43]. We take the standard GR expressions for the radiated energy and angular momentum to still be valid for this modified gravity waveform; these expressions indeed remain valid for a significant number of modified theories [44]. We find that the final mass and spin falls outside the region allowed in GR in many cases, though there is a region of overlap for $\alpha_2 = 20$ and mass ratio $q := m_1/m_2 \gtrsim 4$; see figure 6. We also plot the fractional difference in the final mass and spin between GR and the kludge waveforms with $\alpha_2 = 20$ in this figure, and see that the

[†] We estimate the eccentricity of the waveforms as in section IV C of [42] [Eq. (4.13) and following], though with two additional higher-order PN terms added to the fitting function, since the eccentricity is small. The results did not change upon adding a third additional PN term. Since this estimate only uses the leading-order GW effects, it does not need to be modified to account for the modification to the flux, which only starts at 2PN.

fractional difference in the final spin is significantly larger than the fractional difference in the final mass.

In addition to enhanced energy and angular momentum fluxes, these kludge modified GR waveforms have an enhanced linear momentum flux. This leads to larger GW recoils than one obtains for nonspinning systems in GR, where the maximum is only $\sim 180 \text{ km s}^{-1}$ (see, e.g., [45]). For the $\alpha_2 = 20$ deviation we consider here, the maximum recoil is $\sim 440 \text{ km s}^{-1}$.[†] We do not include any potential effects (e.g., the Doppler shifts discussed in [46]) from these enhanced recoils, which are still small (at most $\sim 10^{-3}$ times the speed of light).[‡]

We have not changed the QNM spectrum of the final black hole in these kludge modified gravity waveforms for simplicity. However, one expects the QNM spectrum to be modified in any alternative theory that predicts a different GW energy loss for a binary black hole, even those for which the unperturbed Kerr metric is a solution of the field equations, since the behavior of perturbations of Kerr will generically be different in an alternative theory (see, e.g., [47]). Extending this analysis to the case where one modifies the QNM spectrum would be a useful additional check of the consistency test presented here, though the requisite catalogs of QNMs are not yet available, even for the case of a Kerr–Newman black hole, which could be used as a stand-in for a modified gravity black hole: See, however, Dias *et al.* [48] for some recent results that could be used to obtain such a QNM catalogue.

4.2. Results from the simulations of modified GR signals

Here we demonstrate that the IMR consistency test is able to identify (at least certain types of) deviations from GR by performing our analysis on a population of simulated signals using the modified GR waveforms described above. In Paper I, we have demonstrated the ability of this test to identify a relatively large modification to the binary’s energy flux ($\alpha_2 = 400$). Such a deviation from GR was easily detectable with high confidence from a single observation of moderate SNR. Here we consider the ability of the test to discern much smaller deviations from GR by combining results from multiple observations. Specifically, we consider the same population of binary parameters considered to simulate the GR case in section 3 except that the waveforms are generated with a modification of the GR energy flux as described in section 4.1. Also, for simplicity, we consider binary black holes with zero spins, since the EOB waveform family that we employ to produce modified GR waveforms is a nonspinning model [40, 41].

Modified GR waveforms for binaries with different mass ratios (as well as distances, sky locations, and other extrinsic parameters) were constructed using the prescription presented in section 4.1, with deviation parameter $\alpha_2 = 20$. An example of a modified GR waveform (along with the corresponding GR waveform from the same binary) is shown in figure 5. Although such a modified GR waveforms appears similar to a GR waveform, we show below that the IMR consistency test is able to identify deviations from GR by combining multiple events. Figure 7 shows the joint posteriors on the parameters that describe deviations from GR. It can be seen that, although individual posteriors are unable to identify deviations from GR, the combined posteriors show a clear departure from the GR predictions, even for the relatively small deviation that we consider.[†] We also see that the combined posterior on $\sigma := \Delta a_f / \bar{a}_f$ differs significantly more from 0 than does the combined posterior on $\epsilon := \Delta M_f / \bar{M}_f$. This might be expected, due to the larger fractional differences from GR in the final spin seen in figure 6.[‡]

[†] The original IHES EOB waveform model (without deviations) already has a maximum recoil of $\sim 350 \text{ km s}^{-1}$.

[‡] Enhanced recoil were not an issue for the larger deviation considered in Paper I because the recoil vanishes, by symmetry, for the equal-mass nonspinning system considered there.

[†] Taking $\alpha_2 = 20$ corresponds to multiplying the 2PN term in the frequency domain phase expression by a factor of ~ 0.4 for equal masses, with the factor decreasing monotonically to ~ -4.5 for a mass ratio of 7 (the largest mass ratio in the detected population).

[‡] This fractional difference in the final spin has the opposite sign to what might naïvely be expected, given that the modGR waveforms correspond to smaller final spins than their GR counterparts. Since most of the angular momentum loss comes

