Investigation on the Effects of Non-Gaussian Noise Transients and Their Mitigations on Gravitational-Wave Tests of General Relativity

Jack Y. L. Kwok,1,* Rico K. L. Lo,2 Alan J. Weinstein,2 and Tjonnie G. F. Li1,3,4

1Department of Physics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
2LIGO, California Institute of Technology, Pasadena, California 91125, USA
3Institute for Theoretical Physics, KU Leuven, Celestijnenlaan 200D, B-3001 Leuven, Belgium
4Department of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

(Dated: October 12, 2021)

The detection of gravitational waves from compact binary coalescence by Advanced LIGO and Advanced Virgo provides an opportunity to study the strong-field, highly-relativistic regime of gravity. Gravitational-wave tests of General Relativity (GR) typically assume Gaussian and stationary detector noise, thus do not account for non-Gaussian, transient noise features (glitches). We present the results obtained by performing parameterized gravitational-wave tests on simulated signals from binary-black-hole coalescence overlapped with three classes of frequently occurring instrumental glitches with distinctly different morphologies. We then review and apply three glitch mitigation methods and evaluate their effects on reducing false deviations from GR. By considering 9 cases of glitches overlapping with simulated signals, we show that the short-duration, broadband blip and tomtite glitches under consideration introduce false violations of GR, and using an inpainting filter and glitch model subtraction can consistently eliminate such false violations without introducing additional effects.

I. INTRODUCTION

Over a century after its formulation in 1915, Einstein’s General Relativity (GR) remains as the accepted theory of gravity, passing all precision tests to date [1]. In the weak-field, slow-motion regime, where the effects of metric theories of gravity can be approximated as higher-order post-Newtonian (PN) corrections to the Newtonian theory [2], GR lies within the stringent bounds set by solar-system tests and pulsar tests [3, 4]. Recent attention has turned to testing GR in the strong-field, highly-relativistic regime [3], which potentially suggests high-energy corrections to the Einstein-Hilbert action [5], making GR compatible with standard quantum field theory [1]. One approach of probing the strong-field regime is through the detection of gravitational waves (GWs), which propagates at the speed of light and carries information about its astrophysical origin [6].

Of all strong-field astrophysical events that could be probed using GWs, the coalescence of stellar-mass binary black holes (BBHs), which can be schematically divided into inspiral, merger and ringdown (IMR) stages, plays a crucial role in testing GR [1]. Since the orbital separation of BBHs can reach far below the last stable orbit before merging, the generated gravitational field can be many order of magnitudes stronger than other astrophysical events observed so far [7–13]. Moreover, GWs emitted by coalescing BBHs offers one of the cleanest test of GR, as matter and electromagnetic fields are negligible for most sources [8, 14], and the emitted GWs essentially propagate through matter unimpeded [8], enabling precision tests of the strong-field dynamics of GR.

Since 2015, Advanced LIGO [15] and Advanced Virgo [16] have jointly announced over 40 confident detections of GWs from coalescing BBHs [17, 18].

Several GW tests of GR using coalescing BBHs are developed to test for generic deviations from GR without the need for signal models from competing theories of gravity [8]. For example, consistency tests search for excess power in the residual noise after subtracting a best-fit GR waveform [19], or compare the source parameters inferred using only high-frequency data to that inferred using only low-frequency data [19]; parameterized tests introduce parameterized deformations to waveform approximations to GR and infer the extent of deviation using Bayesian parameter estimation [9]. To this date, no evidence for violations of GR has been identified using GWs emitted by coalescing BBHs [20, 21].

Aside from GWs, output from GW detectors is attributed to many independent sources of random noise [22]. Assuming that noise characteristics remain stationary over observation timescales, detector noise is typically modeled as stationary and Gaussian in GW data analysis in light of the central limit theorem [23, 24]. However, these assumptions cannot account for transient, non-Gaussian noise features, commonly referred to as glitches [25–27]. Glitches pose significant problems to GW searches [26] and may bias GW data analysis by violating the noise model. Three glitches from commonly-seen glitch classes during the O3 observing run are shown in Fig. 1.

Many efforts are made to identify and classify glitches [26, 28–32]. Once a glitch is identified, the data containing the glitch can be removed using various mitigation methods [33–37]. The effects of glitches and their mitigations to the inference of source parameters have been studied in the context of glitches similar to that affecting...
GW170817 [38]. It is of interest to extend the study to parameterized tests of GR, as the additional degree(s) of freedom introduced by parameterized deformations of the signal model may enhance such effects.

This article is structured as follows: Sec. II describes the typical data model used in GW data analyses [23, 24], which composes of a GW signal in additive stationary and Gaussian noise. Sec. III introduces a parameterized test of GR involving the phase parameterization of an IMR waveform model [40]. Sec. IV introduces three glitch mitigation methods to be applied in our investigation, namely band-pass filtering, inpainting and glitch model subtraction. Sec. V presents the results obtained by performing the parameterized test of GR to glitch-overlapped BBH-coalescence GW signals before and after glitch mitigations.

II. DATA MODEL

A GW detector is designed to respond linearly to the fractional change in arm length, or strain [22]. The time series of detector output data $d$, sampled at time $t_k$ at constant sampling interval $\Delta t$, can thus be expressed as a linear superposition of a time series of the GW strain signal $h$ and a time series of detector noise $n$:

$$d(t_k) = h(t_k) + n(t_k) .$$

In Eq. (1) and in subsequent discussion, boldface denotes the matrix representation of specified quantities.

