Model systematics in time domain tests of binary black hole evolution

Shilpa Kastha [1,2] Collin D. Capano, [1,2,3] Julian Westerweck, [1,2] Miriam Cabero, [4] Badri Krishnan, [1,2,5] and Alex B. Nielsen [6]

[1]*Albert-Einstein-Institut, Max-Planck-Institut für Gravitationsphysik,
Callinstraße 38, 30167 Hannover, Germany*
[2]*Leibniz Universität Hannover, 30167 Hannover, Germany*
[3]*Department of Mathematics, University of Massachusetts, Dartmouth, Massachusetts 02747, USA*
[4]*Department of Physics and Astronomy, The University of British Columbia,
Vancouver, British Columbia V6T 1Z4, Canada*
[5]*Institute for Mathematics, Astrophysics and Particle Physics, Radboud University,
Heyendaalseweg 135, 6525 AJ Nijmegen, Netherlands*
[6]*Department of Mathematics and Physics, University of Stavanger, NO-4036 Stavanger, Norway*

We perform several consistency tests between different phases of binary black hole dynamics; the inspiral, the merger, and the ringdown on the gravitational wave events GW150914 and GW170814. To perform these tests we excise the data explicitly in the time domain to avoid any spectral leakage between the different phases using gating and inpainting method. We estimate the posterior distributions on the mass and spin of the initial black holes and the final black hole separately and independently from the different phases in frequency-domain. We also compute the initial areas of the two individual black holes and the final area from the parameters describing the remnant black hole. This facilitates a test of Hawking's black hole area theorem. We use different waveform models to quantify systematic waveform uncertainties for the area increase law with the two events. We find that these errors may lead to overstating the confidence with which the area theorem is confirmed. For example, we find $> 99\%$ agreement with the area theorem for GW150914 if a damped sinusoid consisting of a single mode is used at merger to estimate the final area. This is because this model overestimates the final mass. Including an overtone of the dominant mode decreases the confidence to $\sim 94\%$; using a full merger-ringdown model further decreases the confidence to ~ 85–90%. We find that comparing the measured change in area to the expected change in area yields a more robust test, as it also captures over estimates in the change of area. We find good agreement with GR when applying this test to GW150914 and GW170814.

I. INTRODUCTION

The first observation of gravitational waves (GWs) by the two detectors of the Laser Interferometer Gravitational-wave Observatory (LIGO), GW150914 [1], was inferred to be produced by two merging black holes (BHs) with masses $\sim 36~M_\odot$ and $\sim 29~M_\odot$ at a luminosity distance of ~ 400 Mpc. Following the first detection, several others [2–13] have also been detected during the three observing runs of the LIGO-Virgo detectors to date. During the second observing run, GW170814 was detected coherently by the Advanced Virgo detector along with the two Advanced LIGO detectors. GW170814 [6] was produced by two BHs with masses $30^{+5.7}_{-3.0}~M_\odot$ and $25.3^{+2.8}_{-4.2}~M_\odot$ at a luminosity distance of 540^{+130}_{-210} Mpc.

These observations have enabled tests of the predictions of Einstein's general relativity (GR) in a highly nonlinear and relativistic regime produced by compact binary mergers [14,15]. Several tests of GR have been developed for these recent GW detections, including parametrized tests of GR waveforms [16–25], parametrized tests of the multipolar structure of binaries [24,26], tests of the parametrized post-Einsteinian formalism [19,27], inspiral-merger-ringdown (IMR) consistency tests [28], and tests of black hole no-hair theorem [29–32].

The frequency-domain IMR consistency test is routinely performed on GW events detected by the LIGO and Virgo detectors [14]. This test checks the consistency of the low frequency part of the observed signal with the high-frequency part. Results are presented in terms of the differences of the final mass and final spin inferred from the two different parts. These values are computed using

extrapolations based on GR models parametrized in terms of the individual black hole masses and spins. Results to date are consistent with these differences being zero, implying that the GR based models are consistent with the data.

A related test uses measurements of the initial and final black holes' parameters as a check of the black hole area increase law. This law states that within classical general relativity, assuming the null energy condition and cosmic censorship, the area of a black hole horizon can never decrease. This was first shown by Hawking [33,34] and was later generalized to include non-differentiable event horizons and a cosmological constant [35]. Quasilocal versions of this law, which do not require cosmic censorship are also known [36–38]. Any stationary, astrophysical black hole is completely described by the Kerr metric in terms of its mass M and angular momentum J if the black hole no-hair theorem holds. The corresponding horizon area is $A = 8\pi M^2(1 + \sqrt{1 - \chi^2})$, where $\chi = J/M^2$ is the dimensionless spin of the black hole. For a binary black hole coalescence, the area theorem states that the final area A_f of the merged black hole will be larger than the combined area $(A_1 + A_2)$ of the two initial black holes,

$$A_1 + A_2 \equiv A_i < A_f. \tag{1}$$

It is additionally expected that sufficiently removed from the dynamical regime, i.e., at very early or late times, these black holes will be well described by the stationary Kerr metric.

Following [39], a concrete proposal for a test of the area increase law was proposed in [40]. The test consists of independent analyses of the early inspiral and the final ringdown stages, leading respectively to independent estimates of the initial and the final masses and spins of the binary. This test is carried out in the time domain, meaning that a portion of the strain data around the merger is excised in the time domain to separate the inspiral and ringdown phases. The data segment around the merger is excised in estimating the parameters, since a violation of the area theorem is perhaps most likely to occur near the merger of the two black holes where the spacetime is highly dynamical. These estimates are then used to obtain the areas of the individual BHs using the Kerr formula. To demonstrate the method, the authors used a simulated GW150914-like binary black hole signal and found that the area theorem could be confirmed at 75% probability.

Recently, Ref. [41] has explored the validity of the area theorem in the time domain and presented observational evidence that actual GW150914 data is consistent with the theorem with a probability of 97% when they do not excise the merger and 95% when they excise 3 ms of the merger. They also provide an estimate of the same by truncating the inspiral at different times before the peak amplitude and find

that the different measurements support the area theorem with probabilities within 88%–97%.

In this paper we apply the test proposed in [40] to two GW events, GW150914 and GW170814. To test the area theorem or the IMR consistency between different phases of the waveform, we need each of these different phases (inspiral/ringdown) to have reasonable signal-to-noise ratio (SNR) to perform the parameter estimations. We choose GW150914 as it has a high network-SNR of ∼24 [1,2], with a ringdown SNR ∼ 8.5 [42] 3 ms after the merger. The estimation of the ringdown SNR is based on the result in Ref. [42] (see Fig. 5 of Ref. [42]), where, among the several ringdown analyses of the GW150914, only those results were consistent with general relativity when the start time was chosen to be 3 ms (or later) after the merger. Furthermore, previously, in Ref. [43], with a simulated GW150914-like system, similar results were shown. The other event, GW170814, showed some support for a deviation from GR in initial analyses ($\Delta M_f/\bar{M}_f$ showed a second peak at higher values away from zero; see Fig. 2 of [14]). To explore the validity of the area theorem we choose GW170814 as our second candidate event.

We separately analyze the data segment before a truncation time, removing later data, to obtain the initial parameters. Similarly, the data before a truncation time are removed, and the remainder is analyzed to estimate the final parameters of each event. The truncation times may be chosen differently between the pre- and post-truncation analyses such that the merger phase is excluded. We employ various waveform models for estimating the parameters to show the effect of waveform systematics on the constraints of the area increase law. We show that ignoring higher overtones in the ringdown waveform model leads to overestimation of the final mass. This yields an inconsistency between the estimated parameters from the pre- and post-truncation analyses, but perversely results in a *better* agreement with the area theorem. A positive change in area, while obeying the area increase law, may still disagree with GR predictions if the area increase is too large. We use the ratio between the measured and the expected change of area, $\mathcal{R} = (A_f^{\text{measured}} - A_i^{\text{measured}})/(A_f^{\text{expected}} - A_i^{\text{measured}})$, as a measure here. For a perfect measurement, obeying GR predictions, $\mathcal{R} = 1$. However, in the presence of noise, we expect $\mathcal{R}$ to follow a Gaussian distribution with unit mean. In order to quantify the agreement of our results with GR, we also compute the mismatch, $C_\mathcal{R}$. It denotes the probability of R lying within the range symmetric about the median of the R-distribution, extending to $R = 1$. For GW150914, we find $C_\mathcal{R} = 6.5\%$ when we avoid 26 ms of data around the merger. Although, for GW170814, we obtain a better measurement yielding $C_\mathcal{R} = 2.7\%$ when ∼13 ms of data is avoided around the merger.

