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

Jack Y. L. Kwok*
Department of Physics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Mentors: Alan J. Weinstein, Rico K. L. Lo
LIGO, California Institute of Technology, Pasadena, California 91125, USA
(Dated: August 3, 2020)

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 false deviations from GR obtained by performing parameterized gravitational-wave tests on simulated signals from binary-black-hole coalescence overlapped with instrumental glitches. We then separately apply three common glitch mitigation methods and evaluate their effect on reducing false deviations from GR.

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 to probe 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].

Since 2015, Advanced LIGO [7] and Advanced Virgo [8] have jointly announced 14 confident detections of GWs, all of which are generated by the *coalescence* of compact binaries [9–12]. The coalescence of BBHs begins as their orbital separation continuously decreases due to emission of GWs during the *inspiral* phase, until the point when the black holes are so close to each other they plunge together close to the speed of light and *merge* into a single black hole, which quickly settles down to a Kerr black hole during the *ringdown* phase [13, 14].

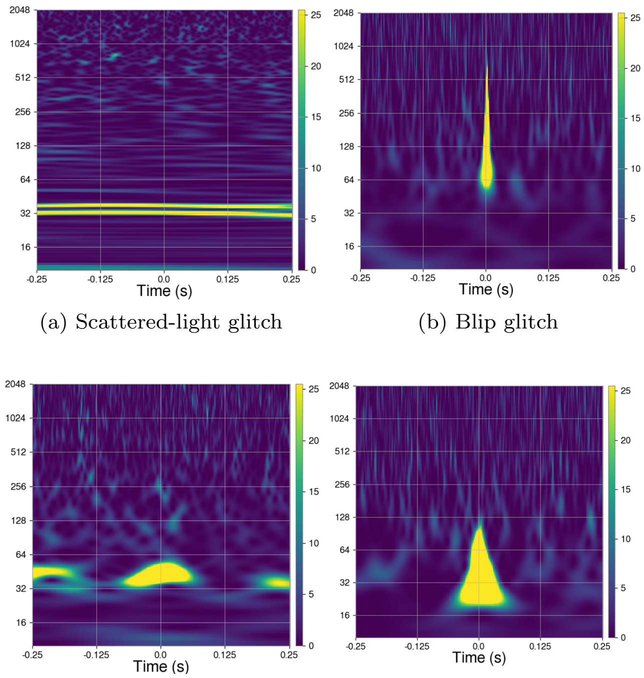
Of all strong-field astrophysical events that could be probed, the coalescence of stellar-mass binary black holes (BBHs) plays a crucial role in testing GR [1]. Since the orbital separation of the BBH can reach far below the last stable orbit before merging, the gravitational field generated can reach many order of magnitudes larger than other observed astrophysical events [14–16]. Moreover, GWs emitted by coalescing BBHs offers one of the cleanest test of GR, as environmental effects such as accretion disks and electromagnetic fields are negligible for most sources [17], enabling precision tests of the strong-field dynamics of GR.

Several *generic* tests of GR using coalescing BBHs are developed: consistency tests search for excess power after subtracting a best-fit GR waveform from the data [18], or compare the source parameters inferred using only high-frequency data to that inferred using only low-frequency data [18]; parameterized tests introduce parameterized deformations to waveform approximations to GR, which is in turn inferred using Bayesian parameter estimation [16]. To this date, no evidence for violations of GR has been identified using GWs emitted by coalescing BBHs [19].

Aside from GWs, a GW detector output can be attributed to many independent sources of random noise [20]. In light of the central limit theorem, and by assuming that noise characteristics remain stationary over timescales of observing GW signals from BBH coalescence, noise in GW detectors are typically modeled to be stationary and Gaussian in tests of GR [21, 22]. However, these assumptions cannot account for transient, non-Gaussian noise features which enter GW detectors, commonly referred to as *glitches* [23–25]. Four classes of commonly-seen glitches in the LIGO detectors during the O3 observing run are shown in Figure 1. If the presence of glitches were not accounted for, one may infer from the detected waveform that a deviation from GR has occurred. The extent to which glitches mimic the effects of a deviation of GR and the effecst of glitch mitigations to tests of GR deserve an investigation.

This report is structured as follows: In Section II we introduce the typical data model used in GW data analyses [21, 22], which composes of a GW strain component and a stationary Gaussian noise component. In Section III we discuss the phase parameterization of an inspiral-merger-ringdown waveform model [26] and its connection with tests of GR. In Section IV we introduce three commonly-used glitch mitigation measures. In Section

* Email: jackkwok@link.cuhk.edu.hk



(c) Fast-scattering glitch (d) Tomte glitch

FIG. 1. Glitches with similar morphology are categorized into different classes. Four spectrogram representations (Q-scans) of commonly-seen classes of glitches in LIGO Hanford and Livingston detector during the O3 observing run are plotted. The colour represents the normalized energy of the signal at each time-frequency bin [24].

V, we describe our methods of preparing data samples with glitches overlapping GW signals, applying mitigation measures and performing parameterized tests of GR.