A. Stationary Gaussian Noise Model

Assuming that a large number of independent noise sources contribute linearly to the detector noise $n$, the central limit theorem states that the probability density distribution of the noise $n$ tends to follow a multivariate Gaussian distribution [41]:

$$P(n) = \frac{1}{\sqrt{(2\pi)^N|\Sigma|}} e^{-\frac{1}{2}(n-\mu)^T\Sigma^{-1}(n-\mu)} ,$$

which is uniquely defined by the covariance matrix $\Sigma_{ij} = E[(n(t_i) - \mu(t_i))(n(t_j) - \mu(t_j))]$ and the mean vector $\mu_i = E[n(t_i)]$, where $E[\cdot]$ and $|\cdot|$ denotes the expectation and determinant operation respectively. The diagonal (off-diagonal) terms of the covariance matrix are the variances at each instance of time (correlations between data from different instances of time).

If the number of samples $N$ is large, it is undesirable to invert the $N \times N$ covariance matrix in Eq. (2). Instead, we consider the joint probability density in frequency domain, which is also a multivariate Gaussian distribution [41]. With the assumption of stationarity, i.e. the joint probability density distribution is time-invariant, the covariance matrix in frequency domain is diagonalized in the infinite-duration limit [42]. This relation can be approximated for the finite-duration discretely-sampled time series, giving the following approximation to the joint probability density in frequency domain [42] (for even $N$), also known as the Whittle likelihood [43] in the context of statistical inference:

$$P(n) \approx \prod_{j=0}^{N/2-1} \frac{2\Delta f}{\pi S_n(f_j)} \exp \left( -\Delta f \frac{2|\tilde{n}_j|^2}{S_n(f_j)} \right) ,$$

where $f_j \equiv j/N\Delta t$. The quantity $S_n(f_j) \equiv 2|\tilde{n}(f_j)|^2/T$ is scaled from the diagonal terms of the covariance matrix in frequency domain, $\Delta f \equiv 1/T$ is the frequency resolution and the tilde denotes a discrete Fourier transformed (DFT) quantity:

$$\tilde{n}_j \equiv \Delta t \text{ DFT}[n(t_k)] = \Delta t \sum_{k=0}^{N-1} n(t_k)e^{-2\pi ijk/N} .$$

To motivate the quantity $S_n(f_j)$, called the one-sided power spectral density (PSD), we invoke Parseval’s theorem [42]:

$$\sum_{j=0}^{N/2-1} S_n(f_j)\Delta f \equiv \frac{2}{T} \sum_{j=0}^{N/2-1} |\tilde{n}(f_j)|^2\Delta f = \frac{1}{T} \sum_{k=0}^{N-1} |n(t_k)|^2\Delta t ,$$

FIG. 1. Glitches with similar morphology can be algorithmically categorized into different classes [26]. A time-frequency representation, called a Q-scan (or Omega scan) [39], where the duration of each time-frequency bins varies inversely with frequency and linearly with a parameter $Q$, is commonly used to visualize glitches [26, 32]. Q-scans of three frequently-occurring glitches (top-left: blip, top-right: tomte, bottom: scattered-light) during the O3 observing run are shown. The value of $Q$ used is 8, 8 and 40 respectively. The colour represents the normalized amplitude (square root of the normalized power) in each time-frequency bin.

A GW detector is designed to respond linearly to the fractional change in arm length, or strain [22]. The time series of detector output data $d$, sampled at time $t_k$ at constant sampling interval $\Delta t$, can thus be expressed as a linear superposition of a time series of the GW strain signal $h$ and a time series of detector noise $n$: $d(t_k) = h(t_k) + n(t_k)$.
and note that the rightmost side of Eq. (5) returns the power of the time series. Since a time series is real, we have \( \tilde{n}(f_j) = \tilde{n}^*(-f_j) \). Consequently, we can sample only the frequency bins from 0 Hz to up to the Nyquist frequency \( 1/2\Delta t \), and introduce the factor of 2 in Eq. (3) and Eq. (5).

### B. Signal Model

Since the two-body self-gravitating problem cannot be solved analytically in GR, we generate simulated GW strain signals from coalescing BBHs using the frequency-domain precessing IMR waveform model IMRPhenomPv2 [40] in virtue of its good match with Numerical Relativity (NR) waveforms [44] and low computational costs.

IMRPhenomPv2 is a phenomenological waveform model constructed by combining PN-like inspiral waveforms with NR-calibrated merger-ringdown ansatz [45]. In natural units, the inspiral stage of IMRPhenomPv2 is modeled up to \( f = 0.018/M \), where \( M \) is the total mass of the system. The region with \( Mf \geq 0.018 \) is subdivided into an intermediate stage with \( 0.018 \geq Mf \geq 0.5f_{\text{RD}} \), which bridges the inspiral stage to the merger-ringdown stage modeled above half the ringdown frequency \( f_{\text{RD}} \) [45]. Fig. 2 illustrates the stages of coalescence of an example IMRPhenomPv2 GW strain and its frequency evolution over time.

The phase of IMRPhenomPv2 composes of terms with known frequency dependence. The coefficients of these terms, denoted as the phase coefficients \( \delta p_i \), are the subjects of parameterized tests of GR to be discussed in Sec. III. The phase coefficients \( \delta p_i \) and the orbital evolution of the BBH depend only on the masses and spin angular momentum vectors of the component black holes [44], denoted as the intrinsic parameters. The phase coefficients \( \delta p_i \) can be categorized into three groups, depending on the stages of coalescence in which they predominantly assert their effect on [9, 45]: (i) the inspiral PN coefficients \( \{ \varphi_0, ..., \varphi_2, \varphi_3, \varphi_4, \varphi_6, \varphi_7 \} \) and phenomenological coefficients \( \{ \sigma_0, ..., \sigma_2 \} \); (ii) the intermediate phenomenological coefficients \( \{ \beta_0, ..., \beta_3 \} \); (iii) the merger-ringdown phenomenological and black hole perturbation theory coefficients \( \{ \alpha_0, ..., \alpha_5 \} \).

Seven additional extrinsic parameters, including the sky location, luminosity distance, polarization angle of the source, and the spatial orientation and orbital phase of the system at a reference frequency and time respectively, are needed to determine the response of the GW detectors.