This paper is organized as follows. In Sec. II we explain the method adopted here to excise the data and perform parameter estimation. In Secs. III and IV we

develop the different pretruncation and post-truncation analyses. In Sec. V we discuss our results for the inspiral-merger-ringdown consistency test and in Sec. VI for the test of the area theorem. Our concluding remarks are presented in Sec. VII.

II. DATA PREPARATION AND PARAMETER ESTIMATION

The gravitational wave strain, $h(t)$, observed by a detector can be schematically written as

$$h(t - t_0) = F_+(\alpha, \delta, \psi) h_+(t - t_0, \phi_0) \qquad (2)$$

$$+ F_\times(\alpha, \delta, \psi) h_\times(t - t_0, \phi_0), \qquad (3)$$

where $F_{+,\times}$ are the antenna pattern functions of the detector. The right ascension, α, and declination, δ, define the sky-location of the source in a geocentric coordinate system and the polarization angle, ψ, defines the relative orientation of the wave frame with respect to the geo-centric coordinate system [44,45]. t_0 is the arrival time of the signal at the detector and ϕ_0 is the phase at t_0. $h_+(t)$ and $h_\times(t)$ are the two independent polarizations of the GW signal, given by

$$h_+(t) = A_+(t) \cos \Phi(t), \qquad (4)$$

$$h_\times(t) = A_\times(t) \sin \Phi(t), \qquad (5)$$

where $A_{+,\times}$ are slowly varying amplitudes and $\Phi(t)$ is a rapidly varying phase.

To estimate the source parameters of a GW signal present in a given data stream, $s(t)$, we use Bayesian inference. We first consider a model for the signal h within GR, para-metrized by the source properties such as masses, spins etc., $\{M_1, \chi_1, \ldots\} \equiv \vec{\vartheta}$. According to Bayes' theorem, the prob-ability distribution of the model parameters given the data $s(t)$, $p(\vec{\vartheta}|s, h)$ (known as the *posterior* distribution) is proportional to the likelihood, $\mathcal{L}(s|\vec{\vartheta}, h)$, of observing the data given $\vec{\vartheta}$ multiplied by a *prior* distribution, $p(\vec{\vartheta})$, on the parameters representing the allowed range and expected distribution on $\vec{\vartheta}$. For a network of N gravitational-wave detectors d, the likelihood function is given by [46,47]

$$\mathcal{L}(s|\vec{\vartheta}, h) \propto \exp\left[-\frac{1}{2} \sum_{d=1}^{N} \langle s_d - h_d(\vec{\vartheta}), s_d - h_d(\vec{\vartheta}) \rangle \right], \qquad (6)$$

where s_d is the data in the dth detector (assuming the noise to be uncorrelated, stationary and Gaussian) and h_d is the waveform model (or *template*). Here, the noise-weighted inner product is defined as [46,47]

$$\langle a|b \rangle = 2\mathrm{Re} \int_{f_{\mathrm{low}}}^{f_{\mathrm{high}}} \frac{\tilde{a}(f)\tilde{b}^*(f) + \tilde{a}^*(f)\tilde{b}(f)}{S_n^{(d)}(f)} df, \qquad (7)$$

where $S_n^{(d)}(f)$ is the noise power spectral density (PSD) of the dth detector and $\tilde{a}(f)$ and $\tilde{b}(f)$ are frequency domain representations of the time-domain data. We use PYCBC INFERENCE[48] to evaluate Eq. (6) over the large, multi-dimensional parameter space defining the waveform model. To sample the parameter space we use the dynesty [49] and parallel-tempered emcee [50,51] sto-chastic samplers. Marginalizing the resulting distribution yields measurements on individual parameters.

Our aim is to analyze separately the early and late parts of the signal to investigate the consistency between the different phases, and to explore the validity of the area theorem. In order to perform these analyses, we excise portions of the templates, keeping a desired time segment intact to perform the parameter estimation. As observed in Ref. [40], merely excising times from the template can lead to biases unless a corresponding part of the signal is also removed from the data. Hence, both the template and the signal need to be excised while performing parameter estimation. We accom-plish this by doing "gating and in-painting" [9,29]. In this method, we zero-out ("gate") the residual $s_d - h_d(\vec{\vartheta})$ over times to be excised, then add a component to the gated times such that the contribution of these times to the likelihood is zero ("in-painting"). This removes biases that arise due to the convolution of the inverse covariance of the detector noise with the residual. The end result is the pretruncation analysis is independent of the post-truncation analysis.

Since the sky location affects the arrival time of the signal in the detectors, the specific time for the gating changes with respect to different sky locations. If the sky location is varied during the analysis, the truncated template favors those sky locations that include more of the signal. This results in an estimation of the parameters shifted from their true values [40]. To avoid this, here we consider a fixed sky location. In order to fix the values for α and δ, we first perform the parameter estimation analysis on the full data stream, using the complete inspiral-merger-ringdown signal. We then use the corresponding maximum likelihood values for the sky location in the area-theorem analysis. Using slightly different values within the range of the posterior distributions on α and δ does not affect our final conclusions.

III. PRETRUNCATION ANALYSIS

In this section we focus on measuring the initial para-meters from the early part of the data. To model the gravitational-wave signal in the early to late inspiral regime, we use three waveform models: IMRPhenomPv2 [52,53], which models precessing binaries; IMRPhenomXPHM [54,55], which models precessing binaries with higher modes; and NRSur7dq4 [56], which also models precessing

binaries with higher modes, mass ratios $q \leq 6$, and spin magnitudes $\chi_1, \chi_2 \leq 0.8$.

For the pre-truncation analysis, we excise the residual after a desired time ("*gate-start-time*") and perform the sampling on the remaining data segment. While the truncation can be started at any time within the inspiral regime, earlier truncation times lead to posterior distributions yielding worse constraints on the estimated parameters due to decreasing SNR of the remaining signal. We choose two different *gate-start-times* in our analysis. In the first case, we use the merger time, t_m (obtained when the parameter estimation performed with the complete data set) as the truncation time. For GW150914, we base the merger time on the estimate in [15], while for GW170814 we use the results from [2], as explained in more detail in the respective sections.

In the second case we use the time corresponding to the hybrid minimum energy circular orbit (hybrid MECO), t_{hmeco} [57]. The hybrid MECO depends on the mass ratio and the spins of the black holes. It corresponds to a time earlier than t_m, which can be considered as the end of the inspiral phase for comparable mass binaries, but before the peak amplitude GW emission. Here the hybrid MECO time is computed using the maximum likelihood values for the masses and the spins of the individual black holes from the full IMR analysis.

In the parameter estimation performed here, we use uniform priors on merger time and the source-frame component masses. For the different analyses we use different prior ranges on the component masses. We keep the interval between the minimum and the maximum values of the component masses sufficiently large so that the posteriors have negligible values on the prior boundaries.

We also assume a distance prior uniform in comoving volume assuming a flat ΛCDM cosmological model. For the spins, we use uniform priors for the magnitude of the spin and isotropic for the orientation. We numerically marginalize over polarization. From each of the inspiral analyses we obtain the posterior distributions on the individual masses and spins. Finally, using the fitting formula in Refs. [58–60], we convert the initial masses and spins to the final mass and spin posterior.

IV. POST-TRUNCATION ANALYSIS

After the two individual black holes merge, the remnant is expected to settle down to a final stable black hole during the ringdown phase. Similar to the pretruncation analysis, we perform post-truncation analyses using the GW signal emitted during this merger and postmerger phase. As opposed to the previous analysis, here we excise the residual all the way up to the analysis start-time ("*gate-end-time*") to perform sampling on the data segment after the *gate-end-time* to obtain the posterior distribution on the final parameters.

We preform two sets of post-truncation analyses. In one, we use the late part of the same IMR waveform models as in the pretruncation analysis to obtain the posterior distribution on the masses and the spins. Using the fitting formula in Refs. [58–60] we then convert the component-object posteriors to a distribution on the final mass and spin of the remnant BH. For these cases we use similar priors as in the pre-truncation analyses: uniform priors on merger time and the source-frame component masses, and a distance prior uniform in comoving volume. For the spins, the priors are uniform in magnitude of the spin and isotropic for the orientation. We also numerically marginalize over polarization.

