

Template-based Gravitational-Wave Echoes Search Using Bayesian Model Selection

Rico K. L. Lo,^{1,2, a} Tjonnie G. F. Li,¹ and Alan Weinstein²

¹*Department of Physics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong*

²*LIGO Laboratory, California Institute of Technology, Pasadena, California 91125, US*

(Dated: November 20, 2018)

The ringdown of the gravitational-wave signal from a merger of two black holes has been suggested as a probe of the structure of the remnant compact object, which may be more exotic than a black hole. It has been pointed out that there will be a train of echoes in the late-time ringdown stage for different types of exotic compact objects. In this paper, we have developed a template-based search methodology using Bayesian statistics to search for echoes of gravitational waves. Evidence for the presence or absence of echoes in gravitational-wave events can be established by performing Bayesian model selection. The Occam's factor in Bayesian model selection will automatically penalize the more complicated model that echoes are present in gravitational-wave strain data because of its higher degree of freedom to fit the data. We found that the search methodology was able to identify gravitational-wave echoes with Abedi et al.'s echoes waveform model about 82.3% of the time in simulated Gaussian noise in Advanced LIGO and Virgo network and about 61.1% of the time in real noise in the first observing run of Advanced LIGO with $\geq 5\sigma$ significance. The analysis technique developed in this paper is independent of the waveform model used, and can be used with different parameterized echoes waveform models to provide more realistic evidence of the existence of echoes from exotic compact objects.

LIGO DCC Number: LIGO-P1800319

I. INTRODUCTION

As of this writing, the Laser Interferometer Gravitational-wave Observatory (LIGO) [1] and the Advanced Virgo [2] have successfully detected five compact binary coalescence events from binary black hole systems: GW150914, GW151226, GW170104, GW170608, GW170814 [3–7], one binary neutron star collision GW170817 [8], and one candidate gravitational-wave event LVT151012 [9]. These discoveries mark the beginning of a new era of gravitational-wave (GW) astronomy and astrophysics, where we can infer and probe the properties and structure of astronomical objects using gravitational waves.

During the inspiral stage of gravitational-wave emission from the coalescence of a compact binary system, for instance a binary black hole system, the two black holes spiral towards each other with an increasing orbital frequency. Eventually, they coalesce in the merger stage to form one single black hole. The final black hole then relaxes to a Kerr black hole during the ringdown stage.

Cardoso, Franzin and Pani [10] first pointed out that the ringdown part of the gravitational-wave signal can be used as a probe of the structure of a compact object. A very compact object, not necessary a black hole, with a light ring will also exhibit a similar ringdown as that of a black hole. Cardoso, Hopper, Macedo, Palenzuela and Pani [11] further showed that a similar ringdown stage will also be exhibited for different types of *exotic compact objects* (ECOs) with a light ring (or a photon sphere), and there will be a train of echoes in the late-time ringdown stage associated with the photon sphere. Examples of ECOs are theoretical alternatives to black holes, such as gravastars and fuzzballs. A common feature of these alternatives is that there is some kind of structure near the

would-be event horizon. The echoes in the late-time ringdown stage are caused by repeated and damped reflections between the effective potential barrier and the reflective structure. Cardoso et al. also showed that the time delay between each echo Δt_{echo} can be used to infer the nature of an ECO [11], namely

$$\Delta t_{\text{echo}} \sim -nM \log\left(\frac{l}{M}\right), \quad (1)$$

where M is the mass of the ECO, $l \ll M$ is the microscopic correction of the location of the ECO surface from the Schwarzschild radius, and n is an integer of the order of 1 which depends on the nature of the ECO.

Abedi, Dykaar and Afshordi published a paper in December 2016, claiming that they have found tentative evidence of Planck-scale structure near the black hole event horizons at a combined 2.9σ significance level [12] on GW150914, LVT151012 and GW151226 using matched filtering technique. However, their analysis methodology, especially the estimation of statistical significance, was questioned [13, 14]. Various teams have also proposed methods to estimate the parameters of the gravitational-wave echoes, and to search for echoes in a morphology-agnostic way [15–18].

In this paper, we present a template-based search methodology using Bayesian inference to search for echoes of gravitational waves in compact binary coalescence events. The analysis technique in this paper can be used with different gravitational-wave echoes waveform models to provide robust evidence of the existence of echoes from ECOs by showing consistent results using different models. This would be a groundbreaking discovery if we detect an exotic compact object as this would revolutionize our understanding of compact objects, and that this can only be unveiled unequivocally by gravitational-wave observations. In parallel to this work, there are efforts to search for gravitational-wave echoes using Bayesian model selection with templates using the inference package PyCBC Inference, and published concurrently with this paper [19].

^a klllo@caltech.edu

This paper is structured as follows: In Sec. II, we first establish the methodology of the search, namely Bayesian model selection and parameter estimation in Sec. II A, gravitational-wave echoes template model in Sec. II B and statistical significance estimation in II C. We then describe ways to evaluate the sensitivity of a search in Sec. II D, and the combination of Bayesian evidence from multiple gravitational-wave echoes events in Sec. II E. In Sec. III, we first describe our implementation in III A, and then we present the results of a Bayesian parameter estimation and model selection of the presence of echoes versus their absence that were performed on simulated data with Gaussian noise in Sec. III C and Sec. III D respectively. Then we evaluate the performance of the search in simulated Gaussian noise and real noise in the first observing run (O1) in Sec. III E. Finally, we demonstrate the idea of combining multiple gravitational-wave echoes events in Sec. III F.

II. METHODS

A. Bayesian model selection and parameter estimation

To search for echoes of gravitational-wave from the coalescence of exotic compact objects, we perform Bayesian model selection analyses on confirmed gravitational-wave events. Here we consider two hypotheses \mathcal{H}_0 and \mathcal{H}_1 , which can also be considered as the null hypothesis and alternative hypothesis in the frequentist language, and they are

$$\begin{aligned}\mathcal{H}_0 &:= \text{No echoes in the data} \Rightarrow d = n + h_{\text{IMR}}, \\ \mathcal{H}_1 &:= \text{There are echoes in the data} \Rightarrow d = n + h_{\text{IMRE}},\end{aligned}$$

where d denotes the gravitational-wave data, n denotes the instrumental noise and h_{IMR} , h_{IMRE} denote the inspiral-merger-ringdown (IMR) gravitational-wave signal and inspiral-merger-ringdown-echo (IMRE) gravitational-wave signal respectively. Note that we *assume there is a gravitational-wave signal in the data* since we perform the search after the gravitational-wave signal has been identified, and we are only interested in knowing whether there are echoes in the data or not. When the null hypothesis \mathcal{H}_0 is true, that means the data contain a GW signal with IMR. When the alternative hypothesis \mathcal{H}_1 is true instead, that means the data contain a GW signal with both echoes and inspiral-merger-ringdown part.

In the context of gravitational-wave data analysis, suppose that the strain data $d(t)$ from a detector only consist of noise $n(t)$, which we assume to be Gaussian and stationary (we will relax these assumptions in what follows). The probability that the noise $n(t)$ has a realization $n_0(t)$ (with zero mean) is given by [20]:

$$p(n_0) = \mathcal{N} \exp \left[-\frac{1}{2} \int_{-\infty}^{+\infty} df \frac{|\tilde{n}_0(f)|^2}{(1/2)S_n(f)} \right], \quad (2)$$

where \mathcal{N} is a normalization constant and $S_n(f)$ is the power spectrum density of noise. We introduce the notion of *noise-weighted inner product*, namely

$$\langle A|B \rangle = 4\Re \int_{f_{\text{low}}}^{f_{\text{high}}} df \frac{\tilde{A}^*(f)\tilde{B}(f)}{S_n(f)}, \quad (3)$$

where f_{low} and f_{high} are the low frequency cut-off and high frequency cut-off respectively. The integration is performed over a finite range because detectors are taking samples at a finite rate, and hence there is a theoretical upper limit on the maximum frequency that one can resolve from the data, and detectors are not sensitive enough below some frequency threshold. Using the inner product, we can rewrite Eq. 2 into

$$p(n_0) = \mathcal{N} \exp \left[-\frac{1}{2} \langle n_0|n_0 \rangle \right]. \quad (4)$$

Now, suppose the strain data $d(t)$ consist of both noise $n_0(t)$ and a GW signal modeled by a template $h(t; \vec{\theta})$, where $\vec{\theta}$ is a set of parameters of the template that describe the signal. That is

$$n_0(t) = d(t) - h(t; \vec{\theta}),$$

then the likelihood $p(d|\vec{\theta}, \mathcal{H}, I)$ for a single detector can be obtained from Eq. 2:

$$p(d|\vec{\theta}, \mathcal{H}, I) = \mathcal{N} \exp \left[-\frac{1}{2} \langle d(t) - h(t; \vec{\theta})|d(t) - h(t; \vec{\theta}) \rangle \right], \quad (5)$$

where I denotes the knowledge known prior to the selection, in this case we knew prior to the model selection that the data contain a GW signal. For the case of multiple detectors (for example, H1 and L1), if we assume that the noise distributions for each detector are all Gaussian and stationary, and more importantly independent of each other, then we have

$$p(d_{\text{H1}}, d_{\text{L1}}|\vec{\theta}, \mathcal{H}, I) = \prod_{i \in \{\text{H1}, \text{L1}\}} \mathcal{N}_i \exp \left[-\frac{1}{2} \langle d_i(t_i) - h(t_i; \vec{\theta})|d_i(t_i) - h(t_i; \vec{\theta}) \rangle \right]. \quad (6)$$