### III. PARAMETERIZED TESTS OF GR

We will focus on a test of GR which tests for parameterized deviations from GR. It assumes the stationary Gaussian noise model, as such this test provides quantitative indicators of whether glitches can result in false deviations of GR and whether glitch mitigations will reduce or amplify them.

In the test, fractional deviations \( \delta p_i \), also known as de-phasing coefficients, are introduced to IMRPhenomPv2 phase coefficients \( p_i \) [9]:

\[
\delta p_i \rightarrow p_i[1 + \delta p_i].
\]

For the exceptional case where \( p_i = 0 \), such as \( \varphi_1 \), an absolute deviation is instead introduced [9]. In practice, we do not allow some of the IMRPhenomPv2 phase coefficients to deviate from their prescribed values as they have large uncertainties or are degenerate with other coefficients or physical parameters [9]. We therefore perform tests with the remaining 14 dephasing coefficients, henceforth denoted as the testing dephasing coefficients [9]:

\[
\{ \delta p_i \} = \{ \delta \varphi_0, ..., \delta \varphi_4, \delta \varphi_5, \delta \varphi_6, \delta \varphi_7, \delta \beta_2, \delta \beta_3, \delta \alpha_2, \delta \alpha_3, \delta \alpha_4 \}.
\]

The frequency dependence of the testing parameters \( \delta p_i \) is shown in Table I [19, 46].

To quantify a deviation from GR, we can infer the most probable values of \( \delta p_i \) through Bayesian parameter estimation, as discussed in the following subsection.

### A. Parameter Estimation

Recall our data model \( d = h + n \). We denote \( \theta(\theta, \delta p_i) \) as the parameter vector generating the signal \( h \). It consists of the parameters \( \theta \) generating the IMRPhenomPv2 waveform, and the testing parameters \( \delta p_i \) generating the phase deviations from the IMRPhenomPv2 waveform. In practice, the testing parameters are introduced once at
we wish to infer the conditional probability density of $\theta$. Assuming that noise from multiple detectors, indexed $l$, where $(\cdot | \cdot)$ is the noise-weighted inner product [47]:

$$P(d_l | \theta, I) \propto \exp\left[ -\frac{1}{2} (d_l - h_l | d_l - h_l) \right].$$  

(10)

The prior $P(\theta | I)$ incorporates our beliefs about $\theta$ prior to the observation. We follow the default choice of prior in LALInference [24], which include uniform priors for the component masses $m_1$ and $m_2$, with $m_2 \leq m_1$, a log-uniform prior for the luminosity distance, an isotropic prior for the sky location of the source and the spin angular momentum vectors of the component black holes, and uniform priors for the remaining parameters. We note that in LALInference, the uniform priors specified for component masses are transformed to non-uniform, correlated priors for the chirp mass $M \equiv (m_1 m_2)^{3/5} (m_1 + m_2)^{-1/5}$ and the mass ratio $q \equiv m_2/m_1$ for more efficient sampling [24].

In parameterized tests of GR, parameters of primary interest are the testing parameters $\delta p_i$, while the posterior distribution spans the full 16-dimensional parameter space. We therefore compute the marginalized posterior distribution for introduced the testing parameter $\delta p_i$:

$$P(\delta p_i | d, I) = \int P(\theta | d, I) d\theta,$$

(11)

where $\theta$ denotes the parameters generating the underlying IMRPhenomPv2 waveform.

### IV. GLITCH MITIGATION METHODS

In this section, we review four methods that can be applied to mitigate data containing glitches, in which three are used in our investigation, including a frequency-domain filtering method of band-pass filtering, a time-domain filtering method of inpainting, and a glitch model subtraction method using the BayesWave algorithm.

#### A. Band-pass filtering in Frequency Domain

Assuming stationary and Gaussian noise, components of the noise-weighted inner product from different frequency bins of equal bandwidth and from different detectors contribute linearly to the log likelihood, as seen from Eq. (10). A direct way of removing the glitch in frequency domain is by excluding the frequency bins containing the glitch from the likelihood calculation. In LALInference, this can be done by specifying the high-pass and low-pass cutoff frequency for the affected detector such that data containing the glitch is filtered out. Only the passed frequency bins are considered in the likelihood calculation. By default, data is high-passed at 20 Hz in LALInference [24].

#### B. Gating and Inpainting in Time Domain

A similar procedure can be done in the time domain, commonly known as gating, in which data containing the glitch is zeroed out by multiplying an inverse window
function. The inverse window function reduces the spectral leakage in frequency domain due to discontinuity of data at the boundary of the region to be zeroed out [48].

Gating is adopted in the mitigation of glitch-overlapped GW170817 signal in LIGO-Livingston during the rapid localization of the source [49], which successfully led to follow-up electromagnetic observations [50]. However, gating was not used for parameter estimation purposes for the first half of the O3 observing run (O3a) [51]. Some of the concerns and limitations of mitigating glitches by gating are discussed below.

As remarked in Ref. [38], gating can introduce errors to parameterized tests of GR, as it affects the signal power in frequency bins that count towards the noise-weighted inner product.

A new method, called inpainting or hole filling [37], is developed to address the noise artifacts and statistical bias that may be resulted from gating. After specifying the time interval to be mitigated, new values are assigned for data within the interval, or hole, according to an inpainting filter, while data outside the hole are unaffected. The inpainting filter depends on the PSD of the stationary Gaussian noise. Inpainted data within the hole is identically zero upon twice-whitening by the same PSD, and the quantity \( (d|h) \) is independent of the template waveform \( h \) inside the interval [37]. Since the hole can be made arbitrarily narrow, inpainting affects the minimal amount of data if the glitch is localized in time.