In the second set of post-truncation analyses we use damped sinusoids as the signal model. The signal emitted during this postmerger phase is conventionally called the "ringdown" signal and can be decomposed into a sum of exponentially damped sinusoids [61]. The gravitational waveform for the ringdown can be schematically written in terms of spin-weighted spheroidal harmonics as

$$ h_+ + ih_\times = \sum_{\ell,m,n} {}_{-2}S_{\ell mn}(\iota, \varphi) A_{\ell mn} e^{i(\Omega_{\ell mn}t + \phi_{\ell mn})}, \quad (8) $$

where the sum is over the various integer quantum numbers, $l \geq 2$, $-\ell \leq m \leq \ell$ and $n \geq 0$ denoting the different quasinormal modes. ι is the inclination angle, φ is the azimuthal angle of the black hole with respect to the observer, and ${}_{-2}S_{\ell mn}(\iota, \varphi)$ are the spin-weighted spheroidal harmonics. The spin-weighted spheroidal harmonics reduce to the usual spin-weighted spherical harmonics, $Y_{\ell m}$ for the nonspinning case. The various mode amplitudes $A_{\ell mn}$ and the phases $\phi_{\ell mn}$ depend on the initial configuration of the binary and on the particular theory of gravitation. In our analysis we use spheroidal harmonics [29] and treat the mode amplitudes and the phases as independent unknown parameters. The complex frequencies $\Omega_{\ell mn}$ consist of the quasinormal mode frequencies ($f_{\ell mn}$) and the damping times ($\tau_{\ell mn}$), which can be determined from the Teukolsky equation [62,63]. The no-hair theorem states that all ($f_{\ell mn}$) and ($\tau_{\ell mn}$) are determined by only two quantities, the mass M_f and spin χ_f of the black hole. Here, we use $\chi_f \in (-0.99, 0.99)$, with positive or negative values referring to perturbations that are co- or counterrotating with respect to the black hole's spin.

One can in principle vary many modes in the ringdown analysis with Bayesian inference. However, this increases the dimension of the parameter space leading to weaker constraints on the measured parameters if the contribution to the signal from additional modes is weak. Here we perform two types of ringdown analyses for each event. First we consider only the dominant fundamental mode, for which $\ell = m = 2$ and $n = 0$ (i.e., the [220] mode). Second, we include one overtone of the dominant mode (i.e.,

[220 + 221]), giving a signal template consisting of two modes. In a similar study of GW150914 in Ref. [31], the authors claimed the existence of the fundamental quasinormal mode and one overtone associated with the dominant angular mode ($\ell = m = 2$) with 3.6σ confidence.

In the ringdown analyses performed here, we vary the following parameters: final mass M_f, final spin χ_f, A_{220}, ϕ_{220}, and inclination ι in the single-mode 220 case, and add two additional parameters A_{221} and ϕ_{221} when considering the additional first overtone. We numerically marginalize over the polarization ψ. Furthermore, we use different analysis start times (or "*gate-end-times*") for the ringdown template. For each of these cases, the priors on final mass and final spin are assumed to be uniform in the following ranges

$$M_f \in [10, 200) \tag{9}$$

$$\chi_f \in [-0.99, +0.99). \tag{10}$$

For A_{220}, we choose a prior uniform in $\log_{10}$. We allow the overtone amplitude, A_{221}, to uniformly vary from zero to ten times that of A_{220} when we start the analysis at the merger time, t_m. When starting the analysis at later times, we choose a uniform prior so that $A_{221} < A_{220}$.

V. CONSISTENCY BETWEEN INSPIRAL-MERGER-RINGDOWN PHASES

Our results for GW150914 are summarized in Fig. 1. In the left panel of Fig. 1, we compare the posterior distributions from the various pretruncation analyses performed with different segments of the data and different waveform models. The innermost contour refers to the full IMR

analysis performed using the complete data segment with the NRSur7dq4 model. We find similar estimates with the IMRPhenomXPHM and IMRPhenomPv2 models. For all these cases we fix the sky location at $\alpha = 1.252$, $\delta = -1.224$, the maximum-likelihood values from [64].

For the pretruncation analyses we consider two different gate start times, after which the residual is excised for each of the three waveform models. The first one is the merger time t_m, where we use the estimate from [15], 1, 126, 259, 462.423 s GPS time at the LIGO Hanford site. Using the sky-location quoted above, this corresponds to 1, 126, 259, 462.411 s in geocentric GPS time. The second one is the hybrid MECO time, $t_{\text{hmeco}} = 1, 126, 259,$ 462.388 s in geocentric GPS time, which is ~23 ms earlier than t_m. Different waveform models provide different constraints on both the final mass and final spin parameters, but all of them are consistent with the full IMR analysis.

For a fixed waveform model, we obtain better constraints for the "gate-start-time" at t_m (denoted by "IMR model before t_m") compared to t_{hmeco} ("IMR model before t_{hmeco}") in left panel of Fig. 1). This is because later times include signal power from the late inspiral regimes. The posterior contours shrink and converge to the full IMR values as more data are included. Comparing between different waveform models, we find that the strongest constraints are obtained with the NRSur7dq4 model.

The different post-truncation analyses are shown in the right panel of Fig. 1. First, we consider the same waveform models as in the pretruncation analyses, where we excise the residual before t_{hmeco}, t_m and $t_m + 3$ ms. We find that all three waveform models provide consistent results for each of the gate-end-times. Hence we only show the posterior arising
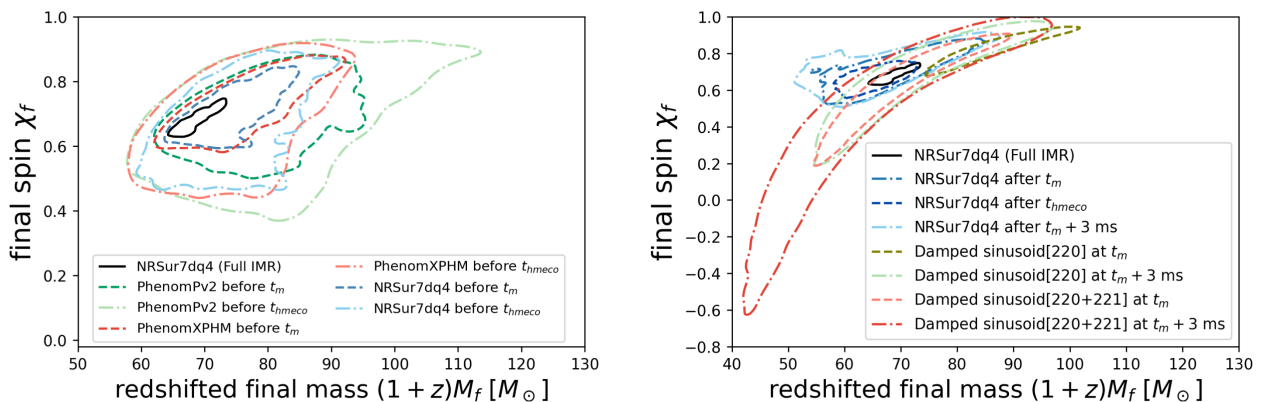

FIG. 1. The 90% posterior contours of the redshifted final mass (detector frame) M_f and the final spin χ_f obtained from the various pre- and post-truncation analyses for GW150914. In the left panel, we show the contours from the two pre-truncation analyses performed with two *gate-start-times*, t_m and t_{hmeco} using three different waveform models, IMRPhenomPv2, IMRPhenomXPHM and NRSur7dq4. On the right panel, we show the final mass-final spin contours obtained using damped sinusoids with only the dominant mode [220] and the dominant mode with one overtone [220 + 221] at different analysis start times (or the "gate-end-times") as the representatives of the post-truncation analyses. We also show three post-truncation analyses with the NRSur7dq4 model for the three "gate-end-times", t_{hmeco}, t_m, and $t_m + 3$ ms. The posterior contour arising from the full IMR analysis with NRSur7dq4 is shown by the innermost contour. The full IMR contour is contained within the contours of all the truncated analyses.

from the NRSur7dq4 model. As is evident from the figure, the posterior distribution arising from choosing t_{hmeco} as "gate-end-time" provides the most stringent constraints on the parameters and the closest to the IMR values among these three. This is expected since in this case we exclude only the early inspiral segment of the data and include the late inspiral, merger, and the ringdown phases.