With the notion of noise-weighted inner product, we can also define the matched-filtering *signal-to-noise ratio (SNR)* ρ , which tells us how strong a signal is with respect to the

noise, as the following

$$\rho^2 = \frac{\langle d|h \rangle^2}{\langle h|h \rangle}, \quad (7)$$

where d denotes the gravitational-wave data and h is a gravitational-wave signal template. If we have multiple detectors, we can define the *network SNR* squared as the sum of matched filtering SNR squared in each detector as

$$\rho_{\text{network}}^2 = \sum_{i \in \{\text{H1}, \text{L1}\}} \rho_i^2. \quad (8)$$

In the optimal case of a template that exactly matches the signal in the data, the matched-filtering SNR is bound by the *optimal SNR*, which is given by

$$\rho_{\text{optimal}}^2 = \frac{\langle h|h \rangle^2}{\langle h|h \rangle}, \quad (9)$$

and can be used as an indication on how strong a signal is.

In Bayesian model selection, we compute the Bayes factor \mathcal{B}_0^1 and odds ratio \mathcal{O}_0^1 , which are defined as

$$\mathcal{B}_0^1 = \frac{p(d|\mathcal{H}_1, I)}{p(d|\mathcal{H}_0, I)}, \quad (10)$$

$$\mathcal{O}_0^1 = \frac{p(\mathcal{H}_1|d, I)}{p(\mathcal{H}_0|d, I)} = \mathcal{B}_0^1 \times \frac{p(\mathcal{H}_1|I)}{p(\mathcal{H}_0|I)}. \quad (11)$$

In the Bayesian language, the odds ratio has the interpretation that when $\mathcal{O}_0^1 > 1$, it means that the data favor the hypothesis \mathcal{H}_1 , and vice versa. For the sake of simplicity, we will drop the superscript and subscript on Bayes factor \mathcal{B} and odds ratio \mathcal{O} from now on when the context is clear. If we assume that each hypothesis is equally likely prior to the model selection, namely

$$p(\mathcal{H}_0|I) = p(\mathcal{H}_1|I) = \frac{1}{2},$$

then the odds ratio is simply the Bayes factor, that is

$$\mathcal{O} = \mathcal{B} = \frac{p(d|\mathcal{H}_1, I)}{p(d|\mathcal{H}_0, I)}. \quad (12)$$

It is often more convenient to work in log space, namely we compute log posterior, log likelihood and log prior. We take natural logarithm on both sides of Eq. 12, we have

$$\begin{aligned} \ln \mathcal{O} &= \ln \mathcal{B} \\ &= \ln p(d|\mathcal{H}_1, I) - \ln p(d|\mathcal{H}_0, I) \\ &= \ln Z_1 - \ln Z_0, \end{aligned} \quad (13)$$

where the term $Z_i \equiv p(d|\mathcal{H}_i, I)$, which is known as the *evidence* for the hypothesis \mathcal{H}_i , can be estimated by numerically integrating over the template parameter space $\vec{\theta}_i$ of hypothesis \mathcal{H}_i using a sampling algorithm such as Parallel-Tempering Markov Chain Monte Carlo with thermodynamic integration or nested sampling. In this paper, we will use nested sampling [21] (or more specifically LALInferenceNest [22]). Apart from the estimate of evidence, we can also obtain a set of posterior samples as byproducts of the nested sampling algorithm, which allow us to perform parameter estimation ¹

with little additional computational cost. We can calculate various estimators of parameters as point estimates from the posterior samples, such as the maximum likelihood estimator (MLE), which is

$$\hat{\vec{\theta}}_{i, \text{MLE}} = \arg \max \mathcal{L}(\vec{\theta}_i|d, \mathcal{H}_i, I),$$

where $\mathcal{L}(\vec{\theta}_i|d, \mathcal{H}_i, I) = p(d|\vec{\theta}_i, \mathcal{H}_i, I)$ is the likelihood as a function of the parameters $\vec{\theta}_i$. Another estimator is the maximum a posteriori estimator (MAP), which is

$$\hat{\vec{\theta}}_{i, \text{MAP}} = \arg \max p(\vec{\theta}_i|d, \mathcal{H}_i, I),$$

where $p(\vec{\theta}_i|d, \mathcal{H}_i, I)$ is the posterior distribution of parameters $\vec{\theta}_i$.

To obtain the evidence for the hypothesis \mathcal{H}_i in the context of gravitational-wave data analysis, we use gravitational-wave waveform templates that assume \mathcal{H}_i being true to compute the log likelihood.

1. Occam's factor

One must be cautious when performing model selection that the model which fits the data best does not imply that the model gives the highest evidence. A more complicated model, i.e. with more free parameters, is easier to be affected by noise in the data than a simpler model, i.e. with less free parameters. This is similar to over-fitting in regression. Suppose there are N data points for fitting, one can always use a degree $N - 1$ polynomial to fit all points, but very likely the fitted polynomial will not generalize well to new data because it was affected by the noise in the data.

Bayesian analysis embodies the *Occam's factor* and penalizes more complicated models automatically. Suppose there are two hypotheses, namely \mathcal{H}_0 and \mathcal{H}_1 , and without loss of generality, that $\dim \vec{\theta}_1 > \dim \vec{\theta}_0$, where $\dim \vec{\theta}_i$ denotes the dimension of the parameter vector $\vec{\theta}_i$ that describes the hypothesis \mathcal{H}_i . If the posterior distribution has a sharp peak at $\vec{\theta}_i = \vec{\theta}_{i, \text{MAP}}$ with width $\sigma_{i, \text{posterior}}$, then the integral for evidence Z_i can be approximated using Laplace's method. We first write the integral for evaluating the evidence of the hypothesis \mathcal{H}_i into the standard form for Laplace's method

$$Z_i = \int \exp \left\{ \ln \left[p(d|\vec{\theta}_i, \mathcal{H}_i, I) p(\vec{\theta}_i|\mathcal{H}_i, I) \right] \right\} d\vec{\theta}_i. \quad (14)$$

Let $f(\vec{\theta}_i) \equiv \ln \left[p(d|\vec{\theta}_i, \mathcal{H}_i, I) p(\vec{\theta}_i|\mathcal{H}_i, I) \right]$, and we expand $f(\vec{\theta}_i)$ about the sharp peak $\vec{\theta}_i = \vec{\theta}_{i, \text{MAP}}$, which gives

$$f(\vec{\theta}_i) = f(\vec{\theta}_{i, \text{MAP}}) + \frac{f''(\vec{\theta}_{i, \text{MAP}})}{2} (\vec{\theta}_i - \vec{\theta}_{i, \text{MAP}})^2 + \mathcal{O}(\vec{\theta}_i - \vec{\theta}_{i, \text{MAP}})^3,$$

with the first derivative f' vanishes and second derivative $f''(\vec{\theta}_{i, \text{MAP}}) < 0$ at the local maximum. Substituting back to

¹ A detailed discussion of parameter estimation can be found in [22]

Eq. 14, we have

$$Z_i \approx \exp f(\vec{\theta}_{i,\text{MAP}}) \int \exp \left[-\frac{|f''(\vec{\theta}_{i,\text{MAP}})|}{2} (\vec{\theta}_i - \vec{\theta}_{i,\text{MAP}})^2 \right] d\vec{\theta}_i, \quad (15)$$

where the integral becomes a Gaussian integral in the limit that the integration is performed over $(-\infty, \infty)$.

Finally we can approximate the evidence Z_i (up to some constant factors) by

$$Z_i \approx p(d|\vec{\theta}_{i,\text{MAP}}, \mathcal{H}_i, I) p(\vec{\theta}_{i,\text{MAP}}|\mathcal{H}_i, I) \sigma_{i,\text{posterior}},$$

note that we have assumed the posterior width $\sigma_{i,\text{posterior}}$ is much smaller than the width of integration limits such that the integral in Eq.15 can be well approximated by a Gaussian integral.