Inpainting is not expected to bias parameterized tests of GR: re-expressing the noise-weighed inner product in the likelihood calculation:

\[
P(d|h) \propto \exp \left[ -\frac{1}{2} (d - h|h) - h|h \right].
\]

\[
= \exp \left[ -\frac{1}{2} (d|h) + (d|h) - \frac{1}{2} (h|h) \right].
\]

Given inpainted data \( d \), only the terms \( (d|h) \) and \( (h|h) \) differ across waveform templates \( h \); between these two terms, only \( (d|h) \) explicitly depends on the inpainted data. As \( (d|h) \) is independent of the template waveform inside the hole, inpainted data inside the hole is not expected to contribute to the outcome of parameterized tests.

C. Glitch Model Subtraction

The BayesWave [35, 36] algorithm models the GW signal and glitches in each detector using a variable number of wavelets, such as sine-Gaussian wavelets. Mitigated, or deglitched, data is generated by subtracting the glitch model from the original data. Using Bayesian inference, the output data in each detector is modeled as a superposition of a GW signal \( h \), stationary Gaussian noise \( n_G \) and glitches \( g \):

\[
d = h + n_G + g.
\]

While both the GW signal and glitches are non-stationary and non-Gaussian, coherent features across data from multiple detectors are modeled by the signal model and independent features are modeled by the glitch model [52]. A trans-dimensional Reversible Jump Markov Chain Monte Carlo algorithm is used to sample models with different number of wavelets or with wavelets of different parameter values. The most probable model is inferred through Bayesian inference by comparing the evidence \( P(d|M_i, I) \) for different models \( M_i \); given data \( d \) and prior information \( I \), we define the odds \( O_2^1 \) between two competing models \( M_1 \) and \( M_2 \) as

\[
O_2^1 \equiv \frac{P(M_1|d, I)}{P(M_2|d, I)} = \frac{P(M_1|I)}{P(M_2|I)} \times \frac{P(d|M_1, I)}{P(d|M_2, I)},
\]

where the equality on the right is obtained by invoking Bayes’ Theorem. The model \( M_1 \) will be more probable than model \( M_2 \) if the odds \( O_2^1 \) is larger than 1. To express our ignorance towards the probability of models prior to observation, we can set the first term on the rightmost of Eq. (14), called the prior odds, to unity. The odds can then be obtained by comparing the evidences of the two models. In BayesWave, the evidences are calculated through thermodynamic integration [35]. Once the most probable glitch+signal model is inferred, the glitch model is subtracted from the data.

The BayesWave algorithm was first used to remove the glitch which overlapped with the GW170817 signal during parameter estimation [49], and was regularly used to mitigate glitch-overlapped signals during O3a [18]. Ref. [38] concluded that parameter recovery results using data reconstructed by BayesWave are unbiased. In the context of tests of GR, which are designed to detect small deviations from GR waveforms, the subtraction of sine-Gaussian wavelets by BayesWave may alter the GW signal to an extent which may be reported as a false violation of GR. This is not observed in our results.

V. RESULTS OF GLITCHES OVERLAPPING A GW190828_065509-LIKE SIGNAL

We are motivated to consider a signal similar to that of the high-mass-ratio BH-BH merger event GW190828_065509 [18], in which the mitigation of potential glitches overlapping the event in L1 through bandpass filtering resulted in pathological features in parameterized tests of GR [53]. Values of some selected generating parameters of the GW190828_065509-like signal is tabulated in Table III.

We first present the expected results of parameterized tests of GR in the absence of glitches by coherently injecting the simulated GW190828_065509-like signal, generated with an IMRPhenomPv2 waveform model, into 5 realizations of simulated stationary, Gaussian noise colored with the representative best (cleaned) PSD of the LIGO-Hanford (H1), LIGO-Livingston (L1) and Virgo (V1) detectors during O3a. The posterior distributions
TABLE II. Key specifications of the three mitigation methods.

| Mitigation Method | Specification                  | Blip   | Tomte  | Scattered-Light |
|-------------------|--------------------------------|--------|--------|-----------------|
|                   | High-pass Cutoff (Hz)          | 20     | 105    | 40              |
| Band-pass         | Low-pass Cutoff (Hz)           | 60     | 511.875| 511.875         |
| Inpainting        | Hole Duration (s)              | 0.005  | 0.040  | N/A*            |
|                   | Sampling Rate (Hz)             | 4096   | 4096   | N/A*            |
| Glitch Model      | Segment Length (s)             | 4      | 4      | 8               |
|                   | High-pass Cutoff (Hz)          | 20     | 20     | 8               |
|                   | Sampling Rate (Hz)             | 2048   | 2048   | 2048            |
|                   | $Q_{\text{max}}$               | 40     | 40     | 200             |
|                   | $D_{\text{max}}$               | 100    | 100    | 200             |

*Inpainting is replaced by disregarding data from the detector which the scattered-light glitch is present.

FIG. 3. Posterior distributions of testing parameters (top) and recovered chirp mass (bottom) obtained by performing parameterized tests of GR on 5 data realizations of a simulated GW190828_065509-like signal in stationary Gaussian noise. The simulated noise is colored by the representative best LIGO-Hanford, LIGO-Livingston and Virgo detector PSD estimates during O3a.

TABLE III. Injected values of some selected generating parameters of a GW190828_065509-like signal using the IMRPhenomPv2 waveform model. The GW190828_065509-like signal is taken to be the maximum likelihood waveform inferred for real GW190828_065509 data using the IMRPhenomPv2 template waveform model. Despite the large injected values for the component spins, the inferred posterior distributions of the component spins are flat throughout the prior range for the GW190828_065509 and simulated GW190828_065509-like signals.

| Waveform Parameter | Value    |
|--------------------|----------|
| Chirp mass $M$ ($M_{\odot}$) | 16.86    |
| Mass ratio $q$     | 0.14     |
| Dimensionless primary spin magnitude $a_1$ | 0.92 |
| Dimensionless secondary spin magnitude $a_2$ | 0.75 |
| Right ascension $\alpha$ (rad) | 2.54 |
| Declination $\delta$ (rad) | -0.84 |

of the testing parameters and the recovered chirp mass are plotted in Fig. 3. Since we are injecting a weak signal [54], the GR value of 0 is excluded from several distributions of testing parameters (e.g. $\delta\phi_4$ for the dotted case, $\delta\alpha_2$ for the dashed case) in the 5 considered data realizations.