II. DATA MODEL

A GW detector is designed to respond linearly to the fractional change in arm length, or *strain* [20]. The time series of detector output data $\boldsymbol{d}$, sampled at time t_k at constant sampling interval Δt, can thus be expressed as a linear superposition of a time series of the GW strain signal $\boldsymbol{h}$ and a time series of detector noise $\boldsymbol{n}$:

$$\boldsymbol{d}(t_k) = \boldsymbol{h}(t_k) + \boldsymbol{n}(t_k)\ . \tag{1}$$

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

A. Gaussian Noise Model

Assume that a *large* number of independent noise sources contributes linearly to the detector noise $\boldsymbol{n}$. Under these assumptions, the central limit theorem states that the probability density distribution of attaining an output value of $n(t_0)$ at an arbitrary time t_0 in the absence of signal tends to be *Gaussian* [27]:

$$P(n(t_0)) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{(n-\mu)^2/2\sigma^2}\ , \tag{2}$$

which is uniquely characterized by the *mean* μ and the *variance* σ^2 at t_0, defined as the ensemble average $E[n(t_0)]$ and $E[(n(t_0)-\mu)^2]$ respectively. The *joint* probability density for N samples of noise collectively attaining values of $n(t_0), n(t_1), ..., n(t_{N-1})$ is given by the *multivariate* Gaussian distribution:

$$P(\boldsymbol{n}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\boldsymbol{n}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{n}-\boldsymbol{\mu})}\ , \tag{3}$$

where $\Sigma_{ij} = E[(n(t_i) - \mu(t_i))(n(t_j) - \mu(t_j))]$ is the *covariance matrix* and $|\boldsymbol{\Sigma}|$ denotes its determinant. The off-diagonal terms of the covariance matrix are measures of the correlations between data from different instances of time.

In addition, the joint probability density distribution is assumed to be time-invariant, which is a reasonably good approximation for Gaussian noise over timescales of observing GW signals from coalescing BBHs [6]. Noise satisfying this assumption is said to be *stationary*. Without loss of generality, we will henceforth set $\boldsymbol{\mu} = \boldsymbol{0}$. For stationary noise, the correlation between data sampled at time t_i and t_j only depend on the time lag $\tau \equiv |t_i - t_j|$. We define the *auto-correlation* $R(\tau)$ as

$$\Sigma_{ij} = E[x(t_i)x(t_j)] = \langle x(t)x(t+\tau)\rangle \equiv R(\tau)\ , \tag{4}$$

where $\langle\cdot\rangle$ denotes the time average over many samples.

If the number of samples N is large, it is undesirable to invert the $N \times N$ covariance matrix in Eq. (3). Instead, we consider the joint probability density in Fourier domain, which is a multivariate Gaussian distribution [27] with a covariance matrix which tends to be *diagonalized* as the discrete time series approach the continuum limit [28]. For even N, we define the *one-sided power spectral density* (PSD) from the real discrete Fourier transform (DFT) of the auto-correlation $R(\tau)$:

$$S_{nj} \equiv \Delta t\ \mathrm{DFT}[R(\tau)] = 2\Delta t \sum_{k=0}^{N-1} R(\tau_k) e^{-i2\pi jk/N}\ , \tag{5}$$

where $j = 0, 1, ..., N/2 - 1$ and the frequencies $f_j \equiv j/N\Delta t$ are sampled from 0 up to the *Nyquist frequency* $1/2\Delta t$. Inverting Eq. (5) for the zero-lag case, we get

$$\sum_{j=0}^{N/2-1} S_{nj}\Delta f = R(0) = \langle n^2(t)\rangle\ , \tag{6}$$

where $\Delta f \equiv 1/T$ is the *frequency resolution*. Summing the PSD over frequency bins as in Eq. (6) returns the *power* of the time series. In terms of the PSD, the joint probability density in Fourier domain is approximately [28]

$$P(\tilde{\boldsymbol{n}}) \simeq \prod_{j=0}^{N/2-1} \frac{2\Delta f}{\pi S_{nj}} \exp\left(-\Delta f \frac{2|\tilde{n}_j|^2}{S_{nj}}\right) , \quad (7)$$

where the frequency series $\tilde{\boldsymbol{n}}$ is similarly defined as

$$\tilde{n}_j \equiv \Delta t \ \mathrm{DFT}[n_k] = 2\Delta t \sum_{k=0}^{N-1} n_k e^{-i2\pi jk/N} . \quad (8)$$

Eq. (7) is also known as the *Whittle likelihood* [29] in the context of statistical inference.

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 inspiral-merger-ringdown waveform model `IMRPhenomPv2` [26] in virtue of its good match with Numerical Relativity (NR) waveforms [30] and low computational costs.

`IMRPhenomPv2` is a phenomenological waveform model constructed by combining PN-like inspiral waveforms with NR merger-ringdown waveforms [31]. Its *inspiral* stage is modeled up to $f \sim 0.018/M$ in natural units, where M is the total mass of the system. The region with $Mf \geq 0.018$ is subdivided into an *intermediate* stage with $0.018 \geq Mf \geq 0.5 f_{\mathrm{RD}}$, which bridges the inspiral stage to the *merger-ringdown* stage modeled above half the ringdown frequency f_{RD} [31]. 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* p_i, are the subjects of parameterized tests of GR in Section III. The phase coefficients p_i can be categorized into three groups, depending on the stages of coalescence in which they predominantly assert their effect on [16, 31]: (i) the *inspiral* PN coefficients $\{\varphi_0, ..., \varphi_5, \varphi_{5l}, \varphi_6, \varphi_{6l}, \varphi_7\}$ and phenomenological coefficients $\{\sigma_0, ..., \sigma_4\}$; (ii) the *intermediate* phenomenological coefficients $\{\beta_0, ..., \beta_3\}$; (iii) the *merger-ringdown* phenomenological and black hole perturbation theory coefficients $\{\alpha_0, ..., \alpha_5\}$.

The phase coefficients p_i depends only on the masses and spin angular momentum vectors of the component black holes [30], denoted as the *intrinsic* parameters. To determine the response of an Earth-based detector, we need to further specify the *extrinsic* parameters, including the sky location and distance, polarization angle, the spatial orientation of the BBH system with respect to the Earth at a reference frequency, and the orbital phase of the system at an arbitrary time.