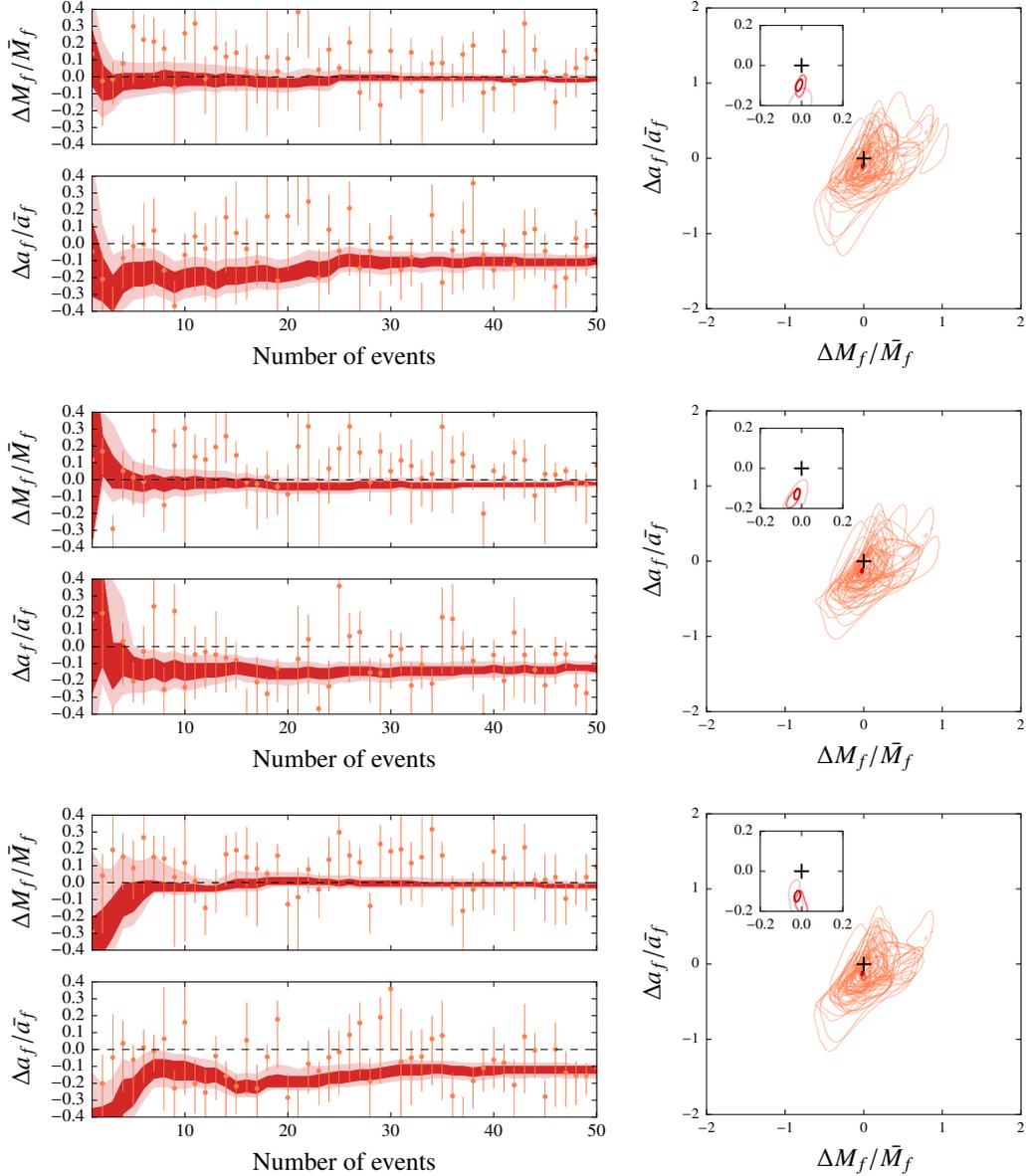


Figure 7. Same as figure 3 except that the test is performed on simulated signals containing a modification from GR described in section 4.2. The combined posteriors from multiple observations show a clear departure from the GR predictions (horizontal dashed lines on the left plots and the plus sign on the right plots).

during the merger and ringdown, it seems reasonable that the inspiral values for the final mass and spin would be reasonably close to their GR values for the system's mass ratio, while those inferred from the ringdown would correspond to the true values used to construct the waveform. However, since the analysis uses the merger–ringdown (and not just the ringdown) in order to have sufficient SNR, we do not expect such a direct correspondence to the values used to construct the waveform. For high-SNR events, the final mass and spin can be inferred directly using just the ringdown, where the only complication is the Doppler shift from any recoil. In this case, the initial masses and spins could also be inferred using only frequencies well below the merger. This would then allow the test to be interpreted as a missing energy experiment, where it could be concluded that additional energy (than predicted by GR) had been lost if the final black hole had a smaller mass than was

$m_1 (M_\odot)$	$m_2 (M_\odot)$	D_L (Mpc)	Optimal SNR	PyCBC SNR	Reduced χ^2	Re-weighted SNR
139.1	69.5	5625.8	8.1	8.6	0.76	8.6
88.1	14.7	803.5	14.3	14.0	0.78	14.0
40.8	46.8	787.0	21.6	21.7	0.81	21.7

Table 1. The table shows the parameters of some of the injected modified GR signals (D_L is the luminosity distance) and how well the PyCBC pipeline is able to detect these injected signals in a single detector search (LIGO Livingston). Optimal SNR is the optimal SNR available in the injection, PyCBC SNR is the matched filter SNR extracted by the PyCBC pipeline, and re-weighted SNR is the chi-squared weighted SNR defined in equation (6) of [50].

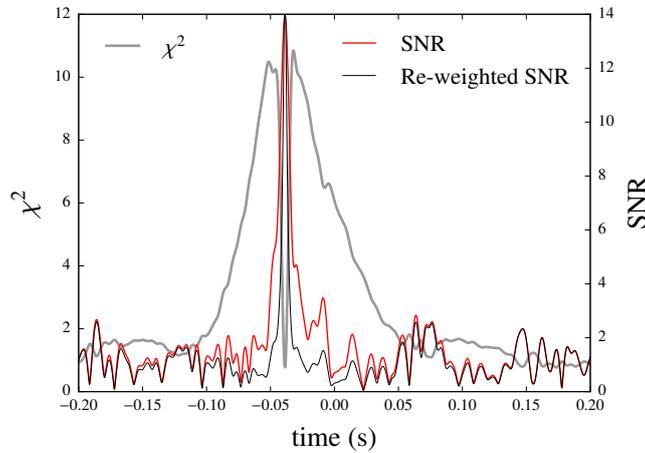


Figure 8. Time series of the matched filter SNR, the reduced χ^2 and re-weighted SNR from an injected modified GR waveform. The re-weighted SNR is the PyCBC detection statistic and is defined in equation (3) of [50]. The injected signal is located in the data at around time -0.05 s in the plot.

We only inject the $(2, \pm 2)$ modes of the waveform here, since these are the only modes included in the waveform model we use for the test, as well as for the injections in section 3. The modifications to the higher modes appear in the $(2, \pm 2)$ modes by their effect on the waveform’s phasing, since the higher modes are used to compute the modified energy flux. We have run the test on two GR injections with higher modes in section 5.4, and found that the inclusion of higher modes does not bias the test in those cases.