Beneath the posteriors of testing parameters, the posteriors of recovered chirp mass are also plotted as an indicator of the sampling performance. With an extra degree of freedom introduced by inspiral testing parameters, broad and occasional multimodal distributions can be observed. Such features can also be observed in the posterior distributions for glitch-mitigated data presented in the following subsections.

The same signal is then injected into real H1, L1 and V1 detector data at times where glitches are present in either H1 or L1 with all three detectors operating in the science mode [55]. Glitches from the blip, tomte, scattered-light classes are chosen, as these classes of glitches have the highest occurrence rates in O3a. The glitches used in our study is further chosen so that their duration and peak frequency are representative of their corresponding glitch classes.

The GW190828_065509-like signal is injected coherently into the three detectors such that each glitch overlaps with the signal at the inspiral, intermediate and merger-ringdown stage in time domain for different data samples. The three stages in time domain are defined as the time intervals when the instantaneous frequencies of the signal are in the corresponding three stages in...
frequency-domain discussed in Sec. II B respectively. The boundaries of the three stages of the signal in time and frequency domain are marked in the Q-scans by vertical and horizontal white lines respectively.

After preparing the data samples, we applied the glitch mitigation methods of band-pass filtering, inpainting and BayesWave glitch subtraction as described in Sec. IV on detector data in which glitches are present. We then performed parameterized tests of GR on the unmitigated and mitigated samples. The specifications of the three glitch mitigation methods are tabulated in Table II. We referred to Ref. [56] for BayesWave specifications, whereas band-pass cutoff frequencies and inpainting hole duration are chosen to exclude time-frequency bins affected by the glitch.

A. Blip Glitch

Blip glitches are short-duration, broadband glitches characterized by their teardrop shape as seen in time-frequency representations. A Q-scan of a blip glitch is shown on the top-left of Fig. 1 [26]. The sources and coupling of blip glitches are not well-understood [18]. The simulated GW190828,065509-like signal is coherently injected into H1, L1 and V1 at GPS times around 1253103382.105, corresponding to the time when a blip glitch is present in H1. The injection time is slightly adjusted so that the glitch overlap with the signal at the inspiral, intermediate and merger-ringdown stages in time domain. Parameterized tests of GR are performed on the unmitigated and mitigated data, the posterior of the testing parameters are plotted on the left and right side of each violin plot in Fig. 4 respectively.

The stages of coalescence where violations of GR are observed show no correlation with those overlapped by the glitch in time or frequency domain: violations of GR can be observed for testing parameters from all stages of coalescence when the blip glitch overlap with the signal in intermediate or merger-ringdown stage in time domain, even though the blip glitch contributes excess power only to intermediate and merger-ringdown frequency bands.

No observable effects on parameterized tests of GR are observed when the blip glitch temporally overlapping the signal at inspiral stage, suggested by the matching posterior distributions without and with the glitch removed through independent methods of low-passing to 60 Hz, deglitching and inpainting. Though, the GR value of 0 is only located at the far tails of distributions for the inspiral testing parameters due to the weakness of the injected simulated signal.

Comparing the unmitigated and mitigated results of the glitch overlapping the intermediate and merger-ringdown stage in time domain, all three mitigation methods of low-passing, inpainting and deglitching can reduce false violations of GR by bringing posteriors of testing parameters from distributions that exclude the GR value of 0 to one that peaks close to 0 (e.g. $\delta \varphi_3, \delta \varphi_4, \delta \beta_2, \delta \beta_3, \delta \alpha_2$ in Fig. 4c). The posterior distribution of the testing parameters for mitigated samples matches each other closely, indicating that the mitigation methods did not contribute extra effects on parameterized tests of GR in these three cases. Significant improvements in parameterized tests of GR after removal of the blip glitch suggests that false violations are attributed to the presence of the glitch.

Upon introduction of an extra degree of freedom by the parameterized deviations, the blip glitch leads to bias in other intrinsic variables such as the chirp mass. This is reflected by the sharp peaks in the chirp mass posteriors which exclude the injected value (e.g. $\delta \alpha_2, \delta \alpha_3, \delta \alpha_4$ in Fig. 4b, $\delta \varphi_2, \delta \varphi_3, \delta \varphi_6$ in Fig. 4c). The independent glitch mitigation methods consistently bring the peak values close to the injected value. Although multimodal features can still be noted in several cases when inspiral testing parameters are introduced, the same features are observed for the case of signals injected into simulated Gaussian noise.

B. Tomte Glitch

Tomte glitches are short-duration, broadband glitches characterized by their triangular shape as seen in time-frequency representations. A Q-scan of a tomte glitch is shown on the top-right of Fig. 1. The sources and coupling of tomte glitches are not well-understood. The simulated GW190828,065509-like signal is coherently injected into H1, L1 and V1 at GPS times around 1252901859, corresponding to the time when a tomte glitch is present in L1. The injection time is slightly adjusted so that the glitch overlap with the signal at the inspiral, intermediate and merger-ringdown stages in time domain. Parameterized tests of GR are performed on the unmitigated and mitigated data, the posterior of the testing parameters are plotted on the left and right side of each violin plot in Fig. 5 respectively.

The stages of coalescence where violations of GR are observed show no correlation with those overlapped by the glitch in time or frequency domain: despite shifting the time of overlapping of the signal with the tomte glitch, which contributes considerable excess power also to the intermediate frequency bands, false violations of GR are only observed in inspiral testing parameters for the unmitigated data samples (left of violins). Exclusions of the GR value of 0 are clearly observed in lower PN orders such as $\delta \varphi_2$ and $\delta \varphi_3$.