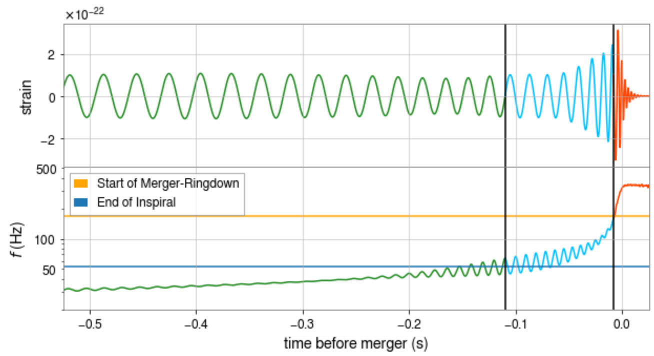


FIG. 2. An example GW strain (upper figure) generated with the `IMRPhenomPv2` approximant and the corresponding instantaneous frequency (lower figure) is plotted against time. The two horizontal lines in the lower figure correspond to the frequencies $Mf = 0.018$ (blue line) and $f_{\mathrm{RD}}/2$ (orange line), which defines the boundaries of the inspiral (green curve), intermediate (light blue curve) and merger-ringdown (dark-orange curve) stages of coalescence for `IMRPhenomPv2`. This figure is reproduced from Fig. 1 of Ref. [16].

III. PARAMETERIZED TESTS OF GR

In this project, we will focus on a parameterized test of GR, which introduces *fractional* deviations δp_i, also known as *de-phasing coefficients*, to `IMRPhenomPv2` phase coefficients p_i [16]:

$$p_i \mapsto p_i[1 + \delta p_i] . \quad (9)$$

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 with other coefficients or physical parameters [16]. We therefore perform tests with the remaining 13 dephasing coefficients, henceforth denoted as the *testing* dephasing coefficients [16]:

$$\{\delta p_i\} = \{\delta\varphi_0, ..., \delta\varphi_4, \delta\varphi_{5l}, \delta\varphi_6, \delta\varphi_{6l}, \delta\varphi_7, \\ \delta\beta_2, \delta\beta_3, \delta\alpha_2, \delta\alpha_3, \delta\alpha_4\} .$$

The frequency dependence of the testing dephasing coefficients δp_i is shown in Table I [18, 32].

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

A. Parameter Estimation

Recall our data model $\boldsymbol{d} = \boldsymbol{h} + \boldsymbol{n}$. Introducing parameterized phase deviations to the signal $\boldsymbol{h}$, we denote $\boldsymbol{\theta}$ as the set of parameters generating the signal, which includes the testing dephasing coefficients δp_i in addition to the intrinsic and extrinsic parameters discussed in Sec II B. In practice, the dephasing coefficients are introduced

Stage of coalescence	δp_i	f-dependence
Inspiral	$\delta\varphi_0$	$f^{-5/3}$
	$\delta\varphi_1$	$f^{-4/3}$
	$\delta\varphi_2$	f^{-1}
	$\delta\varphi_3$	$f^{-2/3}$
	$\delta\varphi_4$	$f^{-1/3}$
	$\delta\varphi_{5l}$	$\log(f)$
	$\delta\varphi_6$	$f^{1/3}$
	$\delta\varphi_{6l}$	$f^{1/3}\log(f)$
	$\delta\varphi_7$	$f^{2/3}$
Intermediate	$\delta\beta_2$	$\log f$
	$\delta\beta_3$	f^{-3}
Merger-Ringdown	$\delta\alpha_2$	f^{-1}
	$\delta\alpha_3$	$f^{3/4}$
	$\delta\alpha_4$	$\tan^{-1}(af+b)$

TABLE I. The frequency dependence of `IMRPhenomPv2` dephasing coefficients used in parameterized tests of GR. The table is reproduced from Table 1 of Ref. [18].

once at a time [18]. A total of 15 parameter estimation runs are thus performed on each data segment.

Given the detector output $\boldsymbol{d}$ and prior information I, we wish to infer the conditional probability density of $\boldsymbol{\theta}$, referred to as the *posterior*, by invoking Bayes' theorem

$$P(\boldsymbol{\theta}|\boldsymbol{d},I) = \frac{P(\boldsymbol{d}|\boldsymbol{\theta},I)\times P(\boldsymbol{\theta}|I)}{P(\boldsymbol{d}|I)}\ , \tag{10}$$

which relates the posterior to three probability densities: the *likelihood* $P(\boldsymbol{d}|\boldsymbol{\theta},I)$, the *prior* $P(\boldsymbol{\theta}|I)$ and the *evidence* $P(\boldsymbol{d}|I)$. During parameter estimation, the evidence, which do not depend explicitly on $\boldsymbol{\theta}$, can be seen as a proportionality constant since $\boldsymbol{d}$ and I are kept fixed. The likelihood and prior is separately discussed below.

Given $\boldsymbol{h}(\boldsymbol{\theta})$, the time series of the output data $\boldsymbol{d}$ uniquely defines a time series of the residual noise $\boldsymbol{d}-\boldsymbol{h}$, which is assumed to be Gaussian and stationary. As such, the likelihood is approximated by the Whittle likelihood in Eq. (7), written in logarithmic form:

$$\log P(\boldsymbol{d}|\boldsymbol{\theta},I) = \sum_{j=0}^{N/2-1}\log\left(\frac{2\Delta f}{\pi S_{nj}}\right) - \frac{1}{2}(\boldsymbol{d}-\boldsymbol{h}|\boldsymbol{d}-\boldsymbol{h})\ , \tag{11}$$

where $(\cdot|\cdot)$ is the *noise-weighted inner product* [33]:

$$(\mathbf{a}|\mathbf{b}) \equiv \sum_{j=0}^{N/2-1} 4\Re\left(\frac{\tilde{a}_j^*\tilde{b}_j}{S_{nj}}\right)\Delta f\ . \tag{12}$$