Finally, almost all of our modified GR signals will be confidently detected by the standard matched-filter based searches for binary black holes, which include the chi-squared discriminatory test [49]. To demonstrate this, we have run the PyCBC [50] matched filter based detection pipeline on these injections (using a single GR template from the SEOBNRv2_ROM_DOUBLESPIN family, corresponding to the parameters of the simulated modGR signal). We see that the chi-squared weighted SNR recovered by the detection pipeline is close to the optimal SNR of the signal (see table 1 and figure 8 for some examples). The SNR maximized over the full template bank is expected to be even larger than this, practically ensuring confident detections.

5. Robustness of the consistency test

We also test the robustness of the IMR consistency test to the choice of various parameters used in the analysis, for example, the cutoff frequency used to demarcate the division between expected from the inspiral.

the low-frequency (inspiral) and high-frequency (merger–ringdown) parts of the waveform, the choice of the particular approximant waveform, and the fitting formula for the mass and spin of the final black hole, etc. For this we use a simulated GR signal with waveform approximant SEOBNRv2_ROM_DOUBLESPIN (except in section 5.3 where we use approximant IMRPHENOMPv2 [51, 52, 53] and in section 5.4 where we use NR waveforms). We also study the effect of neglecting higher modes and spin precession on the IMR consistency test. These are performed by injecting NR waveforms from the publicly available Simulating eXtreme Spacetimes (SXS) catalog [54]. All these studies are performed assuming a binary black hole signal with parameters close to that of the first LIGO event GW150914 — masses $m_1 = 36M_\odot$, $m_2 = 29M_\odot$ and aligned spins, $a_{1,z} = -0.32$, $a_{2,z} = 0.58$, with optimum sky position and orientation, producing an optimal SNR of 25 in the Advanced LIGO Hanford–Livingston and Advanced Virgo network. These studies demonstrate the robustness of the IMR consistency test only for the case of signals with parameters similar to that of GW150914, that is, with moderate mass ratios. If we observe signals with large mass ratios and large, misaligned spins in the future, the robustness of the IMR consistency test will need to be reinvestigated. Also, these robustness studies have been restricted to the case of single events: when a large number of events are combined to produce precise constraints on the deviations from GR, we will have to worry about even small systematic errors affecting our analysis. Such studies have to be performed in the near future, in anticipation of the large number of binary black hole signals that Advanced LIGO is expected to observe.

5.1. Cutoff frequency between the inspiral and merger–ringdown

There is no well defined transition frequency (or time) between the inspiral and post-inspiral (merger–ringdown) parts of the waveform [55]. We have chosen the ISCO frequency of the final Kerr black hole with mass and spin inferred from the full IMR waveform at this cutoff frequency. Increasing this cutoff frequency will increase (decrease) the SNR of the inspiral (post-inspiral) part and hence will improve (worsen) the parameter estimation from the inspiral (post-inspiral) analysis, thus affecting our statistical errors. There thus is likely an optimal choice for the cutoff frequency as far as the test’s sensitivity to a given deviation from GR is concerned. However, if our analysis is free from major systematic errors, the two independent estimates always have to be consistent with each other when it is applied to a binary black hole coalescence in GR. Here we illustrate that our analysis produces consistent results for reasonable choices of the cutoff frequency.

The ISCO frequency of a Kerr black hole with mass $M_f = 62M_\odot$ and dimensionless spin $a_f = 0.68$ (remnant of the merger of two black holes with masses $m_1 = 36M_\odot$, $m_2 = 29M_\odot$ and aligned spins, $a_{1,z} = -0.32$, $a_{2,z} = 0.58$) is 146.4 Hz. We repeat the analysis with 4 different choices of the cutoff frequency in the interval 50–150 Hz. The results are shown in figure 9 (top panels). It can be seen that, while the spread of the posteriors (that is, the statistical errors) depends on the choice of the cutoff frequency, the inspiral and post-inspiral estimates are always consistent with each other. Also, the posteriors on the parameters describing deviations from GR are always consistent with zero, indicating the robustness of the test against the specific choice of the cutoff frequency.

5.2. Waveform approximant

The Bayesian inference described in section 2 is performed by employing semi-analytical gravitational waveforms (computed by a combination of analytical and numerical relativity) as the GR model. The inherent assumption is that these are faithful representations of the actual signals produced by nature. However, depending on the particular methods used to construct these waveforms, there can be minor differences between different gravitational waveform families (approximants). Here, we demonstrate the robustness of the IMR consistency test employing two IMR waveform families, namely, the non-precessing spin reduced-order EOB model

SEOBNRv2_ROM_DOUBLESPIN [32, 33] and the precessing single effective-spin phenomenological model IMRPHENOMPv2 [51, 52, 53]. The systematic errors of these models in the vicinity of GW150914 and their effects on parameter estimation were studied in [56].

Figure 9 (second row) shows the results from simulated GR signals from a binary black hole with parameters described in the previous section. Apart from performing the analysis where the same waveform family is employed in estimating the parameters from the inspiral and post-inspiral parts, the figure also demonstrates the robustness of the consistency test when we employ one waveform family for estimating the parameters from the inspiral part and another to estimate the same from the post-inspiral part.

5.3. Fit formulas for the mass and spin of the final black hole

The consistency test we have developed relies on NR fitting formulas to map the binary’s initial masses and spins into the mass and spin of the remnant black hole. There are a number of such fits available in the literature, of varying degrees of accuracy and generality. Here we consider a variety of recent fits for the final mass and spin.[†] We consider one set that is only applicable to nonspinning binaries (so we do not use the estimated spins in applying this) given in equations (29) of Pan *et al.* [58], and two sets applicable to binaries with aligned spins. The first set is from Healy, Lousto and Zlochower (HLZ) [37] and involves reasonably complicated implicit expressions given in Eq. (14) and Eq. (16) in that paper, with coefficients given in Table XI (we use the fourth-order fits). The second, simpler, though somewhat less accurate set is from Eq. (3.6) and Eq. (3.8) in Husa *et al.* [59] and only uses a single effective spin. The latter two fits were used in [3, 4] to infer the final mass and spin of GW150914 and the HLZ fits were used in the implementation of the current test in Paper I and [8]. For these fits, we use the projection of the spins along the orbital angular momentum when using a precessing waveform model, e.g., IMRPhenomPv2. The final mass and spin used to determine the ISCO frequency is not changed throughout the analysis, even when we are changing the fitting formulas. The calculation of the ISCO frequency uses the HLZ fits.