Comparing the unmitigated and mitigated results, both inpainting and glitch model subtraction can consistently reduce the false violations in the lower PN order testing parameters, resulting in strong support for the GR value of 0 in most testing parameters. Meanwhile, high-passing up to 105 Hz is not a robust glitch mitigation method, as false deviations of GR can be amplified (e.g. $\delta \varphi_2, \delta \varphi_3$ in Fig. 5a) or introduced by the mitigation (e.g. $\delta \beta_2, \delta \beta_3$ in Fig. 5b). Improvements in parameterized
FIG. 4. Top sub-figures: Q-scans of the unmitigated and mitigated data samples. Vertical and horizontal white lines denote the boundaries of different stages of coalescence in time and frequency domain respectively. Bottom sub-figures: Posterior distributions of testing parameters (top) and the recovered chirp mass (bottom) obtained by performing parameterized tests on unmitigated (left of violin plot) blip-glitch-overlapped signals during a three-detector observation. The corresponding mitigated cases (right of violin plot) with band-pass filtering (solid line), BayesWave glitch model subtraction, also called de-glitching (dotted line), and inpainting (dashed line) are also plotted. The GR value of the testing parameters and the injected value of chirp mass are indicated by vertical black lines.
FIG. 5. Similar to Fig. 4, but for blip-glitch-overlapped-signals.
FIG. 6. Similar to Fig. 4, but for scattered-light–glitch-overlapped-signals and with inpainting replaced with using only data from detectors without glitches.
tests of GR after removal of the glitch by inpainting and glitch model subtraction suggest that the false violations in the inspiral parameters are attributed to the presence of the tomte glitch, which contributes significant excess power in inspiral frequency bands.

Again, we can observe bias in the inference of chirp mass when the tomte glitch overlaps the simulated signal, such that the posterior distributions peak far away and exclude the injected value (e.g. $\delta \varphi_3$ in Fig. 5a, $\delta \varphi_2$ in Fig. 5b and $\delta \varphi_4$ in Fig. 5c). The independent glitch mitigation methods consistently bring the peak values close to the injected value, yet multimodal features can be observed, as noted in previous subsections.

C. Scattered-light Glitch

Scattered-light glitches are produced by laser light scattering out and re-entering the main laser beam, and their correlation with seismic motion is well-understood [18]. Scattered-light glitches are characterized by their arch shape as seen in a time-frequency representation such as the bottom $Q$-scan in Fig. 1 [26]. The simulated GW190828_065509-like signal is coherently injected into H1, L1 and V1 at GPS times around 1253416025, corresponding to the time when scattered-light glitches are present in H1. The injection time is slightly adjusted so that the glitch predominantly overlap with the inspiral, inspiral-intermediate-merger-ringdown and merger-ringdown stages of the signal in time domain respectively. Parameterized tests of GR are performed on the unmitigated and mitigated data, the posterior of the testing parameters are plotted on the left and right side of each violin plot in Fig. 6 respectively.

Since the typical timescale of scattered-light glitches (> 1 s) are large compared to that of BBH-coalescence GW signals (in the LIGO band), it is likely for an overlapping scattered-light glitch to overlap with the entire GW signal. As such, we choose not to apply time-domain filtering methods, such as inpainting, to the study of scattered-light-glitch-overlapped signals. To compare the results of the other mitigation methods, we replace inpainting with using only data only from the remaining two detectors which glitches are not present. However, such a method is not preferred in general as much useful information is lost, and is omitted from later discussions of suitable mitigation methods.

For all three cases of glitch-overlapping, the value of 0 is not excluded from the posterior distributions of testing parameters for the unmitigated samples (left of violin plots). Furthermore, the glitch has no observable effect on parameterized tests of GR when it temporally overlaps with the inspiral stage of the signal (Fig. 6a), as the posterior distributions of all testing parameters for the unmitigated case matches with that with the glitch removed in three independent methods.

In most cases, the posterior distributions of testing parameters obtained by the three independent glitch mitigation methods match closely with each other. This is a good indication that the mitigation methods do not introduce additional effects to the results. In particular, in the case where the scattered-light glitch overlaps all stages of coalescence in time domain (Fig. 6b), the glitch removal through high-passing to 40 Hz improves the inference of inspiral testing parameters most significantly (see $\delta \varphi_4$, $\delta \varphi_{61}$, $\delta \varphi_{66}$, $\delta \varphi_{72}$).

VI. CONCLUSION AND OUTLOOK

We overlapped a simulated high-mass-ratio coalescing BBH signal with three glitches from most frequently occurring glitch classes in O3a. We then investigated the effects on parameterized tests of GR of the glitches and their mitigations through band-pass filtering, inpainting and BayesWave glitch model subtraction. Although the number of glitches considered in this investigation is not sufficient for us to give quantitative statements about the effects of certain glitch classes or mitigation methods on tests of GR, our analysis covered all stages of BBH coalescence in time and frequency domain, and we are able to identify the effects case by case by comparing the unmitigated results with that mitigated by independent methods and expected GR results when the noise model is not violated.

No false violations of GR are identified for data samples overlapped with the scattered-light glitch, while false violations are observed for that overlapped with the tomte and blip glitches. For the latter cases, we found no clear correlation between the stages of coalescence which false violations occurred and those overlapped by the glitch in time or frequency domain.

Out of the three mitigation methods, we find that inpainting and BayesWave glitch model subtraction consistently reduces false violations of GR, and the results match closely with each other. This indicates that the two methods did not introduce additional effects to parameterized tests, and suggests successful glitch removals. Band-pass filtering, on the other hand, can also reduce false violations in most cases. However, false violations are amplified or new violations are introduced in more than one case after high-passing to 105 Hz. We suggest the application of inpainting or BayesWave glitch model subtraction for glitch mitigation, as they are found to be effective even when an extra degree of freedom is involved with the introduction of parameterized deviation to the signal model.