For a more complicated model, more parameters are needed to describe the observed data. For example, in our case the hypothesis \mathcal{H}_1 that there are echoes in the data, we need to introduce extra parameters (discussed in the next section) such as the time delay between each echo Δt_{echo} in the model selection analysis. Suppose the prior distribution of parameters $\vec{\theta}_i$ for each hypothesis is uniform over a width $\sigma_{i,\text{prior}}$ such that

$$p(\vec{\theta}_i|\mathcal{H}_i, I) = \begin{cases} \frac{1}{\sigma_{i,\text{prior}}} & \text{within the range} \\ 0 & \text{otherwise} \end{cases}.$$

The ratio $\sigma_{i,\text{posterior}}/\sigma_{i,\text{prior}}$ hence serves as a penalty to down-weight the evidence Z_1 of the more complicated model \mathcal{H}_1 which has a larger prior volume, i.e. $\sigma_{1,\text{prior}} > \sigma_{0,\text{prior}}$ to account for the uncertainty of the extra parameters. This ratio, sometimes referred as *Occam's factor* [23], allows the analysis to bias the less complicated IMR-only model in a natural way.

B. Phenomenological waveform model of echoes

In this paper, we used the phenomenological waveform model of echoes proposed by Abedi et al. in [12] to search for echoes of gravitational waves. It should be noted the methodology we proposed here is *independent of the gravitational-wave echoes templates we used*, and different parameterized waveform models can be readily used instead of the model by Abedi et al. when more physical models are available in the future [24–28]. Their model was motivated by the numerical results in [11]. There are five free parameters in their waveform model, with the phase change between each echo due to the reflection on an ECO surface being π . The descriptions of these 5 parameters are tabulated in Table I.

Using the notations in [12], the echo template $\mathcal{M}_{\text{TE,I}}(t)$ in time-domain is given by

$$\mathcal{M}_{\text{TE,I}}(t) \equiv A \sum_{n=0}^{\infty} (-1)^{n+1} \gamma^n \times \mathcal{M}_{\text{T,I}}(t + t_{\text{merger}} - t_{\text{echo}} - n\Delta t_{\text{echo}}, t_0), \quad (16)$$

Parameter	Description
Δt_{echo}	The time interval between each echo
t_{echo}	The time of arrival of the first echo
t_0	The time of truncation of the GW IMR template $\mathcal{M}_I(t)$ to produce the echo template $\mathcal{M}_{\text{TE,I}}(t)$
γ	The damping factor
A	The amplitude of the first echo relative to the IMR part of the template

Table I. The five free parameters and the corresponding descriptions of the phenomenological gravitational-wave echoes waveform model proposed by Abedi et al. [12]. In particular, Δt_{echo} is of the most astrophysical interest because it encapsulates the compactness of the exotic compact object that we are observing as shown in Eq. 1. Physically Δt_{echo} is related to the distance between the effective potential barrier and the reflective surface that gravitational-wave echoes need to travel. Also, A can inform us the typical strength of the gravitational-wave echoes emitted from exotic compact objects.

where t_{merger} is the time of merger² and $\mathcal{M}_{\text{T,I}}(t)$ is a smooth activation of the GW inspiral-merger-ringdown (IMR) template given by

$$\begin{aligned} \mathcal{M}_{\text{T,I}}(t) &\equiv \Theta(t, t_0) \mathcal{M}_I(t) \\ &\equiv \frac{1}{2} \left\{ 1 + \tanh \left[\frac{1}{2} \omega_1(t) (t - t_{\text{merger}} - t_0) \right] \right\} \mathcal{M}_I(t), \end{aligned}$$

where $\omega_1(t)$ denotes the angular frequency evolution of the IMR waveform as a function of time, and $\mathcal{M}_I(t)$ is the IMR waveform. The smooth activation $\Theta(t, t_0)$ essentially selects the ringdown, which is the part of a waveform that one might expect to see in echoes [12]. Note that the time of merger t_{merger} is the only time reference, and therefore we measure all time-related echo parameters t_0 , t_{echo} and Δt_{echo} with respect to t_{merger} . The top and bottom panel of Fig. 1 show a truncated IMR time-domain waveform $\mathcal{M}_{\text{T,I}}(t)$ used to generate the echo template and a GW150914-like IMRE time-domain waveform with three echoes, respectively.

In particular, the parameter that represents the time interval between each successive echo Δt_{echo} is of the most astrophysical interest because it encapsulates the compactness of the exotic compact object that we are observing as shown in Eq. 1. Physically Δt_{echo} is related to the distance between the effective potential barrier and the reflective surface that gravitational-wave echoes need to travel. The relative amplitude A can also inform us the typical strength of the gravitational-wave echoes emitted from exotic compact objects.

² Or equivalently time of coalescence, denoted by t_c .

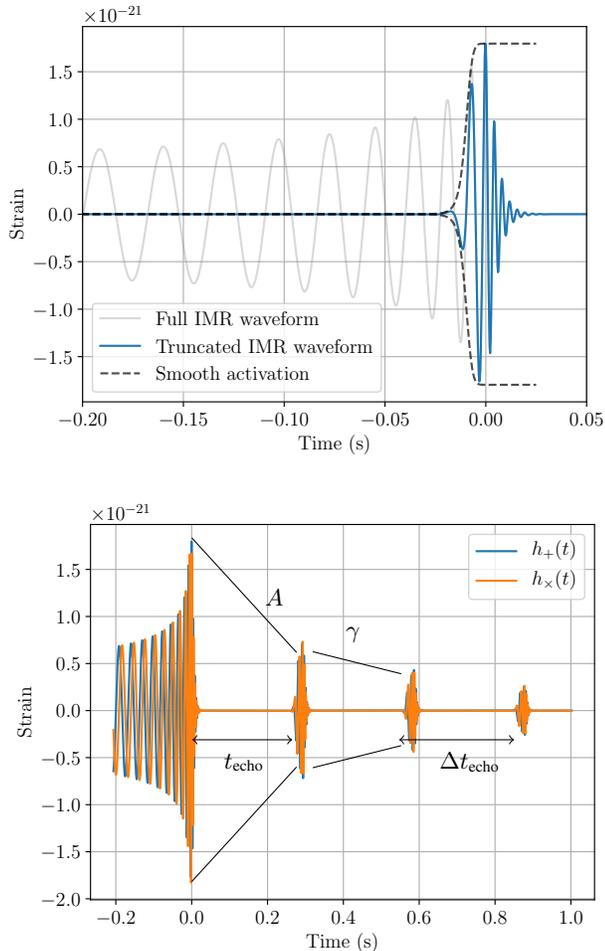


Figure 1. (Color online) *Top panel*: To generate a template of a gravitational-wave echo (in blue) we truncate the ringdown part of the inspiral-merger-ringdown (IMR) part (in gray) of a waveform by applying the smooth activation function $\Theta(t, t_0)$ (in black dashed lines) to get the truncated IMR waveform. *Bottom panel*: A plot of an inspiral-merger-ringdown-echoes (IMRE) template generated using the phenomenological waveform model of echoes proposed by Abedi et al. . We see that the first echo, which is the truncated IMR template (shown in the top panel) scaled by the parameter A , starts at t_{echo} after the merger. Subsequent echoes, which are further scaled down by the parameter γ due to the energy loss when the echo reflects off an ECO surface, are separated with each other in time by Δt_{echo} .

C. Detection statistic and background estimation for statistical significance

In this paper, the log Bayes factor $\ln \mathcal{B}$ (Eq. 10, 13) is the *detection statistic* to decide whether we claim there is an IMRE signal or an IMR signal in data. If the log Bayes factor $\ln \mathcal{B}_0^1$, or equivalently log odds ratio $\ln \mathcal{O}_0^1$, is greater than 0, we can conclude, from the Bayesian point of view, that the data favor the alternative hypothesis that the data contain an IMRE signal more than the null hypothesis that the data con-

tain an IMR signal, thus serving the function of distinguishing which hypothesis is more supported by the data.

After we have obtained a detection statistic, a natural question to ask is that how significant statistically is the detection statistic. Simply put, how likely is the detection actually caused by an IMRE signal but not due to noise? In the Bayesian school, there are different empirical scales, such as Jeffreys' scale, to interpret the strength of the Bayes factor. However, they are subjective and not universally applicable. Therefore, we are not going to use any of them in this paper.

Calculating the posterior probability of a hypothesis is certainly better than using a subjective scale to determine the strength of the Bayes factor. However, the Bayesian posterior probability fails to tell us how likely the evidence is simply due to random background noise, since we only consider one set of data. The frequentist approach can answer the following question: Given the null hypothesis \mathcal{H}_0 is true, how likely (i.e. the probability) are the data going to be *as extreme as or more extreme than* the observed data? The probability that we are looking for is exactly the frequentist p -value. We can also interpret this p -value as the *false-alarm probability*.