Second, we perform the ringdown analysis as described in Sec. IV. Here we choose two different start times for the ringdown models, t_m and $t_m + 3$ ms and for each of these, we do a single mode [220] analysis with only the dominant $\ell = m = 2, n = 0$ harmonic and another with the dominant harmonic and the first overtone [220 + 221], $\ell = m = 2$, $n = 1$. We find that at t_m, the dominant mode [220] analysis provides estimates with higher final mass and higher final spin than the IMR result. The 90% posterior contours are completely disjoint. This is expected: applying the ringdown analysis too early leads to too low a ringdown frequency and results in overestimating the final mass. However, including the first overtone we recover estimates consistent with the full IMR analysis. Starting both the single- and two-mode analyses at 3 ms, we find that the dominant mode analysis provides tighter constraints compared to the [220 + 221] analysis.

We find no support in our posterior for final masses above 100 $M_\odot$, but masses below 50 $M_\odot$ are supported for the quasinormal mode analysis after 3 ms with [220 + 221]. We also find that as we end the truncation at later times, the remaining signal becomes quieter and the constraints on final mass and final spin get broader. In comparison with Isi *et al.* [41], we use the same merger time, $t_m = 1, 126, 259, 462.423$ s GPS time at the Hanford detector site, but perform the analysis with a different sky location (see Ref. [31] for comparison). We find that changing the sky location does not affect our final result. We also find similar final mass and final spin estimates from the [220]-analysis at t_m and the [220 + 221] analysis at $t_m + 3$ ms as quoted in [41]. However, the estimates from the pretruncation analysis carried out before t_m with NRSur7dq4 differ from the results quoted in [41]. In this particular case, we find the final mass and final spin estimates to be $(1 + z)M_f \sim 73.3^{+9.3}_{-6.6}$ $M_\odot$ and $\chi_f \sim 0.72^{+0.08}_{-0.1}$ respectively.

We perform a similar set of analyses with GW170814 and present our results in Fig. 2. We fix the sky-location parameters of GW170814 to $\alpha = 0.8$, $\delta = -0.8$, the maximum likelihood values from the IMR analysis in [2]. We use two truncation times to start or end the gating here, the merger time t_m and hybrid MECO time t_{hmeco}. For the merger time, we use the maximum likelihood value of the coalescence time t_c from [2]. The hybrid MECO time is calculated from the waveform with the maximum likelihood parameters also from [2]. These are $t_m = 1, 186, 741, 861.527$ s and $t_{\text{hmeco}} = 1, 186, 741, 861.5136$ s in geocentric GPS time, corresponding to $t_m = 1, 186, 741, 861.531$ s and $t_{\text{hmeco}} = 1, 186, 741, 861.513$ s at the Hanford detector site.
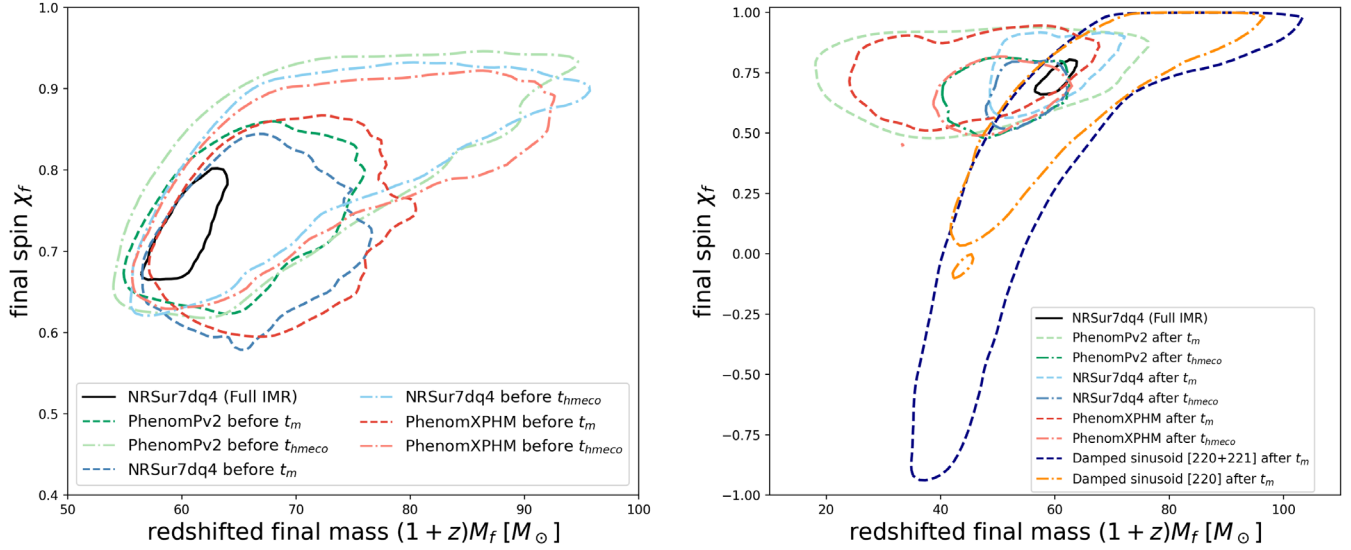

FIG. 2.   The 90% posterior contours of the redshifted final mass (detector frame) M_f and the final spin χ_f obtained from the various pre- and post-truncation analyses for GW170814. On the left panel we show the contours from the pretruncation analyses performed with two *gate-start-times*, t_m and t_{hmeco} using three different waveform models, IMRPhenomPv2, IMRPhenomXPHM and NRSur7dq4. On the right panel we show the 90% posterior contours of the redshifted final mass (detector frame) M_f and the final spin χ_f obtained using the damped sinusoids with only dominant mode [220] and the dominant mode with one overtone [220 + 221] at the merger time t_m as the representatives of the post-truncation analyses for GW170814. We also show the results for the two post-truncation analyses with two "gate-end-times", t_m and t_{hmeco} for IMRPhenomPv2, IMRPhenomXPHM and NRSur7dq4 waveform models. The ringdown contour is larger when the 221 mode is included because inclusion of the overtone increases the dimension of the parameter space. The posterior contour arising from the full IMR analysis with NRSur7dq4 waveform model is shown by the innermost solid black contour.

The estimated final mass-final spin contours from the pre-truncation analyses before each of the gate-times are presented in the left panel of Fig. 2. We also show the IMR results from the analysis of the full data segment with NRSur7dq4 waveform model. This is found to be consistent with all the different analyses.

In the case of the pre-truncation analyses, for a fixed gate-start-time, the different waveform models provide very similar results. The post-truncation analyses are presented in the right panel of Fig. 2. Here we find that at fixed gate-end-time, the NRSur7dq4 model provides the most stringent constraints. In addition to these IMR waveform models we show our estimates from the two ringdown analyses with the [220] and the [220 + 221] modes performed at t_m in right panel of Fig. 2. We find that both estimates are consistent with the IMR values. However the analysis with only the dominant mode provides better constraints on the parameters. Due to the post-merger signal being quiet, we do not

find reasonable constraints on the final mass and final spin for any analysis performed beyond t_m and omit them in the figure.

VI. AREA THEOREM

In this section we investigate the validity of the area theorem from the estimated parameters described in the previous sections. In Table I we report the 90% bounds on the fractional change of the black hole horizon area, $\Delta A = (A_f^{\text{measured}} - A_i^{\text{measured}})/A_i^{\text{measured}}$, for the five (four) different gate-times and three different waveform models for GW150914 (GW170814). Here $A_i^{\text{measured}} = A_1^{\text{measured}} + A_2^{\text{measured}}$ is computed using the initial component masses and spins from the pretruncation analyses and A_f is the final area estimated from the post-truncation parameters. A graphical representation of the same is provided in Fig. 3 for GW150914 (left panel) and GW170814 (right panel). In

TABLE I. The median values and 90% bounds on the fractional change of the BH horizon areas with different IMR waveform models at different truncation times. The first column corresponds to the different combinations of pre- and post-truncation analyses as indicated by the various rows. Here "IMR model" indicates the different IMR waveform models used in parameter estimation, which is denoted in the header of the remaining three different columns. The percentages of the posteriors that have positive change in area are given in parentheses. For GW170814, due to low SNR ($\sim$7) of the signal in post-truncation segment after t_m, we find larger posterior bounds on the parameters from IMRPhenomXPHM and IMRPhenomPv2 model, resulting in negative relative change in black hole horizon area.