A major improvement on the LIGO detectors are expected to be completed in a few years, doubling the detector sensitivities [57]. The increased sensitivity in turn suggests more frequent occurrence of glitches overlapping signals. As mitigating signals overlapped with glitches may become a regularity in the future, a systematic study on the effects of glitches and their mitigations to tests of GR, consisting of repeated applications of similar methodologies to that presented by this work.
to different classes of glitches, will be crucial to the next generation of GW astronomy, making this work an important first study on this subject.

VII. ACKNOWLEDGMENTS

JK thanks Derek Davis and Jonah Kanner for insightful discussions. We thank the National Science Foundation (NSF) and NSF REU for supporting the LIGO SURF program. The LIGO SURF Program is supported by NSF award PHY-1852081. Computing resources for this study was provided by the LIGO Laboratory and supported by the NSF Grants PHY-0757058 and PHY-0823459. The work described in this paper is partially supported by grants from the Research Grants Council of the Hong Kong (Project No. CUHK 24304317), The Croucher Foundation of Hong Kong and Research Committee of the Chinese University of Hong Kong. This research has made use of data, software and/or web tools obtained from the Gravitational Wave Open Science Center (https://www.gw-openscience.org/), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. LIGO Laboratory and Advanced LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. This paper carries LIGO Document Number LIGO-P2100294.
[17] B. Abbott, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt, et al., Gwte-1: a gravitational-wave transient catalog of compact binary mergers observed by ligo and virgo during the first and second observing runs, Physical Review X 9, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].

[18] R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. Adhikari, V. Adya, C. Affeldt, et al., Gwte-2: Compact binary coalescences observed by ligo and virgo during the first half of the third observing run, Phys. Rev. X 11, 021053 (2021), arXiv:2010.14527 [gr-qc].

[19] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al., Tests of General Relativity with GW150914, Phys. Rev. Lett. 116, 221101 (2016), arXiv:1602.03841 [gr-qc].

[20] B. Abbott, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. Adhikari, V. Adya, C. Affeldt, et al., Tests of General relativity with the binary black hole signals from the ligo-virgo catalog gwtc-1, Physical Review D 100, 104036 (2019), arXiv:1903.04467 [gr-qc].

[21] R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. Adhikari, V. Adya, C. Affeldt, et al., Tests of General relativity with binary black holes from the second LIGO-Virgo Gravitational-Wave Transient Catalog, Phys. Rev. D 103, 122002 (2021), arXiv:2010.14529 [gr-qc].

[22] P. R. Saulson, Fundamentals of Interferometric Gravitational Wave Detectors (2017).

[23] LIGO Scientific Collaboration, LIGO Algorithm Library - LALSuite, free software (GPL) (2018).

[24] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Ayloff, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C.-J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O’Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. Van Der Sluys, A. Vecchio, W. Vousden, and L. Wade, Parameter estimation for compact binaries with ground-based gravitational-wave observations using the lalinferece software library, Phys. Rev. D 91, 042003 (2015), arXiv:1409.7215 [gr-qc].

[25] L. K. Nuttal, T. Massinger, J. Areeda, J. Betzwieser, S. Dwyer, A. Effler, R. Fisher, P. Fritschel, J. Kissel, A. Lundgren, et al., Improving the data quality of advanced ligo based on early engineering run results, Classical and Quantum Gravity 32, 245005 (2015), arXiv:1508.07316 [gr-qc].

[26] M. Zevin, S. Coughlin, S. Bahaadini, E. Besler, N. Rohani, S. Allen, M. Cabero, K. Crowston, A. K. Katsaggelos, S. L. Larson, et al., Gravity spy: integrating advanced ligo detector characterization, machine learning, and citizen science, Classical and Quantum Gravity 34, 064003 (2017), arXiv:1611.04596 [gr-qc].

[27] B. P. Abbott, R. Abbott, T. Abbott, M. Abernathy, F. Acernese, K. Ackley, M. Adamo, C. Adams, T. Adams, P. Addesso, et al., Characterization of transient noise in advanced ligo relevant to gravitational wave signal gw150914, Classical and Quantum Gravity 33, 134001 (2016), arXiv:1602.03844 [gr-qc].

[28] J. R. Smith, T. Abbott, E. Hirose, N. Leroy, D. MacLeod, J. McIver, P. Saulson, and P. Shawhan, A hierarchical method for vetoing noise transients in gravitational-wave detectors, Classical and Quantum Gravity 28, 235005 (2011), arXiv:1107.2948 [gr-qc].

[29] T. Isogai, L. S. Collaboration, V. Collaboration, et al., in Journal of Physics: Conference Series, Vol. 243 (IOP Publishing, 2010) p. 012005.

[30] R. Essick, L. Blackburn, and E. Katsavounidis, Optimization of vetoes for gravitational-wave transient searches, Classical and Quantum Gravity 30, 155010 (2013), arXiv:1303.7159 [astro-ph.IM].

[31] R. Biswas, L. Blackburn, J. Cao, R. Essick, K. A. Hodge, E. Katsavounidis, K. Kim, Y.-M. Kim, E.-O. Le Bigot, C.-H. Lee, J. J. Oh, S. H. Oh, E. J. Son, Y. Tao, R. Vaulin, and X. Wang, Application of machine learning algorithms to the study of noise artifacts in gravitational-wave data, Phys. Rev. D 88, 062003 (2013), arXiv:1303.6984 [astro-ph.IM].

[32] F. Robinet, N. Arnaud, N. Leroy, A. Lundgren, D. Macleod, and J. McIver, Omicron: a tool to characterize transient noise in gravitational-wave detectors, arXiv preprint arXiv:2007.11374 (2020), arXiv:2007.11374 [astro-ph.IM].