The p -value, where we denote it as simply p , is related to the *null distribution* of detection statistic $\ln \mathcal{B}$ by

$$\begin{aligned} p &= \Pr(\ln \mathcal{B} \geq \ln \mathcal{B}_{\text{detected}} | \mathcal{H}_0) \\ &= 1 - \int_{-\infty}^{\ln \mathcal{B}_{\text{detected}}} p(\ln \mathcal{B} | \mathcal{H}_0) d \ln \mathcal{B}, \end{aligned} \quad (17)$$

where $\ln \mathcal{B}_{\text{detected}}$ is the detection statistic obtained in an analysis on a segment of data, and $p(\ln \mathcal{B} | \mathcal{H}_0)$ is called the null distribution of $\ln \mathcal{B}$, i.e., the distribution of $\ln \mathcal{B}$ given that \mathcal{H}_0 is true.

Hence, from the null distribution, we can compute the detection statistic threshold $\ln \mathcal{B}_{\text{threshold}}$ corresponding to a certain statistical significance, e.g. 5σ and hence we can claim a detection of gravitational-wave echoes if the detection statistic of a candidate exceeds or equals to the predetermined threshold.

D. Evaluation of search sensitivity

Apart from getting the statistical significance of a particular candidate event of gravitational-wave echoes, we are also interested in investigating the sensitivity and accuracy of this search methodology using Bayesian model selection.

1. Sensitive parameter space

One of the methods to quantify the sensitivity of a search is the fraction of the parameter space of echo parameters that the search can determine whether the data contain echoes or not, given a threshold on the detection statistic $\ln \mathcal{B}_{\text{threshold}}$ and gravitational-wave detectors operating at a specific sensitivities. If a search is sensitive, then it should be able to cover a

reasonable fraction of parameter space possible for astrophysical exotic compact objects, which is schematically defined as

$$\{\vec{\theta}_{\text{echoes}} = (A, \gamma, t_0, t_{\text{echo}}, \Delta t_{\text{echo}}) \mid \ln \mathcal{B}(\vec{\theta}_{\text{echo}}) \geq \ln \mathcal{B}_{\text{threshold}}\}, \quad (18)$$

so that the search is able to detect the existence of echoes in the data with echo parameters in the sensitive parameter space.

2. Search efficiency

Another method to quantify the search sensitivity is to compute the probability that the existence of echoes will be detected given a detection statistic threshold, which is also known as the *efficiency* ζ . It is defined as

$$\zeta = \int_{\ln \mathcal{B}_{\text{threshold}}}^{\infty} p(\ln \mathcal{B} | \mathcal{H}_1) d \ln \mathcal{B}, \quad (19)$$

where $p(\ln \mathcal{B} | \mathcal{H}_1)$ is the *foreground distribution*, i.e. the distribution of $\ln \mathcal{B}$ given that \mathcal{H}_1 is true. If a search is sensitive, then it should have a high value of efficiency ζ .

E. Combining Bayesian evidence from a catalog of detection events

Bayesian model selection provides us a natural way to combine evidence of the existence of exotic compact objects from multiple detection events of gravitational-wave echoes. In the following analysis, we do not assume GW events are described by the same set of echo parameters. Suppose now we have a catalog of N_{cat} independent events so that we have a set of N_{cat} data denoted by $d = \{d_1, d_2, \dots, d_{N_{\text{cat}}}\}$, the odds ratio for the catalog of sources is given by

$$\begin{aligned} \mathcal{O}_0^1 &= \frac{p(d_1, d_2, \dots, d_{N_{\text{cat}}} | \mathcal{H}_1, I) p(\mathcal{H}_1 | I)}{p(d_1, d_2, \dots, d_{N_{\text{cat}}} | \mathcal{H}_0, I) p(\mathcal{H}_0 | I)} \\ &= {}^{(\text{cat})} \mathcal{B}_0^1 \times \frac{p(\mathcal{H}_1 | I)}{p(\mathcal{H}_0 | I)}, \end{aligned} \quad (20)$$

where

$${}^{(\text{cat})} \mathcal{B}_0^1 = \frac{p(d_1, d_2, \dots, d_{N_{\text{cat}}} | \mathcal{H}_1, I)}{p(d_1, d_2, \dots, d_{N_{\text{cat}}} | \mathcal{H}_0, I)}$$

is the *catalog Bayes factor*. Since each event is independent of each other, we can write the catalog Bayes factor as

$$\begin{aligned} {}^{(\text{cat})} \mathcal{B}_0^1 &= \frac{p(d_1, d_2, \dots, d_{N_{\text{cat}}} | \mathcal{H}_1, I)}{p(d_1, d_2, \dots, d_{N_{\text{cat}}} | \mathcal{H}_0, I)} \\ &= \prod_{i=1}^{N_{\text{cat}}} \frac{p(d_i | \mathcal{H}_1, I)}{p(d_i | \mathcal{H}_0, I)} \\ &= \prod_{i=1}^{N_{\text{cat}}} {}^{(i)} \mathcal{B}_0^1, \end{aligned} \quad (21)$$

where ${}^{(i)} \mathcal{B}_0^1$ is the Bayes factor obtained when performing the Bayesian model selection analysis on the i^{th} candidate of

gravitational-wave echoes event. Also, we can define the *catalog log Bayes factor*, which is simply

$$\ln {}^{(\text{cat})} \mathcal{B}_0^1 = \sum_{i=1}^{N_{\text{cat}}} \ln {}^{(i)} \mathcal{B}_0^1. \quad (22)$$

Hence, by multiplying the Bayes factor or adding the log Bayes factor from a catalog of gravitational-wave echoes events, we can combine the evidence of the existence of echoes in gravitational-wave data. Note that if the events share the same value of a parameter, e.g. Δt_{echo} , then the analysis is more complicated but still possible to do.

III. RESULTS

Before performing Bayesian model selection analyses on real gravitational-wave strain data, it is necessary to validate the performance of the search methodology by performing analyses on simulated strain data first, namely strain data with Gaussian noise and an IMRE signal of known parameters³. By recovering the injected signal and inferring the parameters correctly, we can validate that the analysis method proposed in this paper will be able to find signals in real strain data. After establishing the validity of the methodology, we can sample the background and foreground distribution of the detection statistic to estimate the statistical significance of a possible gravitational-wave echoes event, and the search efficiency in simulated Gaussian noise and real noise in the first observing run (O1) of Advanced LIGO. The gravitational-wave strain data in O1 of Advanced LIGO are publicly available in Gravitational Wave Open Science Center [29, 30].

A. Implementation

In this paper, we have made use of the software package `LALSuite` developed by the LIGO and Virgo collaboration [31]. In particular, we extensively used the module `LALSimulation` for its waveform generation interface and `LALInference` for its stochastic sampler [22]. We implemented the phenomenological waveform model of gravitational-wave echoes described in Sec. II B in `LALSimulation`, and we have used the IMR approximant `IMRPhenomPv2` [32–34] during the echo waveform generation. We have also modified `LALInference` so that the five extra echo parameters will be sampled by the program.

It should also be noted that in theory there should be infinitely-many gravitational-wave echoes. However, they are damped after each reflection from an ECO surface and more practically we are analyzing a finite segment of gravitational-wave data, making the detection of all the echoes in an event impossible. Therefore we will only put three echoes in the template during a search.

³ We call the ‘fake’ signals that were manually added to the data as injections

B. Details of the validation analysis

We performed our proposed search on a 8 second-long data with an IMRE signal injected into *simulated Gaussian noise* with Advanced LIGO-Virgo network to validate both the methodology and the implementation. We have chosen the prior distribution of the echo parameters to be uniform over a range (i.e. the *prior range*), and that the echoes will not overlap in time-domain. The prior range of the parameters used are listed in Table II. The IMR parameters of *this particular injected signal*, such as masses and spins, were chosen to be close to the inferred values of GW150914 [35] and the injected echo parameters were chosen *randomly* over the prior range. It should be noted that in this work *we do not assume IMR parameters were known a priori and they were allowed to vary during the validation analysis together with the echo parameters*. This is because the IMR parameters would affect the determination of the echo waveform used and thus the uncertainties in inferring IMR parameters would also propagate to the the search of echoes. For this particular injection, it has a log Bayes factor of 11.5, and a network optimal SNR of 63.8.

C. Parameter estimation

As an output of our search methodology, the set of posterior samples allow us to perform the parameter estimation on the simulated data. The search needs to accurately recover the injected IMRE signal if we are to use the proposed search methodology to search for gravitational-wave echoes in real data. A visualization of the sampled posterior distributions, i.e. a corner plot, that shows the estimated 1D marginal posterior probability distribution for each parameter and joint posterior probability distribution for each pair of parameters, is shown in Figure 2. We see that the inferred values of the echo parameters are both accurate (close to injected value) and precise (narrow posterior distribution), especially for those time-related parameters. For example, we see from the 1D histogram of Δt_{echo} in Figure 2 that the MAP is very close to the injected value (represented by the vertical blue solid line), and the 90% Bayesian credible interval ([0.2921, 0.2925] s) is much narrower than the prior range ([0.05, 0.5] s), that means the range is shrunk by about 99.91%.