It is also possible to augment aligned-spin fits for the final spin to include the contribution from the in-plane spins, as was done for the HLZ fit in [60] and significantly increases its accuracy for precessing systems. As that reference shows for the HLZ fit, the basic aligned-spin final mass fit is already accurate for precessing systems when evaluated using the components of the spins along the orbital angular momentum. This augmentation was applied to the HLZ fit to infer the final spin of the binary black hole events from Advanced LIGO’s first observing run in [5, 6, 61]. We thus also consider this extension of the final spin expression for the HLZ and Husa *et al.* fits (HLZ aug and Husa *et al.* aug), as well as a new final spin fit for precessing systems from Hofmann, Barausse and Rezzolla (HBR) [62] (we use the one with the smallest number of coefficients, which the authors recommend), which we pair with the HLZ final mass fit. We also consider the augmented versions of both new aligned-spin fits that extend those in Husa *et al.* to include the spin difference (UIB aug) [57], as well as a small update to the HLZ fits from Healy and Lousto (HL aug) [45].

In figure 9 (third row), we show that for the simple test case we consider, the two aligned-spin fits give almost identical results, as do the extensions using the in-plane spins, while the nonspinning fits still give a posterior that is consistent with zero. Here we use the precessing waveform model IMRPhenomPv2 [51, 52, 53] for both the injection and recovery. We use it for the recovery in order to have nonzero in-plane spins, and use it (in its aligned-spin limit) for the injection so we are only testing different fits, not different waveform models here. Since we apply the same fitting formulas to infer the final mass and spin from both the inspiral and post-inspiral portions of the waveform, systematic errors in the fitting formulas do not translate into systematic errors in the test.

[†] While there are fits for the final mass and spin used in the two waveform models we employ, we do not just use those here, since they are not the most accurate and up-to-date ones: see Jiménez-Forteza *et al.* [57], which compares the final mass fit used in SEOBNRv2 and the final mass and spin fits used in IMRPhenomPv2 with more accurate fits.

Nevertheless, having more accurate fitting formulas might help the test distinguish modified GR signals from GR signals.

5.4. Effects of spin precession and higher modes

Although GW emission in the leading order is quadrupolar, higher order (nonquadrupole) modes can contain nonnegligible power if the system has large asymmetries (e.g., large mass ratios or large, misaligned spins). Due to the unavailability of waveform models that are accurate and computationally fast to generate, currently the parameter estimation is mostly performed making use of waveform templates that only model the quadrupole ($\ell = 2, m = \pm 2$) modes. This can potentially introduce systematic errors in the estimated parameters, which could be falsely taken for a deviation from GR. However, the systematic errors due to neglecting nonquadrupole modes are expected to be negligible for the case of binary black holes with moderate mass ratios and aligned/antialigned spins [56, 63]. In this paper’s population study we considered binaries with spins aligned/antialigned with the orbital angular momentum so that there is no precession. However, fast-to-evaluate waveform templates that take into account the dominant precession effects are available [51, 52, 53], which have been used for performing the IMR consistency test on the first LIGO event, GW150914 [8].

We investigate the robustness of the IMR consistency test against the effects of precession and higher modes, making use of injected NR waveforms from the SXS waveform catalog [64, 54]. We select two unequal mass ($m_1/m_2 = 1.228$) waveforms used in [56]: SXS:BBH:0307, which has aligned spins ($a_{1,z} = 0.32, a_{2,z} = -0.5798^\dagger$) binary black hole system, and SXS:BBH:0308, which is a precessing system with spin magnitudes $a_1 = 0.3406, a_2 = 0.6696$ and aligned components $a_{1,z} = 0.3224, a_{2,z} = -0.5761$. We inject both systems with total masses of $65M_\odot$ (so their individual masses are 36 and $29M_\odot$, the same as for the previous injections), and with inclination angles of 0. For parameter estimation, we use SEOBNRv2_ROM_DOUBLESPIN as our template. The results are summarized in figure 9 (bottom panel): the test is robust against the presence of higher modes and spin precession in the comparable mass regime.

5.5. Effect of splitting the signal in the frequency domain

In the IMR consistency test, we check the consistency of the posteriors of the final mass and spin estimated from the early part of the signal (inspiral) with the same estimated from the late part of the signal (merger–ringdown). The split between the inspiral and merger–ringdown parts is done in the Fourier domain. This choice is purely made for convenience, since, in the Fourier domain, this amounts to simply setting the lower and upper limits of the likelihood integral in equations (2) and (3).[‡] There is a possibility that a fraction of the power from the inspiral can get deposited at high Fourier frequencies, or the power from the merger–ringdown parts to get deposited at low Fourier frequencies. This effect will be particularly pronounced for the ringdown signal, whose power is spread over a large range of Fourier frequencies. One may worry that this spectral leakage would cause the posteriors derived from the two regions of Fourier frequency to be automatically consistent (although, we have already demonstrated that when certain kinds of deviations from GR are present in the signal, the IMR consistency test is able to identify them; see section 4.2).

Below we show that the effect of spectral leakage is small for the case of cutoff frequencies that we choose. In order to demonstrate this, we do the following: We compute the Fourier transform of the NR waveform SXS:BBH:0253 (an aligned-spin system with $m_1/m_2 = 2$ and $a_{1,z} = a_{2,z} = 0.5$) [64, 54] making use of the *stationary phase approximation* (SPA) [e.g., equation (3.7) of [65]]. The approximation here is that the power at a Fourier frequency f comes entirely from the time t_f (the saddle point) when the instantaneous frequency $F(t) := (d\varphi(t)/dt)/2\pi$

[†] The aligned spin components for the NR injections used in section 5.4 have signs opposite to the ones used in the rest of the section; this is not a typographical error.

[‡] This is different from splitting the signals in time domain and then taking the Fourier transform, which will introduce additional artifacts in the Fourier transform due to edge effects or windowing.