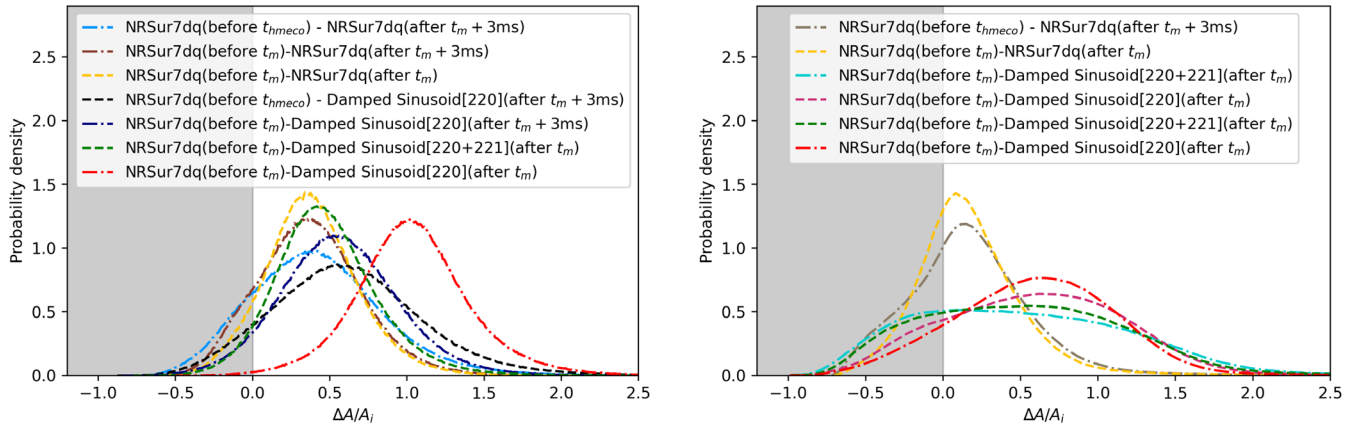| GW150914	$\Delta A/A_i^{\text{measured}}$ (NRSur7dq4)	$\Delta A/A_i^{\text{measured}}$ (IMRPhenomXPHM)	$\Delta A/A_i^{\text{measured}}$ (IMRPhenomPv2)
IMR model before t_m-Damped Sinusoid [220] after t_m	$1.03^{+0.66}_{-0.58}(99.7\%)$	$0.99^{+0.69}_{-0.66}(99.2\%)$	$0.83^{0.74}_{-0.82}(95.1\%)$
IMR model before t_m-Damped Sinusoid [220 + 221] after t_m	$0.46^{+0.61}_{-0.48}(94.3\%)$	$0.43^{+0.63}_{-0.52}(91.3\%)$	$0.31^{+0.66}_{-0.60}(79.3\%)$
IMR model before t_m-Damped Sinusoid [220] after $t_m + 3$ ms	$0.56^{+0.66}_{-0.60}(93.8\%)$	$0.52^{+0.68}_{-0.62}(91.4\%)$	$0.38^{+0.73}_{-0.67}(81.3\%)$
IMR model before t_m-Damped Sinusoid [220 + 221] after $t_m + 3$ ms	$0.29^{+0.81}_{-0.64}(74.6\%)$	$0.25^{0.83}_{-0.64}(71.7\%)$	$0.13^{+0.85}_{-0.62}(61.0\%)$
IMR model before t_m-IMR model after t_m	$0.37^{+0.54}_{-0.47}(90.5\%)$	$0.33^{+0.59}_{-0.51}(85.5\%)$	$0.20^{+0.61}_{-0.57}(71.7\%)$
IMR model before t_m-IMR model after $t_m + 3$ ms	$0.37^{+0.59}_{-0.52}(86.6\%)$	$0.31^{+0.66}_{-0.58}(80.0\%)$	$0.18^{+0.68}_{-0.59}(67.6\%)$
IMR model before t_{hmeco}-Damped Sinusoid [220] after t_m	$1.1^{+0.94}_{-0.85}(98.6\%)$	$1.08^{+0.98}_{-0.99}(96.3\%)$	$0.76^{+1.12}_{-1.01}(85.2\%)$
IMR model before t_{hmeco}-Damped Sinusoid [220 + 221] after t_m	$0.51^{+0.81}_{-0.64}(90.1\%)$	$0.49^{+0.83}_{-0.73}(87.2\%)$	$0.25^{+0.92}_{-0.73}(67.7\%)$
IMR model before t_{hmeco}-Damped Sinusoid [220] after $t_m + 3$ ms	$0.6^{+0.87}_{-0.73}(90.6\%)$	$0.58^{+0.90}_{-0.81}(87.7\%)$	$0.32^{+1.01}_{-0.77}(70.0\%)$
IMR model before t_{hmeco}-Damped Sinusoid [220 + 221] after $t_m + 3$ ms	$0.4^{+0.78}_{-0.63}(83.6\%)$	$0.29^{+1.02}_{-0.74}(71.4\%)$	$0.06^{+1.09}_{-0.66}(54.4\%)$
IMR model before t_{hmeco}-IMR model after t_m	$0.41^{0.73}_{-0.61}(86.4\%)$	$0.38^{+0.78}_{-0.68}(82.3\%)$	$0.16^{+0.86}_{-0.67}(61.7\%)$
IMR model before t_{hmeco}-IMR model after $t_m + 3$ ms	$0.4^{+0.78}_{-0.63}(83.6\%)$	$0.36^{+0.85}_{-0.72}(78.0\%)$	$0.12^{+0.93}_{-0.67}(59.0\%)$
GW170814			
IMR model before t_m-Damped Sinusoid [220 + 221] after t_m	$0.52^{+1.12}_{-0.95}(77.3\%)$	$0.45^{+1.1}_{-0.90}(74.9\%)$	$0.67^{1.21}_{-1.05}(81.3\%)$
IMR model before t_m-Damped Sinusoid [220] after t_m	$0.63^{+0.81}_{-0.87}(87.8\%)$	$0.55^{+0.81}_{-0.83}(85.5\%)$	$0.8^{+0.86}_{-0.99}(90.4\%)$
IMR model before t_m-IMR model after t_m	$0.16^{+0.66}_{-0.46}(71.8\%)$	$-0.33^{+0.67}_{-0.41}(17.4\%)$	$-0.23^{1.15}_{-0.47}(31.9\%)$
IMR model before t_{hmeco}-IMR model after t_m	$0.17^{+0.71}_{-0.59}(69.1\%)$	$-0.35^{+0.75}_{-0.42}(19.1\%)$	$-0.31^{1.17}_{-0.45}(27.6\%)$
IMR model before t_{hmeco}-Damped Sinusoid [220 + 221] after t_m	$0.49^{+1.26}_{-0.96}(74.7\%)$	$0.40^{+1.29}_{-0.93}(70.8\%)$	$0.52^{1.29}_{-1.01}(74.7\%)$
IMR model before t_{hmeco}-Damped Sinusoid [220] after t_m	$0.62^{+0.93}_{-0.94}(83.9\%)$	$0.53^{+0.97}_{-0.94}(79.6\%)$	$0.66^{+0.94}_{-1.02}(83.5\%)$

FIG. 3. Fractional change in the black hole horizon area, $\Delta A/A_i^{\text{measured}}$, due to the BBH merger [GW150914 on the left and GW170814 on the right] for different pre- and post truncation analyses performed with NRSur7dq4 or the damped sinusoid models. The change in the area is denoted by $\Delta A = (A_f^{\text{measured}} - A_i^{\text{measured}})$ with A_i^{measured} being the initial area inferred using the parameters from the pretruncation analysis and A_f^{measured} is the final area measured with the estimated parameters form the post-truncation analysis.

these figures we only provide the results obtained using combinations of NRSur7dq4 and the quasinormal mode model for different data segments.