In practice, the PSD S_{nj} of the data segment of interest is typically estimated using adjacent data segments [22]. The first term on the right side of Eq. (11) do not depend on $\boldsymbol{h}$ thus could be seen as a proportionality constant. Assuming that noise from multiple detectors, indexed l, are uncorrelated, the joint likelihood takes the form

$$P(\boldsymbol{d}_l|\boldsymbol{\theta},I) \propto -\frac{1}{2}\sum_l(\boldsymbol{d}_l-\boldsymbol{h}_l|\boldsymbol{d}_l-\boldsymbol{h}_l)\ . \tag{13}$$

The prior $P(\boldsymbol{\theta})|I)$ incorporates our beliefs about $\boldsymbol{\theta}$ prior to the observation. We follow the default choice of prior in `LALInference` [22], which include uniform priors for the component masses m_1 and m_2, with $m_2 < m_1$, 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 rest of the parameters. In `LALInference`, the uniform priors specified for component masses are transformed to non-uniform, correlated priors for the chirp mass $\mathcal{M} \equiv (m_1 m_2)^{3/5}(m_1+m_2)^{-1/5}$ and the mass ratio $q \equiv m_2/m_1$ for more efficient sampling [22].

In parameterized tests of GR, parameters of primary interest are the testing dephasing coefficients δp_i, while the posterior distribution spans the full parameter space. We therefore compute the *marginalized* posterior distribution for the testing dephasing coefficient δp_i which we introduced into the waveform:

$$P(\delta p_i|\boldsymbol{d},I) = \int P(\boldsymbol{\theta}|\boldsymbol{d},I)d\theta_{\rm int}d\theta_{\rm ext}\ , \tag{14}$$

where $\theta_{\rm int}$ and $\theta_{\rm ext}$ denotes the intrinsic and extrinsic parameters which generates the underlying `IMRPhenomPv2` waveform respectively.

IV. GLITCHES AND THEIR MITIGATION

Many efforts are made to develop algorithms that identify glitches [34–37], which play an important role in gravitational-wave searches. Once a glitch is identified, the data around the glitch could be zeroed out either automatically by search pipelines [38, 39] or manually by multiplying an *inverse* window function [38, 39]. This process, known as *gating*, is illustrated in Figure 3.

A similar procedure can be done in the frequency domain: if the glitch is localized in certain intervals of frequency, zeroing out the corresponding frequency bins through band-pass filtering would eliminate the glitch. In `LALInference`, data is high-passed at 20 Hz by default [22], which can be specified to a higher value to high-pass the frequency bins affected by the glitch.

A more sophisticated approach introduced by `BayesWave` [40, 41] infers the most probable glitch model, constructed using a variable number of sine-Gaussian wavelets, using Bayesian inference. This glitch model is then subtracted from the data. This procedure, known as *de-glitching*, was employed for the glitch-contaminated GW170817 data [42] as illustrated in Figure 3.

In our study, we will separately apply the three mitigation measures of 1) gating, 2) band-pass filtering and 3) de-glitching to data samples.

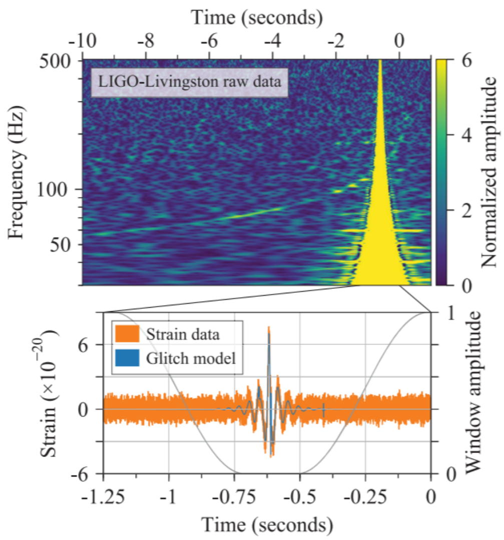


FIG. 3. The output data from the LIGO-Livingston detector during GW170817 is plotted over time in the bottom figure (orange curve). A glitch was identified around the time $t = -0.75$ s to -0.5 s in the figure. To infer the sky location of the event during rapid sky localization, data was multiplied by an inverse Tukey window function (black curve) [42]. To infer the source properties during parameter estimation, a glitch model (blue curve) reconstructed with `BayesWave` is subtracted from the data [42]. The upper figure shows a spectrogram of the raw LIGO-Livingston data. The figure is retrieved from Abbott et al. [42]

V. METHODOLOGY

Our goals are to investigate the extent to which glitches mimic the effects of a deviation of GR in parameterized tests of GR, and evaluate the effect of common glitch mitigation methods on reducing false deviations from GR. To this end, we first prepare data samples by injecting simulated `IMRPhenomPv2` signals coherently into Hanford (H1), Livingston (L1) and Virgo (V1) detector segments where glitches are present. We then perform gating, band-pass filtering, and de-glitching as outlined in Sec IV on the data samples. Lastly, we perform parameter estimation on the mitigated and unmitigated data samples using `LALInference`, where the dephasing coefficients δp_i are allowed to vary one at a time.

A. Preparing Data Samples

On the one hand, the simulated signals in all data samples are all chosen to be the maximum likelihood `IMRPhenomPv2` waveform for the GW event S190828l, which is a BBH merger with total mass of $\sim 44 M_{\odot}$ and a low mass ratio of ~ 0.4. A blop of excess power in L1 overlapped the inspiral stage of the GW signal as shown in Fig 4. The mitigation of the excess power through band-pass filtering lead to pathological features in parameterized tests of GR [43], motivating us to reproduce the situation in our study.