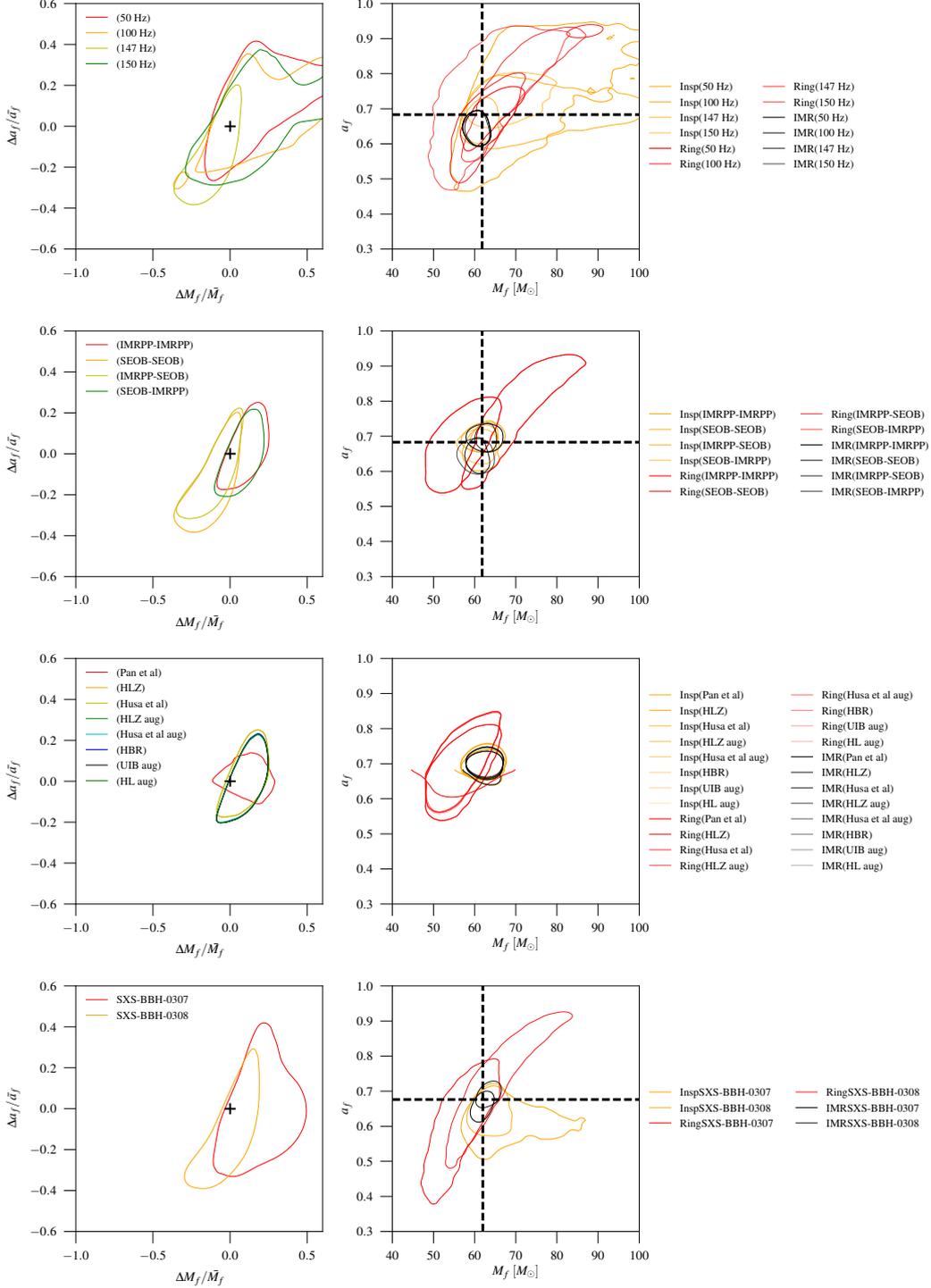


Figure 9. Robustness studies against the choice of cutoff frequency (top row), approximants (second row), fitting formulas (third row) and effects of higher modes and precession (bottom row). Left panels show the 68% credible region of the posterior distributions $P(\epsilon, \sigma)$. The GR value is marked by a + sign. Right panels show the 68% credible region of the posterior distributions $P_I(M_f, a_f)$, $P_{MR}(M_f, a_f)$ and $P_{IMR}(M_f, a_f)$ of the mass and spin of the final black hole estimated from the inspiral, merger–ringdown parts and the full IMR signal for each simulated case, respectively. In the third panel, the contours for the five fits that include the contributions from in-plane spins to the final spin (HLZ aug [37], Husa *et al.* aug [59], HBR [62], UIB aug [57] and HL aug [45]; “aug” denotes augmentation following [60]) lie almost on top of each other, thus making the individual contours indistinguishable. The same is true for the contours for the purely aligned-spin fits (HLZ [37] and Husa *et al.* [59]). The nonspinning Pan *et al.* [58] fit has a distinct contour.

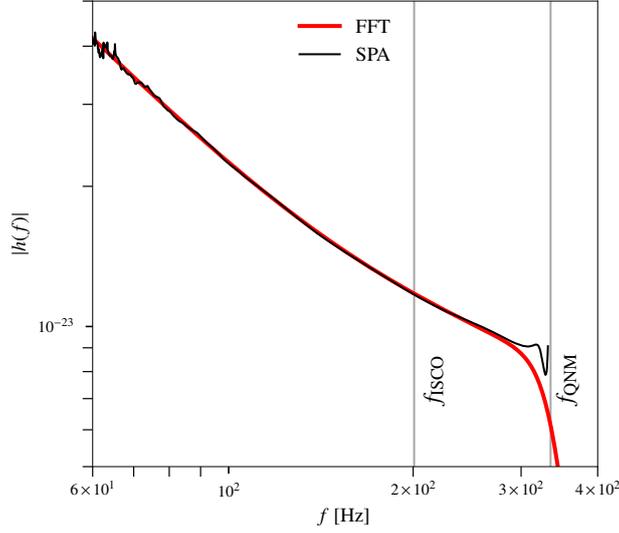


Figure 10. Comparison of the Fourier transform of the NR waveform SXS:BBH:0253 computed using the SPA with the exact Fourier transform computed using the FFT. The corresponding ISCO frequency f_{ISCO} is also shown, which is used to demarcate the inspiral and merger–ringdown parts. The plot suggests that the power at Fourier frequencies less than f_{ISCO} can be fully explained to be coming from the early times of the waveform. Also shown is the dominant QNM frequency, for reference. At high frequencies, the agreement between SPA and FFT becomes poor, indicating spectral leakage. The noise in the SPA at lower frequencies is due to the numerical noise in computing the frequency derivative.