As for amplitude-related parameters A and γ , the parameter estimation is not as accurate and precise as those time-related parameters. For instance, we see that the range for A is not shrunk as much as compared with Δt_{echo} , only by about 60%. This is not surprising because those time-related parameters can be inferred using the coherence of the strain with a template, while those amplitude-related parameters can only be inferred using noisy strain data.

By examining the corner plot for recovered IMR parameters (not shown here), we conclude that the parameter estimation of IMR parameters was not significantly affected by the introduction of five extra parameters.

Therefore, from the parameter estimation, we conclude that we have correctly implemented the gravitational-wave echoes

Parameter	Prior range
A	[0.0,1.0]
γ	[0.0,1.0]
t_0 (s)	[-0.1,0.01]
t_{echo} (s)	[0.05,0.5]
Δt_{echo} (s)	[0.05,0.5]

Table II. The prior range of the echo parameters. The prior distribution of each parameter is *uniform* over the respective prior range. The prior range for A was chosen as listed above because we do not expect the amplitude of echoes to be greater than the amplitude of the inspiral-merger-ringdown part of a signal. However, this is not a stringent requirement and can be easily relaxed. As for the prior range for γ , it was chosen as listed above because we expect echoes to be damped after each reflection from an ECO surface. The prior range for t_0 was chosen such that we are truncating approximately the ringdown part of a signal. As for the prior range for both t_{echo} and Δt_{echo} , they were chosen such that their predicted values for GW150914 as calculated in [12] fall within the corresponding prior range since we are injected a GW150914-like signal into simulated data.

waveform model and modified the sampler, and more importantly the search methodology is able to infer the values of echo parameters in actual analyses on candidate gravitational-wave echoes events as it has successfully recovered the injected IMRE signal accurately and precisely in this validation analysis.

D. Model selection

1. Statistical significance estimation of a candidate gravitational-wave echoes event

To estimate the statistical significance of a gravitational-wave echoes candidate, we sampled the null distribution $p(\ln \mathcal{B} | \mathcal{H}_0)$ of the detection statistic by performing background runs, i.e. data *with an IMR signal injected* (so that the null hypothesis \mathcal{H}_0 is true). The IMR parameters of the injection set used to estimate the background distribution were chosen to be representative of what the Advanced LIGO and the Advanced Virgo would detect, and *were not fixed to be the same as a particular gravitational-wave event*. We will discuss this choice in Sec. IV A.

The histogram of the sampled null distribution of *individual log Bayes factor* for simulated Gaussian noise (with 192 samples) and real noise during O1 of Advanced LIGO (with 953 samples)⁴ are shown in the left and right panel of Figure 3 respectively. The gray-scale bar in the top panel shows the statistical significance corresponding to the detection statistic.

⁴ The number of samples for background distribution in simulated Gaussian noise is less than that for real O1 noise because of the lack of computational resources

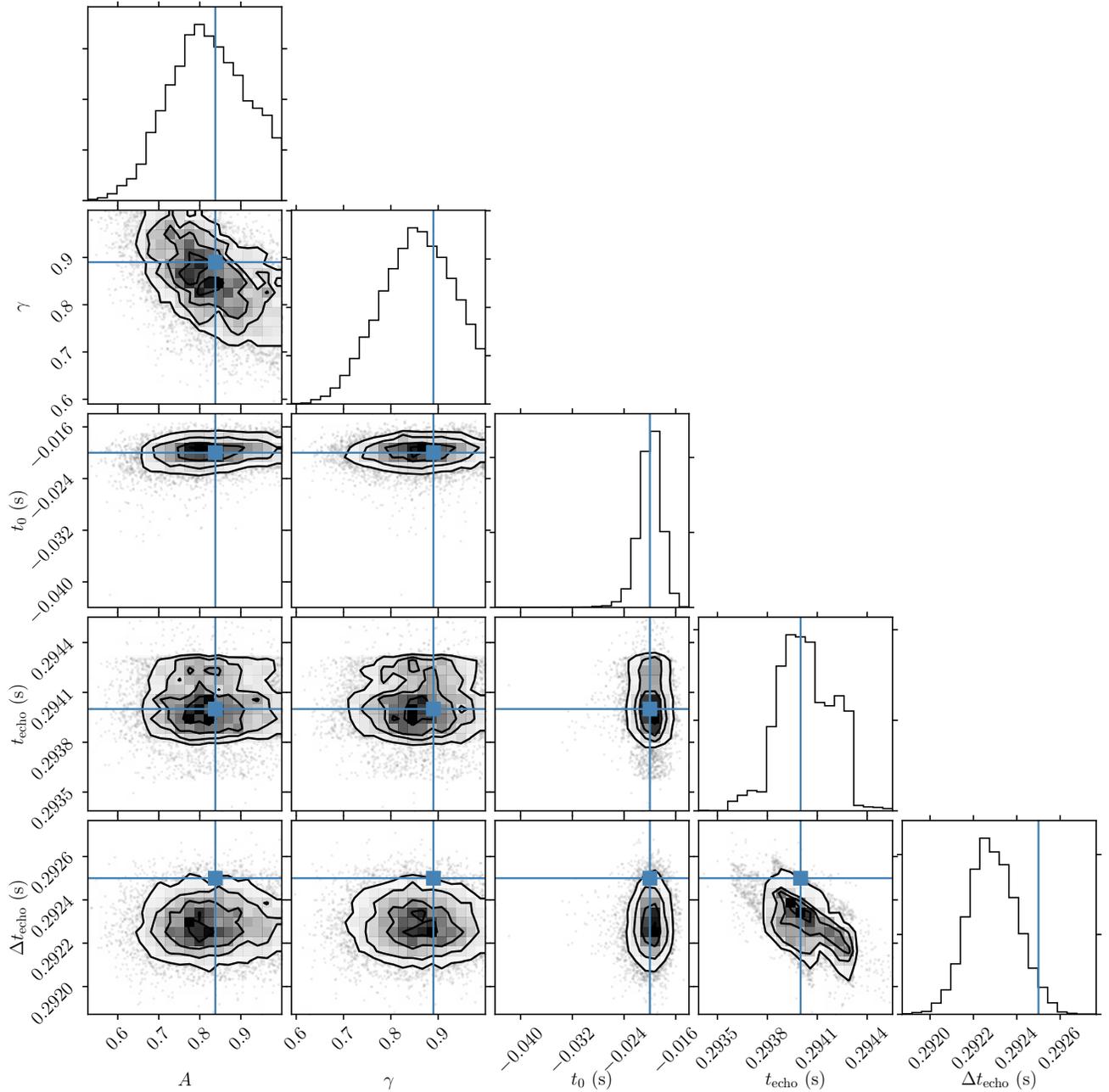


Figure 2. (Color online) A corner plot of the posterior samples from the parameter estimation on simulated data as described in Sec. III B and III C. If we are to use the proposed search methodology to search for gravitational-wave echoes in real data, then the search needs to accurately recover the injected IMRE signal. The blue solid lines represent the injected values for each parameter. Along the diagonal are histograms of the estimated 1D marginal posterior probability distribution for each parameter. The histograms show that the recovered parameters are both accurate (close to injected value) and precise (narrow posterior distribution). For example, the 1D marginal posterior probability density of Δt_{echo} is very narrow compared to its prior range tabulated in Table II, and the peak of the posterior probability distribution is very close to the injected value. The off-diagonal plots are the 2D histograms of the estimated joint posterior probability distribution of each pair of parameters, which show the correlation between pairs of parameters. We conclude that the search methodology is able to infer the values of the echo parameters in actual analyses on candidate gravitational-wave echoes events as it has successfully recovered the injected IMRE signal in the validation analysis accurately and precisely.

It should be noted that the p -value was obtained by extrapolation for $\gtrsim 3\sigma$ region, as sampling the $> 5\sigma$ region would require roughly 10^7 samples. For the case of simulated Gaussian noise in the left panel, we see that the null distribution peaks at about $\ln \mathcal{B} \approx -1$, and the tail of the distribution extends only slightly to $\ln \mathcal{B} > 0$. This means that it is unlikely for Gaussian noise to mimic gravitational-wave echoes. For the case of real noise during O1 in the right panel, we see that the distribution also peaks roughly at $\ln \mathcal{B} \approx -1$. However, the noise extends the tail of the distribution more significantly than the case in Gaussian noise. This means that it is more likely for real detector noises to mimic the effects of gravitational-wave echoes.

In particular, we injected an IMRE injection with echo parameters that *Abedi et al. claimed to have found in GW150914* into simulated Gaussian noise with Advanced LIGO-Virgo network, the detection statistic was found to be

$$\ln \mathcal{B}_{\text{detected, Gaussian}} = -0.2576 < 0.$$

That means in the Bayesian point of view, the data slightly favor the null hypothesis that the data do not contain echoes. We compute the p -value and the corresponding statistical significance, given that the noise is Gaussian, as the following:

$$\begin{aligned} p\text{-value} &= 0.01275, \\ \text{statistical significance} &= 2.234\sigma. \end{aligned}$$

This suggests that what was claimed to be found by *Abedi et al.* in GW150914, even for gravitational-wave detectors operating at design sensitivities and Gaussian noise, does not have the sufficient statistical significance to claim a detection (i.e. $\geq 5\sigma$) in the frequentist approach, and it is also inconclusive whether the data favor the existence of echoes in the data in the Bayesian approach. From the fact that in Figure 3 the null distribution for O1 real noise is more skewed to the right, we can expect that the statistical significance of what *Abedi et al.* had found is small and consistent with noise.