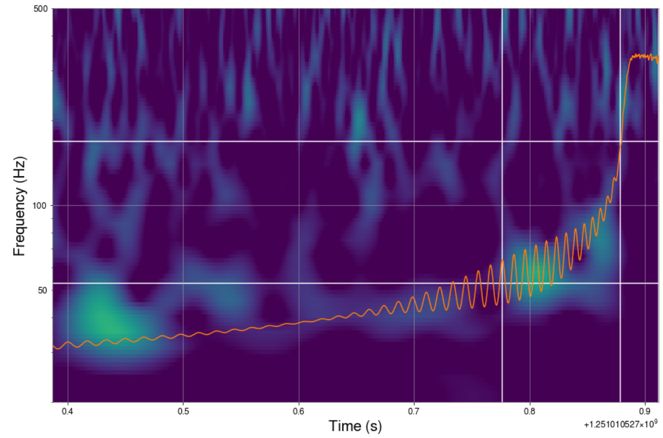


FIG. 4. A Q-scan of the whitened output L1 data for the event S190828l. The white grid lines mark the boundaries for the inspiral, intermediate and merger-ringdown stage of coalescence in time (left to right) and frequency (bottom to top). A blob of excess power can be seen overlapping the inspiral stage. The instantaneous frequency of the simulated S190828l-like signal is plotted on top of the Q-scan (red curve).

On the other hand, one scattered-light glitch, shown in Fig. 1a, and one tomte glitch, shown in Fig. 1d, are chosen to overlap the signal. As seen from Fig. 5, scattered-light glitches (blue) and tomte glitches (red) have relatively high rates of occurrence throughout the O3 observing run. The two classes of glitches differs greatly in morphology: tomte glitches have short duration with a median of 0.625 s and typically affect the data at ~ 20 Hz to ~ 130 Hz, spanning the inspiral and intermediate stages in frequency. Whereas scattered-light glitches have a longer duration with a median of 1.75 s, and a large population of H1 scattered-light glitches are localized at a frequency range of ~ 30 Hz, intersecting the signal track during the early inspiral. The duration of four classes of commonly-seen glitches are retrieved from the search pipeline *Gravity Spy* [24] and plotted in Fig. 5.

The S190828l-like signal is generated and injected into output data across multiple detectors during the times when the chosen H1 scattered-light or L1 tomte glitch were present. The injection are done *coherently* across detectors, taking into account the detector responses and the arrival time delays of the GW. The injection time of the simulated signal is slightly adjusted so that the glitches overlap with the inspiral, intermediate and merger-ringdown stage of the signal, producing in total six data samples of glitch-overlapped signals. We developed and validated a specialized injection program to automate the above process, discussed in Appendix **??**. The Q-scans of these six data samples are plotted in Fig. 6.

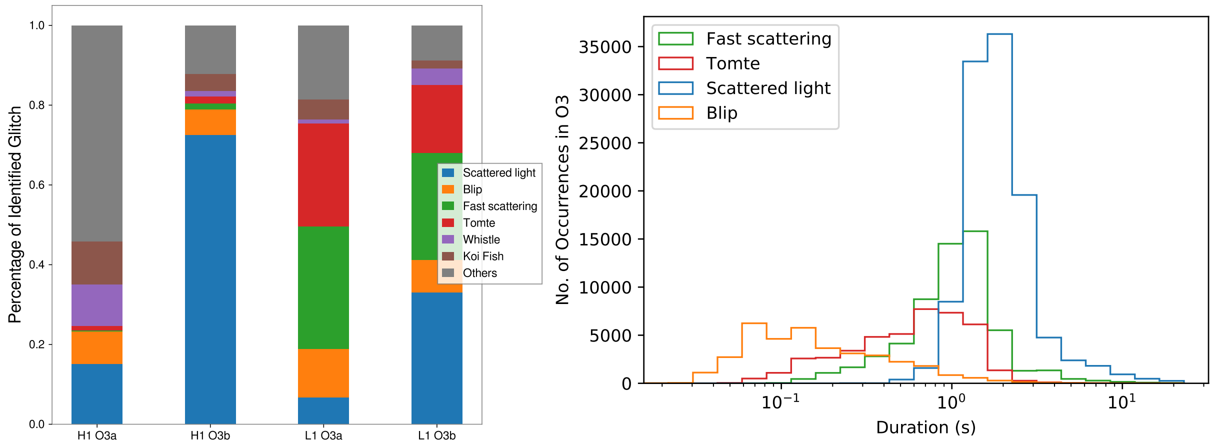


FIG. 5. Left: A distribution of Hanford (H1) and Livingston (L1) detector glitches identified and classified by *Gravity Spy* at 95% confidence during the O3a and O3b observing runs. Different colors denote different classes of glitches. Glitches which occur rarely (< 5% in O3) or at lower frequencies < 20 Hz are categorized into the "Others" class. Right: Probability densities of the duration of four classes of glitches retrieved from *Gravity Spy*. The duration of glitches are plotted in logarithmic scale. Tomte glitches are preferred over fast-scattering glitches in our study due to their shorter duration.