For GW150914 we obtain a 99% agreement with the area theorem when NRSur7dq4 waveform model before t_m is the pretruncation model and a single [220] damped sinusoid after t_m as the post-truncation model. The agreement reduces to 98% when the pretruncation analysis is limited to times before t_{hmeco}. This high agreement with the area theorem is due to the fact that the single [220] damped sinusoid model after t_m overestimates the final mass and final spin, and hence overestimates the final area.

A weaker constraint $\sim 94\%$ on the validity of the area theorem is obtained when the pretruncation analysis is carried out before t_m with the post-truncation analysis being the simple ringdown analyses with $[220 + 221]$ at t_m and just [220] at $t_m + 3$ ms (see the second and the third entry in Table I). These two estimates are slightly lower than the results in Isi *et al.* [41]. Here, in the pretruncation analysis before t_m, the prior ranges on the component masses (uniform between $11 M_\odot - 120 M_\odot$) are chosen to ensure that the posteriors have negligible support at the prior boundaries. Using tighter priors (uniform between $17 M_\odot - 76 M_\odot$, which excludes some part of posterior on the component masses), we find improved constraints on the area theorem, up to $\sim 97\%$, as found in Ref. [41]. We also find that using uniform priors on the component masses provides similar results if we use uniform priors on total mass and mass ratio, as was done in Ref. [41]. To be precise, we find that the difference in the prior distribution function (whether uniform on component masses or uniform in total mass and mass ratio) does not greatly affect our result, but that a larger prior boundary weakens the constraints on the area theorem.

As might be expected, an earlier truncation time (t_{hmeco}) for the pretruncation analysis or a later truncation time

(3 ms) for the post-truncation analysis gives weaker constraints on the area change.

Comparing different waveform models, we find much stronger constraints on the area theorem for GW150914 using NRSur7dq4 or IMRPhenomXPHM than using IMRPhenomPv2. This may be due to the fact that NRSur7dq4 and IMRPhenomXPHM models include subdominant modes, whereas IMRPhenomPv2 does not.

For comparison, we also provide a similar study on GW170814 data (see Table I). Compared to the GW150914 results, we find weaker bounds on the area theorem for GW170814. This is due to fact that GW170814 has a quieter post-merger signal. When the pretruncation analysis is extended only until t_{hmeco}, we find that IMRPhenomPv2 provides similar constraints to NRSur7dq4. On the other hand, when the pretruncation analysis is extended until the merger, the IMRPhenomPv2 analysis provides a stronger bound. We also find a negative change in black hole area when we use IMRPhenomPv2 or IMRPhenomXPHM in the post-truncation analyses after t_m. Due to lower SNR (~ 7) for the post-truncation data segment, we obtain larger posterior bounds on the final mass, having support from very low masses (as low as $\sim 20 M_\odot$). This leads to a negative change in the BH horizon area.

In order to test the consistency of our results with the actual (positive) change of area predicted by GR, we plot the ratio between the measured change and the expected change in area, $\mathcal{R} = (A_f^{\text{measured}} - A_i^{\text{measured}})/(A_f^{\text{expected}} - A_i^{\text{measured}})$, for GW150914 and GW170814 in Fig. 4. The expected final area, A_f^{expected} is computed by converting the initial parameters (used to compute A_i^{measured}) from the pre-truncation analyses to the final parameters assuming GR. All the pretruncation analyses are performed before t_{hmeco} to completely avoid the merger regime. For demonstration purposes
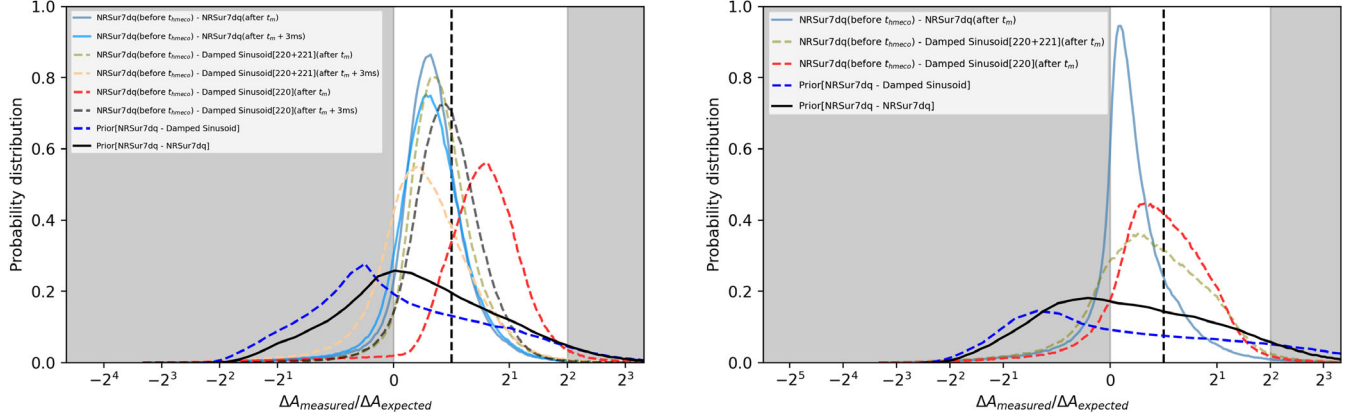
FIG. 4. Ratio of the measured and the expected change in black hole horizon area, $\mathcal{R} = (A_f^{\text{measured}} - A_i^{\text{measured}})/(A_f^{\text{expected}} - A_i^{\text{measured}})$, for GW150914 (left) and GW170814 (right). Here, A_i^{measured} is the initial area inferred using the parameters obtained from the pretruncation analysis. Shown are results using the NRSur7dq waveform model at different "*gate-start-times*," as indicated in the legend. The measured final area A_f^{measured} is computed from the post-truncation analysis, using either damped sinusoids ([220], [220 + 221]) or the NRSur7dq waveform model. The expected final area A_f^{expected} is computed by converting the initial parameters from the pretruncation analysis (used to compute A_i^{measured}) into expected final mass and spin via fits to GR waveforms [58–60]. The vertical black dashed line denotes the ideal scenario when the expected and the measured change of area are the same. We have also plotted the two different prior distributions used by the blue dashed and the black solid lines.

here we have only used NRSur7dq4 as the IMR model in all of the pretruncation analyses.

If the GR prediction is true, we expect $\mathcal{R}$ to be exactly 1 for a perfect measurement. Due to statistical uncertainties, the probability distribution on $\mathcal{R}$ is expected to be a Gaussian with mean equal to one. The priors on masses and spins used for the different pre- and post-truncation analyses are equivalent to a prior distribution on $\mathcal{R}$ that can be both negative as well as a positive, with $\mathcal{R} > 4$ corresponding to a violation of the conservation of energy (see the black solid and the blue dashed lines in Fig. 4).

For GW150914, the post-truncation analysis using the [220] damped sinusoid mode starting at t_m produces a probability distribution on $\mathcal{R}$ peaking at a value greater than 1. Consistently, for this particular case we find a better agreement with the area theorem (∼99% [see Table I]). This

is expected, as the area theorem only requires the final area to be larger than the initial area, so the agreement is improved when the distribution for final area is shifted to higher values. The final mass and spin are overestimated for this case as seen in the right panel of Fig. 1, resulting in overestimating the final area, with the bias visible in Fig. 4.

On the other hand, when we consider the first overtone with the dominant mode after $t_m + 3$ ms, we find a shift in the distribution to the opposite direction and the agreement with the area theorem drops. To quantify the agreement with GR, we also quote the mismatch, $C_{\mathcal{R}}$ for each of the curves in Fig. 4 in Table II. Here the mismatch $C_{\mathcal{R}}$ denotes the probability of getting R within the range of values symmetric about the median value of R, extending to $R = 1$ for each of the curves. To be precise, $C_{\mathcal{R}}$ denotes the area under the curve bounded by $\mathcal{R} = 1$ and symmetric around

TABLE II. Mismatch $C_{\mathcal{R}}$ for the recovery of $\mathcal{R} = 1$ using the various pre- and post-truncation analyses for GW150914 and GW170814. Lower value of the mismatch $C_{\mathcal{R}}$, denotes greater agreement with GR predictions. The IMR model used here is NRSur7dq.