is equal to f . Here $\varphi(t)$ is the phase of the waveform. In this approximation, the magnitude of the Fourier transform of the NR waveform can be computed as

$$|\tilde{h}(f)| = \frac{A(t_f)}{\sqrt{\dot{F}(t_f)}} \quad (10)$$

where $A(t)$ is the time-domain amplitude of the NR waveform, and a dot denotes a time derivative.

If the Fourier transform computed using the SPA agrees well with numerically computed Fast Fourier Transform (FFT), this gives a strong indication that the spectral leakage is negligible. Figure 10 compares the Fourier transform computed using the SPA with the FFT. The corresponding ISCO frequency f_{ISCO} is also shown, which is used to demarcate the inspiral and merger–ringdown parts. The plot suggests that the power at Fourier frequencies less than f_{ISCO} can be fully explained to be coming from the early times of the waveform (i.e., with instantaneous frequency $F(t) < f_{\text{ISCO}}$). The excellent agreement between the SPA and FFT at frequencies less than f_{ISCO} suggests that there is no appreciable spectral leakage between the two bands — almost all the power in the $f < f_{\text{ISCO}}$ ($f > f_{\text{ISCO}}$) band should come from the early (late) times. However, if we choose a transition frequency much higher than f_{ISCO} to demarcate the inspiral and merger–ringdown (say, the dominant QNM frequency, which is also shown in the figure) we expect significant spectral leakage between the two bands. Thus, if we need to perform this same analysis purely on the ringdown part (without including the merger), the analysis must be performed in the time domain. There is ongoing work in this direction.

6. Conclusions and future work

In this paper we have provided a detailed description of a test of GR based on the consistency between the inspiral, merger and ringdown parts of observed GW signals from coalescing binary black holes. First presented in [9], based on the idea proposed by Hughes and Menou [22], this was one of the handful of tests used to establish the consistency of the GW150914 event with a binary black hole system described by GR [8]. Here, we demonstrate how the likelihoods from multiple binary black hole events can be combined to produce tighter constraints on parameters describing a deviation from GR. In order to illustrate how this test might be able to detect certain types of deviations from GR, we generate internally consistent kludge waveforms which disagree with the GR prediction of the radiated energy and angular momentum, by increasing the flux radiated into GWs. We then simulate a population of binaries modeled by these modified GR waveforms. By combining the results from multiple events, we demonstrate how this test might be able to detect small deviations from GR. For the case of single events with parameters similar to LIGO's first binary black hole observation, we also demonstrate the robustness of the consistency test against specific choices of various analysis parameters, such as the choice of the transition frequency used to demarcate the inspiral and merger–ringdown parts, the specific waveform approximant, fitting formulas for the mass and spin of the remnant black hole, etc.

While we have investigated the reliability of our test to ensure its robustness, there remain further extensions of the test and these analyses to pursue. In particular, we have assumed that black holes in binaries have nonprecessing spins in the populations we simulate in this paper. We have also neglected the effect of nonquadrupole modes in the parameter estimation (although the fitting formulas for the remnant mass and spin include their contributions to the energy and angular momentum loss). While the effects of precession and nonquadrupole modes are expected to be negligible for the parameter estimation of comparable-mass binaries (see, e.g., [56, 63]) their effects might be appreciable for binaries with large mass ratios and high spins. Waveform templates describing the spin precession through an effective spin parameter, notably the IMRPHENOMPv2 template [51, 52, 53], are already available and have been used in parameter estimation. IMRPHENOMPv2 has been employed in the application of this test to GW150914 in [8] and in some of the robustness tests in section 5. More recently, waveform templates describing the full double-spin precession effects have also become available [66, 67]. These templates can be employed in parameter estimation when precession effects are expected to be apparent. Fitting formulas for the remnant mass and spin valid for the case of precessing initial spins are already available; some of them have been used in this paper itself. However, waveform templates for binaries of spinning black holes that include the effect of nonquadrupole modes are currently under development. We anticipate them to be available in the near future, and they may help to avoid possible systematic errors due to neglecting nonquadrupole modes.

The robustness studies performed in this paper are restricted to the case of single events with modest SNR. When a large number (~ 100) of events are combined to produce precise constraints on deviations from GR, even small systematic errors could become significant. Apart from the aspects discussed above, possible sources of errors include the finite accuracy of the GR waveform templates and the calibration of the GW detectors. We already have preliminary indications that combined posteriors from a large number of events could be dominated by such errors. Therefore, a careful characterization of various systematic errors is required before we accumulate a large number of detections to analyze. We leave this as future work.

Acknowledgments. We are grateful to M. Vallisneri for his insightful comments on the manuscript and many useful discussions. We thank J. Veitch, A. Nagar and P. Schmidt for assistance with the LALINFERENCE, IHES EOB and NR injection codes, respectively. We also thank the SXS Collaboration for creating a public archive of NR waveforms. We have benefited from useful discussions with K. G. Arun, A. Buonanno, N. Christensen, M. K. Haris, B. R. Iyer, S. Kumar,

A. K. Mehta, C. Messenger, A. Mukherjee, B. S. Sathyaprakash, C. Van Den Broeck, S. Vitale and several members of the LIGO Scientific and Virgo Collaboration’s working group for strong gravity. Ar. G., N. K. J.-M., and P. A. acknowledge support from the AIRBUS Group Corporate Foundation through a chair in “Mathematics of Complex Systems” at ICTS. P. A.’s research was, in addition, supported by a Ramanujan Fellowship from the Science and Engineering Research Board (SERB), India, the SERB FastTrack fellowship SR/FTP/PS-191/2012, and by the Max Planck Society and the Department of Science and Technology, India through a Max Planck Partner Group at ICTS. W. D. P. was partly supported by a Leverhulme Trust research project grant. C. P. L. B. was supported by the Science and Technology Facilities Council. Computations were performed at the ICTS clusters Mowgli, Dogmatix, and Alice. This is LIGO document LIGO-P1700006.