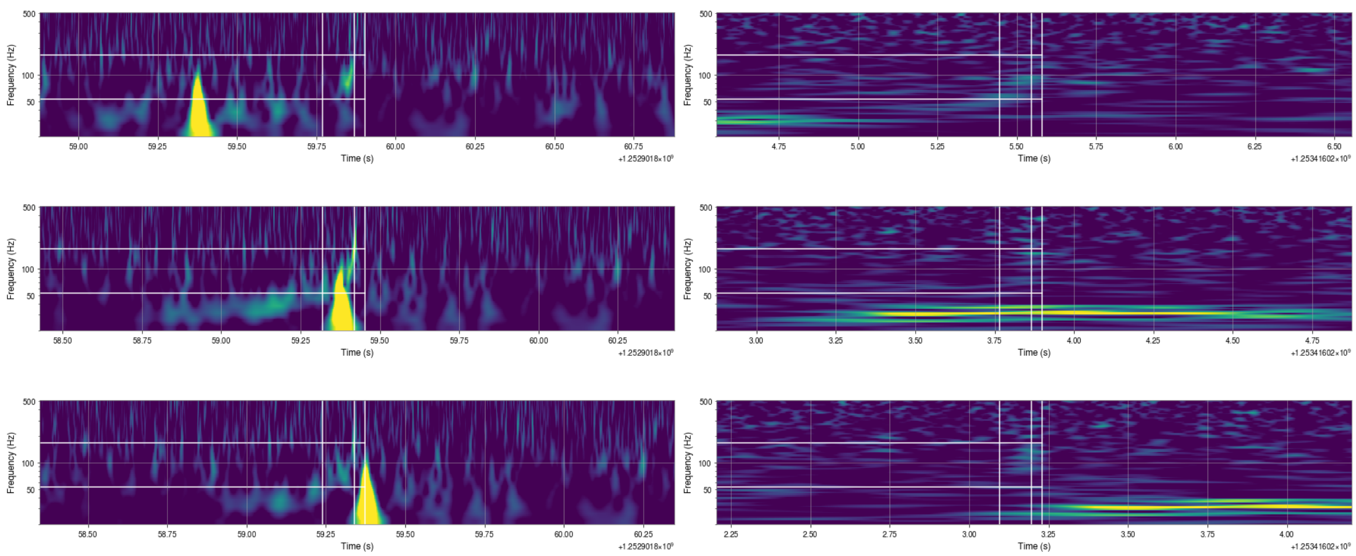


FIG. 6. Six data samples are prepared by injecting simulated `IMRPhenomPv2` waveform generated with the maximum likelihood parameters for S190828l coherently into detector outputs when a L1 tomte glitch (left) and a H1 scattered light glitch (right) are present. By slightly adjusting the time of injection, the glitches are made to affect the inspiral (top row), intermediate (middle row) and merger-ringdown (bottom row) stage respectively. The white grid lines mark the boundaries for the inspiral, intermediate and merger-ringdown stage of coalescence in time (left to right) and frequency (bottom to top).

Appendix A: Injection Tool

We have developed an injection program, `injhelper`, which automates the multi-step process of calculating detector responses and time delays, generating GW strain waveforms, retrieving detector output data and performing injections. Under the hood, it is a wrapper which passes source parameters to `LALSimulation` [21] and manipulates the output waveform using `GWpy` [44].

The program is validated with two independent ways. "Blind" injections were performed using `injhelper` and were successfully recovered (see Fig. 7). Bayesian parameter estimation were performed on injected data and the parameter values recovered are in fair agreement with the injected values (see Fig. 9).

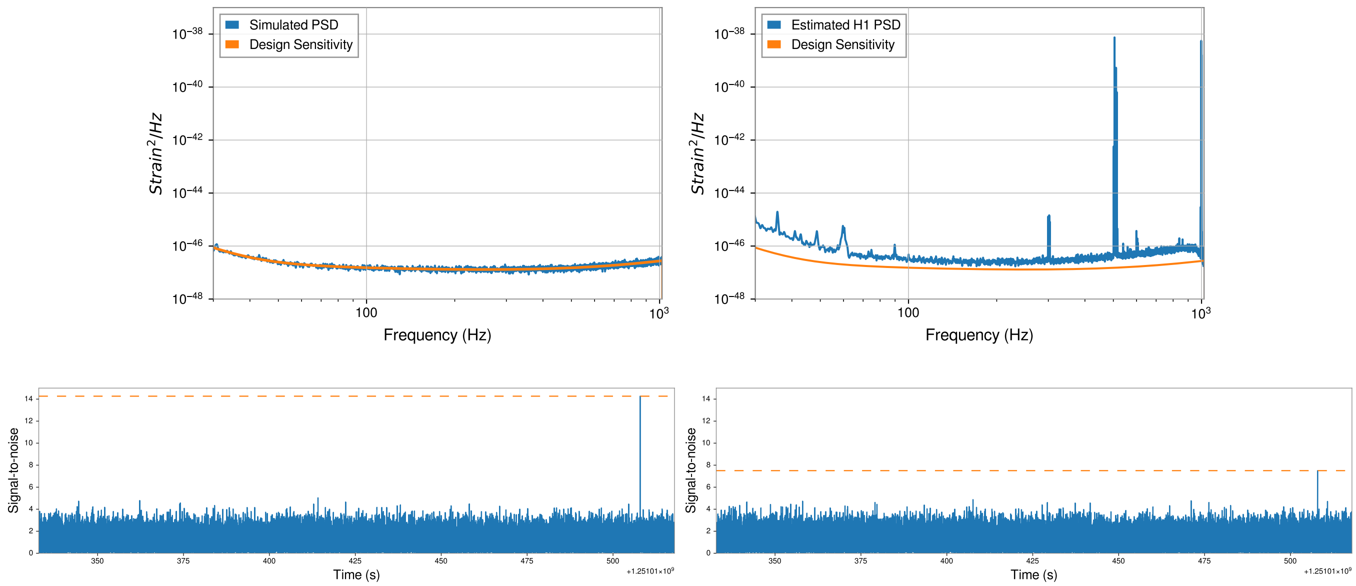


FIG. 7. "Blind" injections are performed on simulated Gaussian noise colored with the design sensitivity (left) and real H1 detector noise (right). Top: the estimated PSD (blue) are plotted with the design sensitivity (orange). Bottom: matched-filtering was performed and the injection times are successfully retrieved from the SNR peaks. This indicates that successful injections are performed by `injhelper`.