Table III tabulates the values of detection statistic $\ln \mathcal{B}$ that correspond to different levels of statistical significance in Gaussian and O1 background. If we want to make a gold-plated detection of gravitational-wave echoes, i.e. having statistical significance $\geq 5\sigma$, we can set the detection threshold as

$$\begin{aligned} \ln \mathcal{B}_{\text{threshold, Gaussian}} &= 1.9, \\ \ln \mathcal{B}_{\text{threshold, O1}} &= 5.7, \end{aligned}$$

in the case of Gaussian noise and real O1 noise respectively, so that any gravitational-wave echoes detection having a detection statistic greater than or equal to this threshold is a $\geq 5\sigma$ detection of echoes.

Statistical significance	Detection statistic (Gaussian noise)	Detection statistic (O1 noise)
1σ	-0.9	0.1
2σ	-0.4	1.5
3σ	1.1	4.0
4σ	1.5	5.4
5σ	1.9	5.7

Table III. The value of detection statistic $\ln \mathcal{B}$ and its corresponding statistical significance in both Gaussian and O1 background. If we want to make a gold-plated detection of gravitational-wave echoes, i.e. having statistical significance $\geq 5\sigma$, we can set the detection statistic threshold as $\ln \mathcal{B}_{\text{threshold, Gaussian}} = 1.9$ in the case of Gaussian noise, and $\ln \mathcal{B}_{\text{threshold, O1}} = 5.7$ so that any gravitational-wave echoes detection having a detection statistic greater than or equal to this threshold is gold-plated.

E. Search sensitivity, efficiency and accuracy

1. Sensitive parameter space of the search

Given the detection statistic threshold $\ln \mathcal{B}_{\text{threshold}}$, we would like to know what part of the parameter space of echoes we are able to see in the *optimal case*: gravitational-wave detectors operating at design sensitivities and that the instrumental noise is Gaussian. To achieve this, we performed analyses on simulated data injected with an IMRE signal with different values of the echo parameters of interest. In this particular study, the IMR parameters were fixed to be *GW150914-like*. We will be focusing on the two parameters that of the most astrophysical interest: the time interval between echoes Δt_{echo} and the relative amplitude A . The two parameters were varied one at a time.

Figure 4 shows plots of detection statistic $\ln \mathcal{B}$ as a function of Δt_{echo} with other echo parameters fixed to ($A = 0.6, \gamma = 0.89, t_0 = -0.02 \text{ s}, t_{\text{echo}} = 0.2940 \text{ s}$) (left) and as a function of A with other echo parameters fixed to ($\gamma = 0.89, t_0 = -0.02 \text{ s}, t_{\text{echo}} = 0.2940 \text{ s}, \Delta t_{\text{echo}} = 0.2925 \text{ s}$) (right) respectively. The horizontal dashed line on each plot corresponds to the detection statistic threshold of 5σ significance. Injections with detection statistic exceeding or equal to the threshold are marked with green ‘Y’, whereas injections with detection statistic lower than the threshold are marked with red ‘X’. From the left panel, we see that there is *no trend* for how the detection statistic is distributed with different values of Δt_{echo} , and that the search is able to detect gravitational-wave echoes with a range of Δt_{echo} (more specifically $[0.05, 0.5] \text{ s}$) as expected since different values of Δt_{echo} only shift the echoes in time, and whether the search is able to find echoes or not should not depend on their time of occurrence as long as they do not overlap. Therefore, fixing the values of time-related echo parameters when investigating the sensitive parameter space of A is justified. As for the relative amplitude A , we see from the right panel that there is a trend that signals with *smaller values of A have smaller values of detection statistic*,

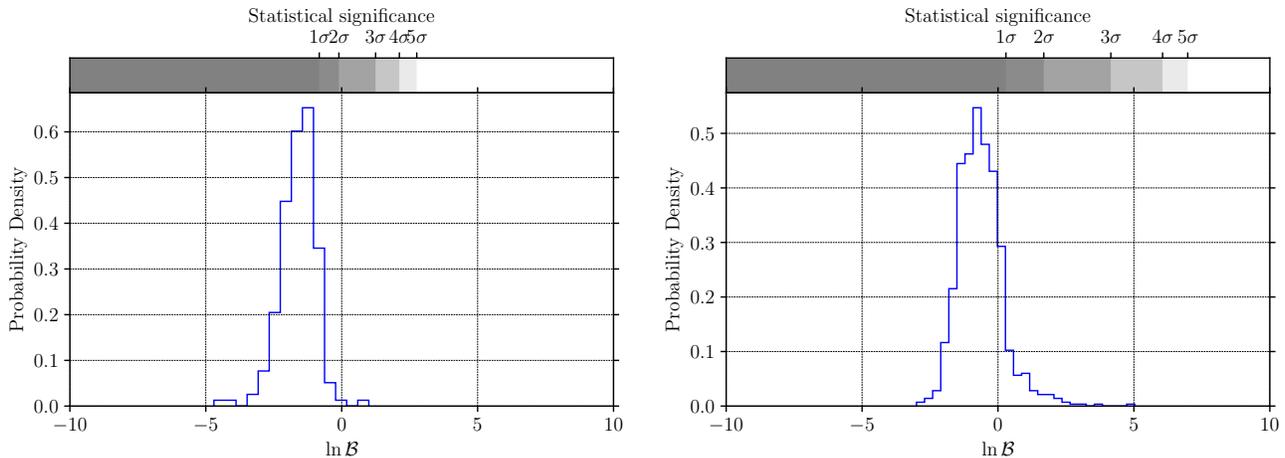


Figure 3. To estimate the statistical significance of a potential gravitational-wave echoes event for simulated Gaussian noise and real noise during O1 of Advanced LIGO, we sampled the null distribution $p(\ln \mathcal{B}|\mathcal{H}_0)$ of the detection statistic by performing background runs, i.e. data with an IMR signal injected. The histogram of null distribution for the case of Gaussian noise with 192 samples and for the case of O1 noise with 953 samples are plotted in the left and right panel respectively. The gray-scale bar in the top panel shows the statistical significance corresponding to the detection statistic (extrapolating from the 3σ region to 5σ region). *Left panel:* For the case of simulated Gaussian noise, we see that the null distribution peaks at about $\ln \mathcal{B} \approx -1$, and the tail of the distribution extends only slightly towards $\ln \mathcal{B} > 0$. This means that it is unlikely for Gaussian noise to mimic gravitational-wave echoes. *Right panel:* For the case of real noise during O1, we see that the distribution also peaks roughly at $\ln \mathcal{B} \approx -1$. However, the noise extends the tail of the distribution more significantly than the case in Gaussian noise. This means that it is likely for real detector noises to mimic the effects of gravitational-wave echoes.

and that the search can only pick up echoes with $A \gtrsim 0.3$ with $\geq 5\sigma$ significance. This is expected because the amplitude of echoes is damped and echoes with small amplitudes are buried in noise. This finding is consistent with what was reported in Westerweck et al. that only injections with strain amplitude $\gtrsim 10^{-22}$ in the echoes part could be recovered with the proper injected amplitude [17]. In Nielsen et al., a plot similar to the right panel of Fig. 4 was also shown [19], but it was unclear in their paper that at what value of their detection statistic (log Bayes factor for signal versus Gaussian noise, which is *different* from what we adopted in this paper) they are claiming a significant detection of echoes, and we will differ the discussion of the differences between two approaches in Sec. IV C 2. It should also be noted that there are cases with echo amplitude $A \gtrsim 0.1$ being found by our search, and the number we quoted for the sensitive parameter space for A is based on the loudest echo injection that were missed.

2. Foreground distribution and search efficiency

To compute the efficiency ζ of the search as described in Sec. IID 2, the foreground distribution $p(\ln \mathcal{B}|\mathcal{H}_1)$ of the detection statistic was sampled by performing foreground runs, i.e. analyses on simulated data with IMRE signals injected. The IMR parameters of the injection set used to estimate the foreground distribution were chosen to be representative of what the Advanced LIGO and the Advanced Virgo would detect, and *were not fixed to be the same as a particular gravitational-wave event*. We will discuss this choice in Sec.

IV A. As for the five echo parameters, their values were drawn randomly from the same distributions described in Table II.