[33] C. Messick, K. Blackburn, P. Brady, P. Brockill, K. Cannon, R. Cario, S. Caudill, S. J. Chamberlin, J. D. E. Creighton, R. Everett, C. Hanna, D. Keppel, R. N. Lang, T. G. F. Li, D. Meacher, A. Nielsen, C. Pankow, S. Privitera, H. Qi, S. Sachdev, L. Sadeghian, L. Singer, E. G. Thomas, L. Wade, M. Wade, A. Weinstein, and K. Wiesner, Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data, Phys. Rev. D 95, 042001 (2017), arXiv:1604.04324 [astro-ph.IM].

[34] S. A. Usman, A. H. Nitz, I. W. Harry, C. M. Biwer, D. A. Brown, M. Cabero, C. D. Capano, T. Dal Canton, T. Dent, S. Fairhurst, et al., The pyche search for gravitational waves from compact binary coalescence, Classical and Quantum Gravity 33, 215004 (2016), arXiv:1508.02357 [gr-qc].

[35] N. J. Cornish and T. B. Littenberg, Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches, Classical and Quantum Gravity 32, 135012 (2015), arXiv:1410.3835 [gr-qc].

[36] T. B. Littenberg and N. J. Cornish, Bayesian inference for spectral estimation of gravitational wave detector noise, Phys. Rev. D 91, 084034 (2015), arXiv:1410.3852 [gr-qc].

[37] B. Zackay, T. Venumadhav, J. Roulet, L. Dai, and M. Zaldarriaga, Detecting gravitational waves in data with non-gaussian noise, arXiv preprint (2019), arXiv:1908.05644 [astro-ph.IM].

[38] C. Pankow, K. Chatzioannou, E. A. Chase, T. B. Littenberg, M. Evans, J. McIver, N. J. Cornish, C.-J. Haster, J. Kanner, V. Raymond, S. Vitale, and A. Zimmerman, Mitigation of the instrumental noise transient in gravitational-wave data surrounding gw170817, Phys. Rev. D 98, 084016 (2018), arXiv:1808.03619 [gr-qc].

[39] S. Chatterji, L. Blackburn, G. Martin, and E. Katsavounidis, Multiresolution techniques for the detection of gravitational-wave bursts, Classical and Quantum Gravity 21, S1809 (2004), arXiv:gr-qc/0412119 [gr-qc].

[40] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Simple model of complete precessing black-hole-binary gravitational waveforms, Physical review letters 113, 151101 (2014).
[41] W. B. Davenport, W. L. Root, et al., *An introduction to the theory of random signals and noise*, Vol. 159 (McGraw-Hill New York, 1958).

[42] J. D. Romano and N. J. Cornish, *Detection methods for stochastic gravitational-wave backgrounds: a unified treatment*, Living reviews in relativity 20, 2 (2017), arXiv:1608.06889 [gr-qc].

[43] P. Whittle, *Hypothesis testing in time series analysis*, Vol. 4 (Almqvist & Wiksells boktr., 1951).

[44] S. Khan, K. Chatziioannou, M. Hannam, and F. Ohme, *Phenomenological model for the gravitational-wave signal from precessing binary black holes with two-spin effects*, Physical Review D 100, 024059 (2019), arXiv:1809.10113 [gr-qc].

[45] S. Husa, S. Khan, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, *Frequency-domain gravitational waves from nonprecessing black-hole binaries. ii. a phenomenological model for the advanced detector era*, Physical Review D 93, 044007 (2016), arXiv:1508.07253 [gr-qc].

[46] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, *Frequency-domain gravitational waves from nonprecessing black-hole binaries. i. new numerical waveforms and anatomy of the signal*, Physical Review D 93, 044006 (2016), arXiv:1508.07250 [gr-qc].

[47] C. Cutler and E. E. Flanagan, *Gravitational waves from merging compact binaries: How accurately can one extract the binary’s parameters from the inspiral waveform?*, Physical Review D 49, 2658 (1994), arXiv:gr-qc/9402014 [gr-qc].

[48] F. J. Harris, *On the use of windows for harmonic analysis with the discrete fourier transform*, Proceedings of the IEEE 66, 51 (1978).

[49] B. P. Abbott, R. Abbott, T. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari, V. Adya, et al., *Gw170817: observation of gravitational waves from a binary neutron star inspiral*, Physical Review Letters 119, 161101 (2017), arXiv:1710.05832 [gr-qc].

[50] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, et al., *Multi-messenger observations of a binary neutron star merger*, The Astrophysical Journal 848, L12 (2017), arXiv:1710.05833 [astro-ph.HE].

[51] J. McIver, M. Millhouse, G. Ashton, D. Davis, J. Veitch, and T. Littenberg, *O3a catalog mitigated data review results and status spreadsheet*, LIGO-T2000260.

[52] N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, *BayesWave analysis pipeline in the era of gravitational wave observations*, Phys. Rev. D 103, 044006 (2021), arXiv:2011.09494 [gr-qc].

[53] Private communication with Rico K. L. Lo.

[54] The event GW190228_065509 has a full-signal optimal SNR of 9.9 - weakest in GWTC-1 and GWTC-2 combined [20, 21].

[55] Calibrated, cleaned data from H1 and L1 are taken from the strain channel DCS-CALIB STRAIN_CLEAN C01. Reproduced data from V1 is taken from the strain channel Hrec_hof_V1O3Repro1A_16384Hz [58].

[56] K. Chatziioannou, N. Cornish, M. Wijngaarden, and T. B. Littenberg, *Modeling compact binary signals and instrumental glitches in gravitational wave data*, Phys. Rev. D 103, 044013 (2021), arXiv:2101.01200 [gr-qc].

[57] J. Miller, L. Barsotti, S. Vitale, P. Fritschel, M. Evans, and D. Sigg, *Prospects for doubling the range of advanced ligo*, Physical Review D 91, 062005 (2015), arXiv:1410.5882 [gr-qc].

[58] R. Abbott et al., *Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo*, SoftwareX 13, 100658 (2021), arXiv:1912.11716 [gr-qc].