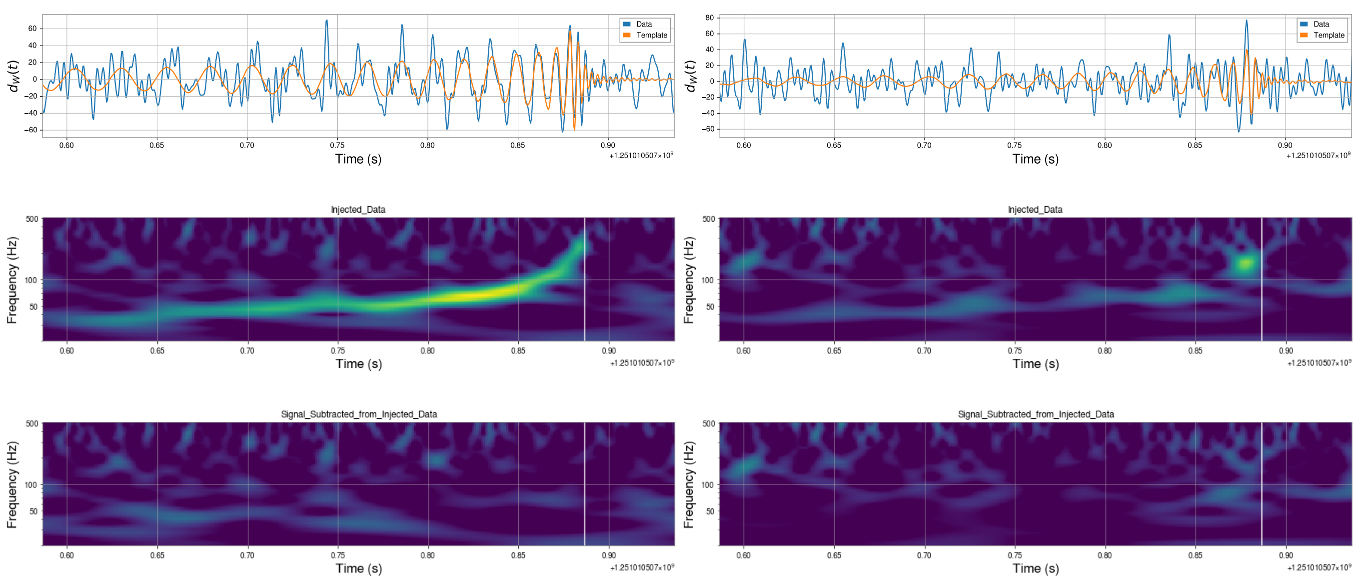


FIG. 8. Top: the simulated signal is aligned to the SNR peaks for simulated Gaussian noise colored with the design sensitivity (left) and real H1 detector noise (right) obtained in Figure 7. The simulated signal (orange) and data (blue) are whitened using the estimated PSD and plotted. Middle: Q-scans of the whitened data containing the injection are plotted. A signal which merges at the specified injection time (white vertical line) can be observed. Bottom: Q-scans of the residual data after subtracting the aligned simulated signal are plotted.

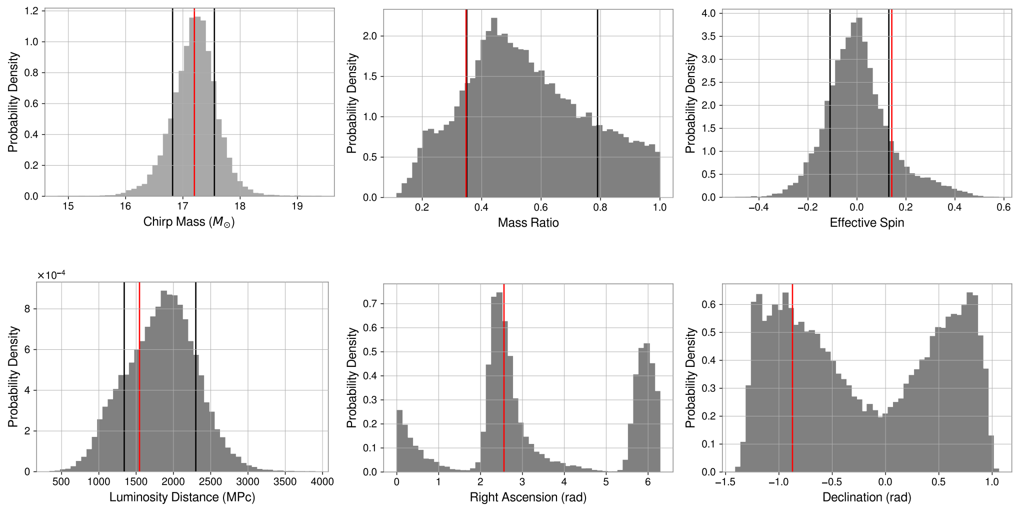


FIG. 9. Parameter estimation was performed on real H1 and L1 detector noise containing an injection. We assume an `IMRPhenomPv2` model for the signal. Marginalized posteriors are plotted for chirp mass, mass ratio, effective spin (top left to right), luminosity distance, right ascension and declination (bottom left to right). The red lines denote the injected value of the parameter, while the black lines for each parameter (except right ascension and declination) correspond to 1σ from the median. Injected values are seen to lie within, or lie near the boundary of the 1σ confidence interval. Bimodal distributions are obtained for right ascension and declination since only two detectors are used, yet the injected values are close to the peaks.

[1] E. Berti, E. Barausse, V. Cardoso, L. Gualtieri, P. Pani, U. Sperhake, L. C. Stein, N. Wex, K. Yagi, T. Baker, *et al.*, Classical and Quantum Gravity **32**, 243001 (2015), arXiv:1501.07274.

[2] C. W. Misner, K. S. Thorne, J. A. Wheeler, *et al.*, *Gravitation* (Macmillan, 1973).