A numerical integration of Eq. 19 on the foreground distribution sampled gives the the efficiency of the search as

$$\begin{aligned}\zeta_{\text{Gaussian}} &= 0.823, \\ \zeta_{\text{O1}} &= 0.611.\end{aligned}$$

That means, in the frequentist language, the search has a probability of 0.823 in the case of Gaussian noise and 0.611 in the case of O1 of detecting the existence of echoes (with parameters drawn uniformly from the priors in Table II) given that the data contain gravitational-wave echoes and that the detection has a $\geq 5\sigma$ significance.

3. Search accuracy

Given the detection statistic threshold $\ln \mathcal{B}_{\text{threshold, Gaussian}} = 1.9$ for the case of simulated Gaussian noise and $\ln \mathcal{B}_{\text{threshold, O1}} = 5.7$ for the case of O1, we make the claim that the data contain echoes only when the detection statistic is greater than or equal to the threshold. To gauge the performance of our proposed search methodology in terms of the ability to classify IMR and IMRE signals, a plot of receiver operating characteristic (ROC) curve is shown in Figure 5 for both the simulated Gaussian noise case and the O1 noise case. The ROC curve shows the fraction of IMRE signals that the search has properly identified as IMRE signals (also known as *true positive rate* for binary classifiers) given

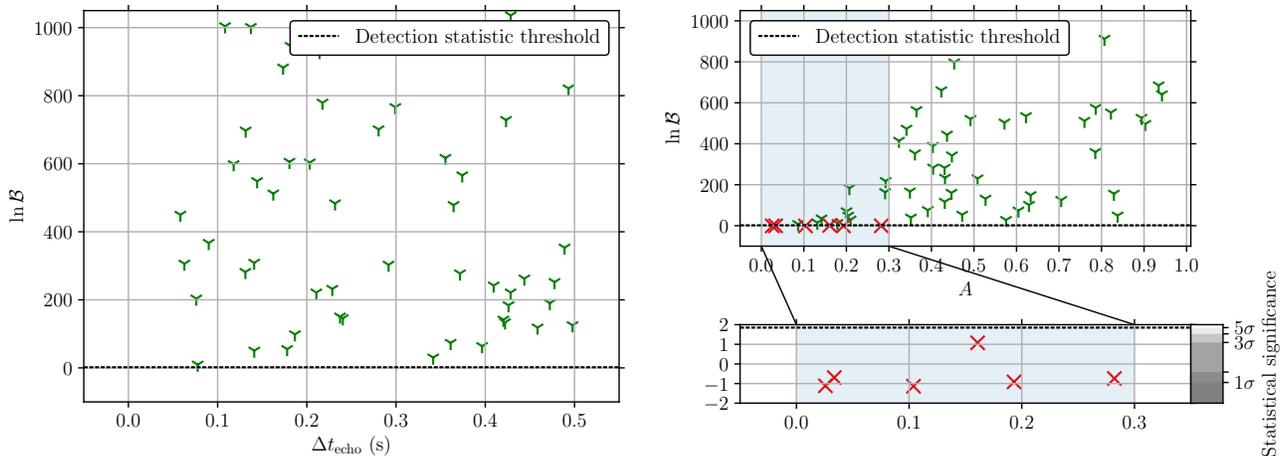


Figure 4. (Color online) To investigate which part of the parameter space of echoes we are able to see in the optimal case, namely gravitational-wave detectors operating at design sensitivities and that the instrumental noise is Gaussian, we performed analyses on simulated data injected with an IMRE signal with different values of the echo parameters of interest. *Left panel:* A plot of the detection statistic $\ln \mathcal{B}$ as a function of Δt_{echo} , with other echo parameters fixed to ($A = 0.6, \gamma = 0.89, t_0 = -0.02 \text{ s}, t_{\text{echo}} = 0.2940 \text{ s}$). *Right panel:* A plot of the detection statistic $\ln \mathcal{B}$ as a function of A , with other echo parameters fixed to ($\gamma = 0.89, t_0 = -0.02 \text{ s}, t_{\text{echo}} = 0.2940 \text{ s}, \Delta t_{\text{echo}} = 0.2925 \text{ s}$). The horizontal dashed line on each plot corresponds to the detection statistic threshold of 5σ significance. Injections with detection statistic exceeding or equal to the threshold are marked with green ‘Y’, whereas injections with detection statistic lower than the threshold are marked with red ‘X’. From the left panel, we see that there is *no trend* for how the detection statistic is distributed with different values of Δt_{echo} , and that the search is able to detect gravitational-wave echoes with a range of Δt_{echo} (more specifically $[0.05, 0.5] \text{ s}$) as expected since different values of Δt_{echo} only shift the echoes in time, and whether the search is able to find echoes or not should not depend on their time of occurrence as long as they do not overlap. Therefore, fixing the values of time-related echo parameters when investigating the sensitive parameter space of A is justified. As for the relative amplitude A , we see from the right panel that there is a trend that signals with *smaller values of A have smaller values of detection statistic*, and that the search can only pick up echoes with $A \gtrsim 0.3$ with $\geq 5\sigma$ significance. This is expected because the amplitude of echoes is damped and echoes with small amplitudes are buried in noise. It should be noted that there are cases with echo amplitude $A \gtrsim 0.1$ being found by our search, and the number we quoted for the sensitive parameter space for A is based on the loudest echo injection that were missed.

the fraction of IMR signals that the search has incorrectly identified as IMRE signals (also known as *false positive rate*). Equivalently, it can also be interpreted as showing how the efficiency of a search changes with detection statistic threshold. A more sensitive search will have a higher true positive rate given a false positive rate. For a random guess, the true positive rate is the same as the false positive rate. Therefore, the ROC curve of a useful search should be on the left of the spaces divided by the diagonal ROC curve. From Figure 5, we see that the search in simulated Gaussian noise performs better than the search in O1 noise as expected, and both of the searches perform better than random.

E. Demonstration of combining evidence from multiple gravitational-wave echoes events

From the investigation of sensitive parameter space of the search in the optimal case as described in Sec. III E 1, we see that the gravitational-wave echoes will need to have a relative strength $A \gtrsim 0.3$ in order to be picked up by our search with $\geq 5\sigma$ significance in the optimal Gaussian noise case. Statistically we can incorporate results from multiple events so that the detection statistic of many weak signals can be added

positively to stand out from the combined background, which adds negatively. Instead of combining the posterior distribution of echo parameters, which assumed the echo parameters for each event to be the same, we add the detection statistic log Bayes factor to form the *catalog log Bayes factor*, which is described in Sec. II E, and this approach only assumes each event to be independent and does not require the parameters of the events to be the identical; indeed, it is assumed that they are all different for each event.

Here we demonstrate how combining multiple events can help us to detect weak echoes and make a detection statement on a *collection of events* (i.e. a catalog). Suppose we have 10 potential gravitational-wave echoes events, which is to say we have $N_{\text{cat}} = 10$ events in the catalog. The catalog log Bayes factor is simply the sum of the log Bayes factor of each individual event according to Eq. 22. We sampled the null distribution for the catalog log Bayes factor $p(\ln^{(\text{cat})} \mathcal{B} | \mathcal{H}_0) = p(\sum_{i=1}^{N_{\text{cat}}} \ln^{(i)} \mathcal{B}_0^1 | \mathcal{H}_0)$ by picking $N_{\text{cat}} = 10$ events from the background distribution and computing the corresponding catalog log Bayes factor. The histogram of the sampled null distribution for the *catalog log Bayes factor* for simulated Gaussian noise and real noise during O1 of Advanced LIGO (both with 10000 samples in the sampled background distributions) are shown in the left and right panel

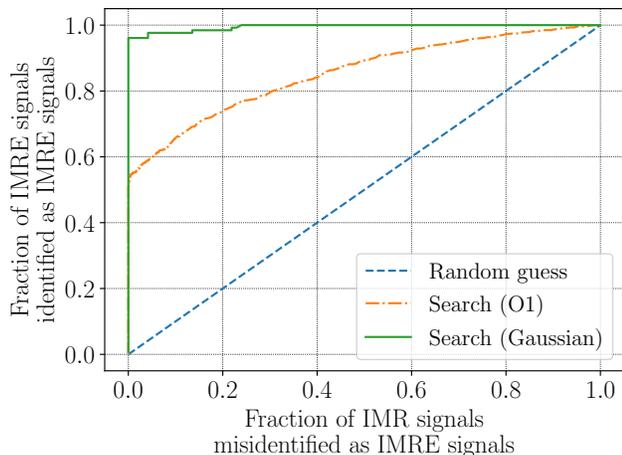


Figure 5. (Color online) To gauge the performance of our proposed search methodology in terms of the ability to classify IMR and IMRE signals, the receiver operating characteristic (ROC) curve for searches in O1 noise (orange dashed-dotted line) and Gaussian simulated noise (green solid line) respectively are shown. We see that the search in simulated Gaussian noise performs better than the search in O1 noise as expected because the fraction of IMRE signals identified as IMRE signals (with echo parameters drawn uniformly from the priors in Table II) for simulated Gaussian noise case is higher than that for O1 noise case for a given fraction of IMR signals misidentified as IMRE signals. Also, both searches in simulated Gaussian noise and O1 noise outperform the random guess (blue dashed line) for the same reason described above.