GW150914	$C_{\mathcal{R}}$ (NRSur7dq4)
IMR model before t_{hmeco}-Damped Sinusoid [220] after t_m	67.8%
IMR model before t_{hmeco}-Damped Sinusoid [220 + 221] after t_m	27.6%
IMR model before t_{hmeco}-Damped Sinusoid [220] after $t_m + 3$ ms	6.5%
IMR model before t_{hmeco}-Damped Sinusoid [220 + 221] after $t_m + 3$ ms	42.3%
IMR model before t_{hmeco}-IMR model after t_m	47.5%
IMR model before t_{hmeco}-IMR model after $t_m + 3$ ms	44.6%
GW170814	
IMR model before t_{hmeco}-IMR model after t_m	61.2%
IMR model before t_{hmeco}-Damped Sinusoid [220] after t_m	2.7%
IMR model before t_{hmeco}-Damped Sinusoid [220 + 221] after t_m	7.9%

the median value. Lower values of $C_{\mathcal{R}}$ refer to higher accuracy in recovery of the GR estimates. We find that excluding the merger and starting the post-truncation analysis with [220] at $t_m + 3$ ms provides a better agreement with GR having $C_{\mathcal{R}} \sim 6.5\%$ as compared to other cases where we use different waveform models (IMR model or the damped sinusoids with [220 + 221]) after t_m or $t_m + 3$ ms.

A similar study for GW170814 is also given in Fig. 4 and Table II. As opposed to the GW150914 result, here we find that the dominant mode post-truncation analysis recovers the GR value with greater accuracy, with the mismatch only $C_{\mathcal{R}} = 2.9\%$. As opposed to this, using the NRSur7dq model for the post-truncation analysis leads to lower agreement with GR.

VII. CONCLUSION

In this paper we provide an extensive study on the validity of Hawking's area theorem and the consistency of different phases of the compact binary signal using GW150914 and GW170814 data. We investigate how the different waveform models, various prior distributions, and use of different data segments affect the final results. We observe that uniform priors on component masses or the total mass and mass ratio yield similar results.

For both GW150914 and GW170814, different waveform models provide different constraints on both the final mass and final spin parameters, but all of them are consistent with the full IMR analysis except one. The ringdown analysis of GW150914 at t_m considering only the dominant mode overestimates the final mass and final spin for the remnant black hole. However, considering one additional overtone provides consistent bounds when compared to the result obtained by analysing the full data segment using IMR waveform models. In the case of GW170814, we find that the dominant mode analysis and the [220 + 221] mode analysis give consistent bounds.

We observe that the different choices for the excision of data and for the waveform models introduce significant systematic errors in the measurements of the validity of the area theorem. In the case of GW150914, for various combinations of excised data, the probability of the validity of the area theorem varies in the range $\sim 74\%$–94% when we use the NRSur7dq4 waveform model in the pretruncation analysis. Using instead the waveform model IMRPhenomXPHM, the range drops by $\sim 3\%$, to $\sim 71\%$–91%. IMRPhenomPv2 leads to an even broader range from 54% to 85%.

In Ref. [40], a study of a similar test of the area theorem was presented based on a simulated GW150914-like signal. This study used IMRPhenomPv2 for the inspiral and a [220] damped sinusoid for the ringdown, finding support for the validity of the theorem of $\sim 74\%$. Our analysis of the GW150914 data shows a similar result of $\sim 70\%$ when

using the corresponding waveform model choices (see 9th row on the 4th column of Table I).

When considering different waveform models along with the different choices of data duration, the probability on the validity of the area theorem varies in the range 54%–94%. We also see that using a damped sinusoid as the waveform model for the post-truncation analysis improves the probability by $\sim 10\%$ on average as compared to one of the updated IMR waveform models. We also see that if we apply the ringdown analysis with the dominant mode [220] at t_m for GW150914, we get a lower ringdown frequency. This overestimates the final mass, making an area theorem test biased in a direction that likely overstates the agreement, with the probability being $\sim 99\%$. For this reason, we believe the ratio of the measured change in area to the expected change $\mathcal{R}$, and the associated mismatch $C_{\mathcal{R}}$, to be a better metric for determining the consistency of the signal with GR. This yields weaker agreement with GR for the [220] ringdown analysis at t_m, for which $C_{\mathcal{R}} \sim 67\%$. This is expected, since a ringdown with only the [220] is not thought to be good model of the signal at merger. For comparison, the [220 + 221] ringdown analysis at $t_m + 3$ ms yields $C_{\mathcal{R}} \sim 6.7\%$.

For GW170814, various data durations used in the analyses lead to larger uncertainty on the probability values as compared to GW150914. For NRSur7dq4 it ranges in between 71%–87%. However, for IMRPhenomXPHM and IMRPhenomPv2 it varies from 17%–85% and 27%–90%, respectively. To compare between a damped sinusoid and an IMR waveform model for the post-truncation analysis, we find that the damped sinusoid improves the probability of a positive change in area by $\sim 15\%$ as compared to the case where the IMR model is NRSur7dq4. But the improvement could be as high as $\sim 60\%$ when the IMR models used are IMRPhenomXPHM or IMRPhenomPv2. The similar findings are reflected in Table II through the values of $C_{\mathcal{R}}$. In this case the best case scenario is obtained using an IMR model before t_{hmeco} and the [220] mode analysis after t_m, for which $C_{\mathcal{R}} = 2.7\%$. However, using IMR model to analyze the post-truncation segment after t_m, $C_{\mathcal{R}}$ drops to $\sim 61\%$.

These large systematic uncertainties highlight the need for binary black hole observations across a longer time period than is possible with current generation detectors. This should become possible in the 2030s with the launch of the space-based Laser Interferometer Space Antenna (LISA) [65] and the beginning of ground-based "3G" detectors, such as Cosmic Explorer (CE) [66] and Einstein Telescope (ET) [67]. With a sensitive frequency band of 0.1–100 mHz, LISA will be able to detect binary black holes $\sim$ a year before their merger are detected by ground-based detectors [25,68–74]. This should yield unprecedented precision measurements of the fundamental laws governing black hole thermodynamics.

[1] B. P. Abbott *et al.* (Virgo, LIGO Scientific Collaborations), Phys. Rev. Lett. **116,** 061102 (2016).

[2] A. H. Nitz, C. D. Capano, S. Kumar, Y.-F. Wang, S. Kastha, M. Schfer, R. Dhurkunde, and M. Cabero, Astrophys. J. **922,** 76 (2021).

[3] B. P. Abbott *et al.* (Virgo, LIGO Scientific Collaborations), Phys. Rev. Lett. **116,** 241103 (2016).

[4] B. P. Abbott *et al.*, Phys. Rev. Lett. **118,** 221101 (2017).

[5] B. P. Abbott *et al.* (Virgo, LIGO Scientific Collaborations), Astrophys. J. **851,** L35 (2017).

[6] B. P. Abbott *et al.* (Virgo, LIGO Scientific Collaborations), Phys. Rev. Lett. **119,** 141101 (2017).

[7] B. P. Abbott *et al.* (LIGO Scientific, Virgo Collaborations), Phys. Rev. X **9,** 031040 (2019).

[8] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, Phys. Rev. D **101,** 083030 (2020).

[9] B. Zackay, L. Dai, T. Venumadhav, J. Roulet, and M. Zaldarriaga, Phys. Rev. D **104,** 063030 (2021).

[10] R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. Adhikari, V. Adya, C. Affeldt *et al.*, Phys. Rev. X **11,** 021053 (2021).

[11] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt *et al.*, Astrophys. J. Lett. **913,** L7 (2021).

[12] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos *et al.*, Astrophys. J. **915,** 86 (2021).

[13] R. Abbott *et al.* (LIGO Scientific, Virgo, KAGRA Collaborations, arXiv:2111.03606 [Phys. Rev. X (to be published)].

[14] B. P. Abbott *et al.* (LIGO Scientific, Virgo Collaborations), Phys. Rev. D **100,** 104036 (2019).

[15] R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. Adhikari, V. Adya, C. Affeldt *et al.*, Phys. Rev. D **103,** 122002 (2021).

[16] K. G. Arun, B. R. Iyer, M. S. S. Qusailah, and B. S. Sathyaprakash, Classical Quant. Grav. **23,** L37 (2006).