Appendix A. Calculation of the posterior of the parameters describing deviations from GR

Here we describe the calculation of the posterior $P(\epsilon, \sigma | d)$ of the parameters (ϵ, σ) describing the fractional difference between the two independent estimates of the mass and spin of the remnant. We first make a change of variables in the joint posterior $P_I(M_f^I, a_f^I | d)P_{MR}(M_f^{MR}, a_f^{MR} | d)$ from $\{M_f^I, a_f^I, M_f^{MR}, a_f^{MR}\}$ to $\{\epsilon, \sigma, \bar{M}_f, \bar{a}_f\}$ and then marginalize over \bar{M}_f and \bar{a}_f . Starting from the definitions given in equations (5) and (7), we find that

$$\begin{aligned} M_f^I &= \left(1 + \frac{\epsilon}{2}\right) \bar{M}_f, & a_f^I &= \left(1 + \frac{\sigma}{2}\right) \bar{a}_f, \\ M_f^{MR} &= \left(1 - \frac{\epsilon}{2}\right) \bar{M}_f, & a_f^{MR} &= \left(1 - \frac{\sigma}{2}\right) \bar{a}_f, \end{aligned} \quad (\text{A.1})$$

so that the Jacobian of the transformation from $\{\epsilon, \sigma, \bar{M}_f, \bar{a}_f\}$ to $\{M_f^I, a_f^I, M_f^{MR}, a_f^{MR}\}$ is $\bar{M}_f \bar{a}_f$. Thus, the final expression is

$$\begin{aligned} P(\epsilon, \sigma | d) &= \int_0^1 \int_0^\infty P_I\left(\left[1 + \frac{\epsilon}{2}\right] \bar{M}_f, \left[1 + \frac{\sigma}{2}\right] \bar{a}_f | d\right) \\ &\quad \times P_{MR}\left(\left[1 - \frac{\epsilon}{2}\right] \bar{M}_f, \left[1 - \frac{\sigma}{2}\right] \bar{a}_f | d\right) \bar{M}_f \bar{a}_f d\bar{M}_f d\bar{a}_f. \end{aligned} \quad (\text{A.2})$$

- [1] Aasi J *et al.* (LIGO Scientific Collaboration) 2015 *Class. Quantum Grav.* **32** 074001 (Preprint [1411.4547](#))
- [2] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Phys. Rev. Lett.* **116** 131103 (Preprint [1602.03838](#))
- [3] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Phys. Rev. Lett.* **116** 061102 (Preprint [1602.03837](#))
- [4] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Phys. Rev. Lett.* **116** 241102 (Preprint [1602.03840](#))
- [5] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Phys. Rev. X* **6** 041015 (Preprint [1606.04856](#))
- [6] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Phys. Rev. Lett.* **116** 241103 (Preprint [1606.04855](#))
- [7] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Astrophys. J. Lett.* **818** L22 (Preprint [1602.03846](#))
- [8] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Phys. Rev. Lett.* **116** 221101 (Preprint [1602.03841](#))
- [9] Ghosh A *et al.* 2016 *Phys. Rev. D* **94** 021101(R) (Preprint [1602.02453](#))
- [10] Vishveshwara C V 1970 *Nature (London)* **227** 936
- [11] Press W H 1971 *Astrophys. J.* **170** L105
- [12] Chandrasekhar S and Detweiler S 1975 *Proc. R. Soc. A* **344** 441
- [13] Berti E, Cardoso V and Starinets A O 2009 *Class. Quantum Grav.* **26** 163001 (Preprint [0905.2975](#))
- [14] Mishra C K, Arun K G, Iyer B R and Sathyaprakash B S 2010 *Phys. Rev. D* **82** 064010 (Preprint [1005.0304](#))
- [15] Agathos M, Del Pozzo W, Li T G F, Van Den Broeck C, Veitch J and Vitale S 2014 *Phys. Rev. D* **89** 082001 (Preprint [1311.0420](#))
- [16] Will C M 1998 *Phys. Rev. D* **57** 2061 (Preprint [gr-qc/9709011](#))