[3] C. M. Will, Living reviews in relativity **9**, 3 (2006), arXiv:0103036.

[4] C. M. Will, Living reviews in relativity **17**, 4 (2014).

[5] K. Stelle, Physical Review D **16**, 953 (1977).

[6] B. Sathyaprakash, Living Rev. Relativity **12**, 2 (2009).

[7] J. Aasi, B. Abbott, R. Abbott, T. Abbott, M. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari, *et al.*, Classical and quantum gravity **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].

[8] F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, N. Allemandou, A. Allocca, J. Amarni, P. Astone, G. Balestri, G. Ballardin, *et al.*, Classical and Quantum Gravity **32**, 024001 (2014), arXiv:1408.3978 [gr-qc].

[9] B. Abbott, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt, *et al.*, Physical Review X **9**, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].

[10] R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt, M. Agathos, *et al.*, arXiv preprint arXiv:2004.08342 (2020).

[11] B. Abbott, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt, *et al.*, The Astrophysical Journal Letters **892**, L3 (2020).

[12] R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt, M. Agathos, *et al.*, The Astrophysical Journal Letters **896**, L44 (2020).

[13] F. Pretorius, in *Physics of relativistic objects in compact binaries: From birth to coalescence* (Springer, 2009) pp. 305–369.

[14] N. Yunes, K. Yagi, and F. Pretorius, Physical review D **94**, 084002 (2016).

[15] N. Yunes and X. Siemens, Living Reviews in Relativity **16**, 9 (2013).

[16] J. Meidam, K. W. Tsang, J. Goldstein, M. Agathos, A. Ghosh, C.-J. Haster, V. Raymond, A. Samajdar, P. Schmidt, R. Smith, *et al.*, Physical Review D **97**, 044033 (2018), arXiv:1712.08772 [gr-qc].

[17] E. Barausse, V. Cardoso, and P. Pani, Physical Review D **89**, 104059 (2014).

[18] L. Scientific, V. Collaborations, B. Abbott, R. Abbott, T. Abbott, M. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, *et al.*, Physical review letters **116**, 221101 (2016).

[19] B. Abbott, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt, *et al.*, Physical Review D **100**, 104036 (2019).

[20] P. R. Saulson, *Fundamentals of interferometric gravitational wave detectors* (World Scientific, 1994).

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

[22] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, *et al.*, Physical Review D **91**, 042003 (2015), arXiv:1409.7215 [gr-qc].

[23] L. K. Nuttall, T. Massinger, J. Areeda, J. Betzwieser,

S. Dwyer, A. Effler, R. Fisher, P. Fritschel, J. Kissel, A. Lundgren, *et al.*, Classical and Quantum Gravity **32**, 245005 (2015), arXiv:1508.07316 [gr-qc].

[24] M. Zevin, S. Coughlin, S. Bahaadini, E. Besler, N. Rohani, S. Allen, M. Cabero, K. Crowston, A. K. Katsaggelos, S. L. Larson, *et al.*, Classical and Quantum Gravity **34**, 064003 (2017), arXiv:1611.04596 [gr-qc].

[25] B. P. Abbott, R. Abbott, T. Abbott, M. Abernathy, F. Acernese, K. Ackley, M. Adamo, C. Adams, T. Adams, P. Addesso, *et al.*, Classical and Quantum Gravity **33**, 134001 (2016), arXiv:1602.03844 [gr-qc].

[26] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Physical review letters **113**, 151101 (2014), arXiv:1308.3271 [gr-qc].

[27] 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).

[28] J. D. Romano and N. J. Cornish, Living reviews in relativity **20**, 2 (2017).

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

[30] S. Khan, K. Chatziioannou, M. Hannam, and F. Ohme, Physical Review D **100**, 024059 (2019).

[31] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Physical Review D **93**, 044007 (2016), arXiv:1508.07253 [gr-qc].

[32] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Physical Review D **93**, 044006 (2016), arXiv:1508.07250 [gr-qc].

[33] C. Cutler and E. E. Flanagan, Physical Review D **49**, 2658 (1994).

[34] J. R. Smith, T. Abbott, E. Hirose, N. Leroy, D. MacLeod, J. McIver, P. Saulson, and P. Shawhan, Classical and Quantum Gravity **28**, 235005 (2011), arXiv:1107.2948 [gr-qc].

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

[36] R. Essick, L. Blackburn, and E. Katsavounidis, Classical and Quantum Gravity **30**, 155010 (2013), arXiv:1303.7159 [astro-ph.IM].

[37] 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, *et al.*, Physical Review D **88**, 062003 (2013), arXiv:1303.6984 [astro-ph.IM].

[38] C. Messick, K. Blackburn, P. Brady, P. Brockill, K. Cannon, R. Cariou, S. Caudill, S. J. Chamberlin, J. D. Creighton, R. Everett, *et al.*, Physical Review D **95**, 042001 (2017), arXiv:1604.04324 [astro-ph.IM].

[39] 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.*, Classical and Quantum Gravity **33**, 215004 (2016), arXiv:1508.02357 [gr-qc].

[40] N. J. Cornish and T. B. Littenberg, Classical and Quantum Gravity **32**, 135012 (2015), arXiv:1410.3835 [gr-qc].

[41] T. B. Littenberg and N. J. Cornish, Phys. Rev. D **91**, 084034 (2015), arXiv:1410.3852 [gr-qc].

[42] B. P. Abbott, R. Abbott, T. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari, V. Adya, *et al.*, Physical Review Letters **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].

[43] Private discussion with Rico.

[44] D. Macleod, A. L. Urban, S. Coughlin, T. Massinger, M. Pitkin, paulaltin, J. Areeda, E. Quintero, T. G. Badger, L. Singer, and K. Leinweber, gwpy.