[17] K. G. Arun, B. R. Iyer, M. S. S. Qusailah, and B. S. Sathyaprakash, Phys. Rev. D **74,** 024006 (2006).

[18] K. G. Arun, Classical Quant. Grav. **29,** 075011 (2012).

[19] N. Yunes and F. Pretorius, Phys. Rev. D **80,** 122003 (2009).

[20] C. K. Mishra, K. G. Arun, B. R. Iyer, and B. S. Sathyaprakash, Phys. Rev. D **82,** 064010 (2010).

[21] M. Agathos, W. Del Pozzo, T. G. F. Li, C. V. D. Broeck, J. Veitch, and S. Vitale, Phys. Rev. D **89,** 082001 (2014).

[22] T. G. F. Li, W. Del Pozzo, S. Vitale, C. Van Den Broeck, M. Agathos, J. Veitch, K. Grover, T. Sidery, R. Sturani, and A. Vecchio, Phys. Rev. D **85,** 082003 (2012).

[23] J. Meidam *et al.*, Phys. Rev. D **97,** 044033 (2018).

[24] S. Kastha, A. Gupta, K. Arun, B. Sathyaprakash, and C. Van Den Broeck, Phys. Rev. D **100,** 044007 (2019).

[25] A. Gupta, S. Datta, S. Kastha, S. Borhanian, K. Arun, and B. Sathyaprakash, Phys. Rev. Lett. **125,** 201101 (2020).

[26] S. Kastha, A. Gupta, K. G. Arun, B. S. Sathyaprakash, and C. Van Den Broeck, Phys. Rev. D **98,** 124033 (2018).

[27] N. Cornish, L. Sampson, N. Yunes, and F. Pretorius, Phys. Rev. D **84,** 062003 (2011).

[28] A. Ghosh *et al.*, Phys. Rev. D **94,** 021101 (2016).

[29] C. D. Capano, M. Cabero, J. Westerweck, J. Abedi, S. Kastha, A. H. Nitz, A. B. Nielsen, and B. Krishnan, arXiv:2105 .05238.

[30] C. D. Capano and A. H. Nitz, Phys. Rev. D **102,** 124070 (2020).

[31] M. Isi, M. Giesler, W. M. Farr, M. A. Scheel, and S. A. Teukolsky, Phys. Rev. Lett. **123,** 111102 (2019).

[32] S. Dhanpal, A. Ghosh, A. K. Mehta, P. Ajith, and B. Sathyaprakash, Phys. Rev. D **99,** 104056 (2019).

[33] S. W. Hawking, Phys. Rev. Lett. **26,** 1344 (1971).

[34] S. W. Hawking, Commun. Math. Phys. **25,** 152 (1972).

[35] P. T. Chrusciel, E. Delay, G. J. Galloway, and R. Howard, Ann. Henri Poincare **2,** 109 (2001).

[36] A. Ashtekar and B. Krishnan, Phys. Rev. Lett. **89,** 261101 (2002).

[37] D. Pook-Kolb, O. Birnholtz, B. Krishnan, and E. Schnetter, Phys. Rev. Lett. **123,** 171102 (2019).

[38] R. Bousso and N. Engelhardt, Phys. Rev. Lett. **115,** 081301 (2015).

[39] S. A. Hughes and K. Menou, Astrophys. J. **623,** 689 (2005).

[40] M. Cabero, C. D. Capano, O. Fischer-Birnholtz, B. Krishnan, A. B. Nielsen, A. H. Nitz, and C. M. Biwer, Phys. Rev. D **97,** 124069 (2018).

[41] M. Isi, W. M. Farr, M. Giesler, M. A. Scheel, and S. A. Teukolsky, Phys. Rev. Lett. **127,** 011103 (2021).

[42] B. P. Abbott *et al.* (Virgo, LIGO Scientific Collaborations), Phys. Rev. Lett. **116,** 221101 (2016).

[43] S. Bhagwat, M. Okounkova, S. W. Ballmer, D. A. Brown, M. Giesler, M. A. Scheel, and S. A. Teukolsky, Phys. Rev. D **97**, 104065 (2018).

[44] T. A. Apostolatos, C. Cutler, G. J. Sussman, and K. S. Thorne, Phys. Rev. D **49**, 6274 (1994).

[45] P. Jaranowski, A. Krlak, and B. F. Schutz, Phys. Rev. D **58**, 063001 (1998).

[46] P. Jaranowski and A. Krolak, Living Rev. Relativity **8**, 3 (2005).

[47] M. H. A. Davis, A review of the statistical theory of signal detection, in *Gravitational Wave Data Analysis*, edited by B. F. Schutz (Springer, Netherlands, Dordrecht, 1989), pp. 73–94.

[48] C. M. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, Publ. Astron. Soc. Pac. **131**, 024503 (2019).

[49] J. S. Speagle, Mon. Not. R. Astron. Soc. **493**, 3132 (2020).

[50] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, Publ. Astron. Soc. Pac. **125**, 306 (2013).

[51] W. D. Vousden, W. M. Farr, and I. Mandel, Mon. Not. R. Astron. Soc. **455**, 1919 (2016).

[52] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Phys. Rev. D **93**, 044006 (2016).

[53] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Prrer, X. J. Forteza, and A. Boh, Phys. Rev. D **93**, 044007 (2016).

[54] G. Pratten, C. Garca-Quirs, M. Colleoni, A. Ramos-Buades, H. Estells, M. Mateu-Lucena, R. Jaume, M. Haney, D. Keitel, J. E. Thompson *et al.*, Phys. Rev. D **103**, 104056 (2021).

[55] LALSuite, LSC Algorithm Library Suite, https://git.ligo .org/lscsoft/lalsuite (2020).

[56] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, Phys. Rev. Research **1**, 033015 (2019).

[57] M. Cabero, A. B. Nielsen, A. P. Lundgren, and C. D. Capano, Phys. Rev. D **95**, 064016 (2017).

[58] J. Healy, C. O. Lousto, and Y. Zlochower, Phys. Rev. D **90**, 104004 (2014).

[59] F. Hofmann, E. Barausse, and L. Rezzolla, Astrophys. J. **825**, L19 (2016).

[60] X. Jimnez-Forteza, D. Keitel, S. Husa, M. Hannam, S. Khan, and M. Prrer, Phys. Rev. D **95**, 064024 (2017).

[61] E. Berti, V. Cardoso, and A. O. Starinets, Classical Quant. Grav. **26**, 163001 (2009).

[62] S. A. Teukolsky, Phys. Rev. Lett. **29**, 1114 (1972).

[63] E. W. Leaver, Proc. R. Soc. A **402**, 285 (1985).

[64] A. H. Nitz, T. Dent, G. S. Davies, S. Kumar, C. D. Capano, I. Harry, S. Mozzon, L. Nuttall, A. Lundgren, and M. Tápai, Astrophys. J. **891**, 123 (2020).

[65] P. Amaro-Seoane *et al.*, arXiv:1702.00786.

[66] D. Reitze *et al.*, Bull. Am. Astron. Soc. **51**, 035 (2019).

[67] M. Punturo, M. Abernathy, F. Acernese, B. Allen *et al.*, Classical Quant. Grav. **27**, 194002 (2010).

[68] A. Sesana, Phys. Rev. Lett. **116**, 231102 (2016).

[69] E. Barausse, N. Yunes, and K. Chamberlain, Phys. Rev. Lett. **116**, 241104 (2016).

[70] A. Toubiana, S. Marsat, E. Barausse, S. Babak, and J. Baker, Phys. Rev. D **101**, 104038 (2020).

[71] C. Liu, L. Shao, J. Zhao, and Y. Gao, Mon. Not. R. Astron. Soc. **496**, 182 (2020).

[72] Z. Carson and K. Yagi, Classical Quant. Grav. **37**, 02LT01 (2020).

[73] G. Gnocchi, A. Maselli, T. Abdelsalhin, N. Giacobbo, and M. Mapelli, Phys. Rev. D **100**, 064024 (2019).

[74] S. Datta, A. Gupta, S. Kastha, K. G. Arun, and B. S. Sathyaprakash, Phys. Rev. D **103**, 024036 (2021).

[74] S. Datta, A. Gupta, S. Kastha, K. G. Arun, and B. S. Sathyaprakash, Phys. Rev. D **103**, 024036 (2021).

[75] https://www.gw-openscience.org/.