- [17] Abbott B P *et al.* 2016 *Living Rev. Relativity* **19** 1 (Preprint [1304.0670](#))
- [18] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Astrophys. J. Lett.* **833** L1 (Preprint [1602.03842](#))
- [19] Acernese F *et al.* (Virgo Collaboration) 2015 *Class. Quantum Grav.* **32** 024001 (Preprint [1408.3978](#))
- [20] Aso Y, Michimura Y, Somiya K, Ando M, Miyakawa O, Sekiguchi T, Tatsumi D and Yamamoto H (The KAGRA Collaboration) 2013 *Phys. Rev. D* **88** 043007 (Preprint [1306.6747](#))
- [21] Iyer B *et al.* 2011 LIGO-India, proposal of the consortium for Indian initiative in gravitational-wave observations (IndIGO) URL <https://dcc.ligo.org/LIGO-M1100296/public>
- [22] Hughes S A and Menou K 2005 *Astrophys. J.* **623** 689 (Preprint [astro-ph/0410148](#))
- [23] Nakano H, Tanaka T and Nakamura T 2015 *Phys. Rev. D* **92** 064003 (Preprint [1506.00560](#))
- [24] Veitch J *et al.* 2015 *Phys. Rev. D* **91** 042003 (Preprint [1409.7215](#))
- [25] Cutler C and Flanagan E E 1994 *Phys. Rev. D* **49** 2658–2697 (Preprint [gr-qc/9402014](#))
- [26] Gregory P C 2005 *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge: Cambridge University Press)
- [27] Skilling J 2006 *Bayesian Anal.* **1** 833
- [28] Veitch J and Vecchio A 2010 *Phys. Rev. D* **81** (Preprint [0911.3820](#))
- [29] LALSuite; <https://www.lsc-group.phys.uwm.edu/daswg/projects/lalsuite.html>
- [30] Krolak A and Schutz B F 1987 *Gen. Rel. Grav.* **19** 1163–1171
- [31] Bardeen J M, Press W H and Teukolsky S A 1972 *Astrophys. J.* **178** 347
- [32] Pürrer M 2016 *Phys. Rev. D* **93** 064041 (Preprint [1512.02248](#))
- [33] Taracchini A *et al.* 2014 *Phys. Rev. D* **89** 061502(R) (Preprint [1311.2544](#))
- [34] Shoemaker D H 2010 Advanced LIGO anticipated sensitivity curves Tech. Rep. LIGO-T0900288 <https://dcc.ligo.org/LIGO-T0900288/public>
- [35] Ghosh A, Del Pozzo W and Ajith P 2016 *Phys. Rev. D* **94** 104070 (Preprint [1505.05607](#))
- [36] Schmidt P, Hannam M and Husa S 2012 *Phys. Rev. D* **86** 104063 (Preprint [1207.3088](#))
- [37] Healy J, Lousto C O and Zlochower Y 2014 *Phys. Rev. D* **90** 104004 (Preprint [1406.7295](#))
- [38] Datta G S and Sweeting T J 2005 *Handbook of statistics* **25** 91–114 (Preprint <https://www.ucl.ac.uk/statistics/research/pdfs/rr252.pdf>)
- [39] Cook S R, Gelman A and Rubin D B 2006 *J. Comp. Graph. Stat.* **15** 675–692
- [40] Damour T, Nagar A and Bernuzzi S 2013 *Phys. Rev. D* **87** 084035 (Preprint [1212.4357](#))
- [41] <http://eob.ihes.fr>; we used the “1202” version of the code
- [42] Dietrich T, Moldenhauer N, Johnson-McDaniel N K, Bernuzzi S, Markakis C M, Brüggmann B and Tichy W 2015 *Phys. Rev. D* **92** 124007 (Preprint [1507.07100](#))
- [43] <http://www.phy.olemiss.edu/~berti/qnms.html>
- [44] Stein L C and Yunes N 2011 *Phys. Rev. D* **83** 064038 (Preprint [1012.3144](#))
- [45] Healy J and Lousto C O 2017 *Phys. Rev. D* **95** 024037 (Preprint [1610.09713](#))
- [46] Gerosa D and Moore C J 2016 *Phys. Rev. Lett.* **117** 011101 (Preprint [1606.04226](#))
- [47] Barausse E and Sotiriou T P 2008 *Phys. Rev. Lett.* **101** 099001 (Preprint [0803.3433](#))
- [48] Dias O J C, Godazgar M and Santos J E 2015 *Phys. Rev. Lett.* **114** 151101 (Preprint [1501.04625](#))
- [49] Allen B 2005 *Phys. Rev. D* **71** 062001 (Preprint [gr-qc/0405045](#))
- [50] Usman S A *et al.* 2016 *Class. Quantum Grav.* **33** 215004 (Preprint [1508.02357](#))
- [51] Hannam M, Schmidt P, Bohé A, Haegel L, Husa S, Ohme F, Pratten G and Pürrer M 2014 *Phys. Rev. Lett.* **113** 151101 (Preprint [1308.3271](#))
- [52] Khan S, Husa S, Hannam M, Ohme F, Pürrer M, Jiménez Forteza X and Bohé A 2016 *Phys. Rev. D* **93** 044007 (Preprint [1508.07253](#))
- [53] Bohé A, Pürrer M, Hannam M, Husa S, Ohme F, Schmidt P *et al.* In preparation
- [54] SXS Gravitational Waveform Database URL <http://www.black-holes.org/waveforms/>
- [55] Buonanno A, Cook G B and Pretorius F 2007 *Phys. Rev. D* **75** 124018 (Preprint [gr-qc/0610122](#))
- [56] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2017 *Class. Quant. Grav.* **34** 104002 (Preprint [1611.07531](#))
- [57] Jiménez-Forteza X, Keitel D, Husa S, Hannam M, Khan S and Pürrer M 2017 *Phys. Rev. D* **95** 064024 (Preprint [1611.00332](#))
- [58] Pan Y, Buonanno A, Boyle M, Buchman L T, Kidder L E, Pfeiffer H P and Scheel M A 2011 *Phys. Rev. D* **84** 124052 (Preprint [1106.1021](#))
- [59] Husa S, Khan S, Hannam M, Pürrer M, Ohme F, Jiménez Forteza X and Bohé A 2016 *Phys. Rev. D* **93** 044006 (Preprint [1508.07250](#))
- [60] Johnson-McDaniel N K, Gupta A, Ajith P, Keitel D, Birnholtz O, Ohme F and Husa S 2016 Determining the final spin of a binary black hole system including in-plane spins: Method and checks of accuracy Tech. Rep. LIGO-T1600168 <https://dcc.ligo.org/LIGO-T1600168/public>
- [61] Abbott B P *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) 2016 *Phys. Rev. X* **6** 041014 (Preprint [1606.01210](#))
- [62] Hofmann F, Barausse E and Rezzolla L 2016 *Astrophys. J. Lett.* **825** L19 (Preprint [1605.01938](#))
- [63] Varma V and Ajith P 2016 (Preprint [1612.05608](#))
- [64] Mroué A H *et al.* 2013 *Phys. Rev. Lett.* **111** 241104 (Preprint [1304.6077](#))
- [65] Damour T, Iyer B R and Sathyaprakash B S 2000 *Phys. Rev. D* **62** 084036 (Preprint [gr-qc/0001023](#))

- [66] Pan Y, Buonanno A, Taracchini A, Kidder L E, Mroué A H, Pfeiffer H P, Scheel M A and Szilágyi B 2014 *Phys. Rev. D* **89** 084006 (*Preprint* [1307.6232](#))
- [67] Babak S, Taracchini A and Buonanno A 2017 *Phys. Rev. D* **95** 024010 (*Preprint* [1607.05661](#))