of Figure 6 respectively. The gray-scale bar in the top panel shows the statistical significance corresponding to the detection statistic. Compared to the histograms for individual log Bayes factor shown in Figure 3, we see that the peak of the null distribution for catalog log Bayes factor for both the case of Gaussian noise and O1 noise shift to be more negative as expected. If we focus on the case of Gaussian noise, and assume that the mean of the log Bayes factor $\langle \ln \mathcal{B} \rangle$ is roughly -1 , then the mean of the catalog log Bayes factor of the size of $N_{\text{cat}} = 10$ should be

$$\langle \ln^{(\text{cat})} \mathcal{B} \rangle \approx N_{\text{cat}} \langle \ln \mathcal{B} \rangle = -10,$$

which is indeed the case in the right panel of Fig. 6. If we assume, for the sake of demonstration, we have observed 10 gravitational-wave echoes events similar to what Abedi et al. claimed to have found in GW150914, namely the individual log Bayes factor is about -0.2576 , then the catalog log Bayes factor will then become roughly

$$\ln^{(\text{cat})} \mathcal{B} \approx N_{\text{cat}} \times -0.2576 = -2.576,$$

which we can make a statement with $\geq 5\sigma$ significance that there are gravitational-wave echoes in one or more events in the catalog, but we will not be able to pinpoint which event has echoes.

From this example, we see that by combining the Bayesian evidence from multiple events, we can statistically make a de-

tection statement whether there are echoes or not in a collection of potential gravitational-wave echoes candidates, which may be too weak to be detected individually.

IV. DISCUSSIONS

A. Background and foreground distribution estimation

During background and foreground estimation, the IMR parameters were not fixed to the same as a particular gravitational-wave event but drawn from distributions that are representative of what the Advanced LIGO and the Advanced Virgo would detect. It is legitimate to do this because in our hypotheses (see Sec II A) we did not require the IMR parameters to be known a priori. By allowing the IMR parameters to be different during background and foreground estimation, our sampled foreground distribution can be used to estimate the efficiency of our search to detect gravitational-wave echoes for a variety of inspiral-merger-ringdown-echo signals. Similarly, our estimated background distribution can be used to estimate the false alarm probability for a variety of inspiral-merger-ringdown signals. Although the background and foreground estimation can also be done specifically for each individual gravitational-wave event, it will soon become computationally too expensive as we will be having more than ten binary black hole mergers in an observing run, for example in the third observing run of Advanced LIGO and Virgo, it was estimated that there will be about 35_{-26}^{+78} binary black hole merger events [36]. One can perform a follow-up analysis on interesting echo triggers found in this search to obtain a more accurate estimate of the false alarm probability and hence statistical significance.

B. Combining evidence from multiple gravitational-wave echoes events

As described in Sec. II E, we can combine Bayesian evidence by simply multiplying the Bayes factor from each independent event to give the catalog Bayes factor. Note that when combining evidence, we do not assume GW events are described by the same set of echo parameters, whereas in Abedi et al. they have assumed that each event has the same value of γ and $t_0/\Delta t_{\text{echo}}$, where Δt_{echo} denotes the average value of Δt_{echo} inferred in the events, which may not be the case in reality [12]. In addition, they have combined the GW events by summing up the ρ^2 of each event, without demonstrating that this is a proper way to combine multiple measurements.

Combining Bayesian evidence from multiple gravitational-wave echoes events can provide tighter constraint on the existence of gravitational-wave echoes than a single event. Note that the null distribution for the catalog Bayes factor (and hence catalog log Bayes factor) can be constructed from the null distributions of log Bayes factor in a catalog of events.

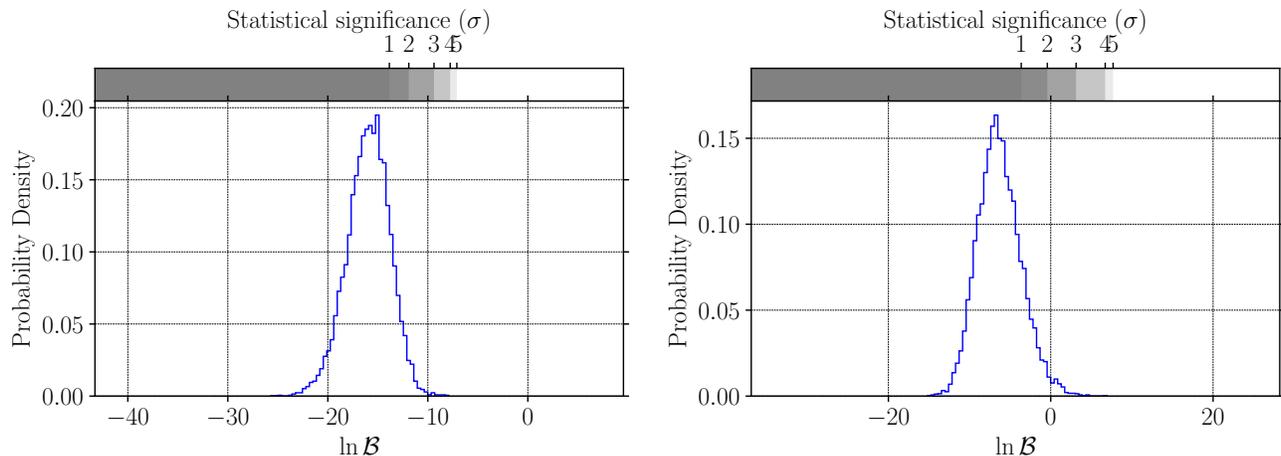


Figure 6. To detect gravitational-wave echoes which may be individually too weak to be detected, we can instead make a detection statement on a *collection of events* (i.e. a catalog). The histogram of the sampled null distribution for the *catalog log Bayes factor* for simulated Gaussian noise and real noise during O1 of Advanced LIGO (both with 10000 samples) are shown in the left and right panel respectively. The gray-scale bar in the top panel shows the statistical significance corresponding to the detection statistic. Compared to the histograms for individual log Bayes factor shown in Figure 3, we see that the peak of the null distribution for catalog log Bayes factor for both the case of Gaussian noise and O1 noise shift to be more negative as expected. Since weak signals add positively while the background events add negatively in the catalog log Bayes factor, weak gravitational-wave echoes signals can be detected as a whole, and statement whether there are echoes or not in a collection of potential gravitational-wave echoes candidates can be made but not individually which event/events in the catalog has/have echoes.

C. Comparisons with other proposed search methodologies for gravitational-wave echoes

1. Model-dependent method proposed by Abedi et al.

As mentioned in the introduction, Abedi et al. have proposed a template-based search methodology for echoes using matched-filtering. The parameter estimation was achieved by maximizing the square of signal-to-noise ratio ρ^2 , which is also defined in Eq. 7. The set of echo parameters that give the highest value of ρ^2 are said to be the inferred values in their analysis. Also, they used ρ^2 as the detection statistic of their search. By finding the number of events, in segments of data without gravitational waves, that have a higher or equal value of the detection statistic found in a candidate, the background distribution of their detection statistic can be estimated.

However, the use of ρ^2 as the detection statistic is sub-optimal because a large short-time instrumental noise fluctuation (also known as a glitch) can easily cause a peak in ρ^2 , and as a result the search will be trying to over-fit the glitch instead of echoes. Also, the addition of five echo parameters when searching for echoes in gravitational-wave data will often make IMRE templates fit the data better than IMR templates when using ρ^2 as the detection statistic as there are more free parameters to be adjusted to fit the noise in the data. The Occam's factor (described in Sec. II A 1) embodied in Bayesian model selection can mitigate the problem described above by penalizing more complicated models (i.e. having more free parameters), making our choice of log Bayes factor more robust against noise than ρ^2 as chosen by Abedi et

al. [12].

2. Model-dependent Bayesian model selection approach proposed by Nielsen et al.

In Nielsen et al., they adopted the same Bayesian model selection framework to search for the existence of gravitational-wave echoes. In their work, they are selecting between two hypotheses: gravitational-wave strain data consist of both echoes plus Gaussian noise versus the data consist only of Gaussian noise, and they are only looking at the post-merger part of a confirmed gravitational-wave signal. In comparison to our work, we are selecting between two hypotheses: gravitational-wave strain data consist of an inspiral-merger-ringdown-echo signal plus Gaussian noise versus an inspiral-merger-ringdown signal plus Gaussian noise model. Although both works concern whether the data contain echoes or not, the hypotheses that are being tested are not equivalent.

They interpreted their detection statistic (log Bayes factor) in the Bayesian way that when it is greater than zero, the data favor the echo signal plus Gaussian noise hypothesis and when it is less than zero, the data favor the pure Gaussian noise hypothesis. However, it is known and also pointed out by Nielsen et al. that the noise in real gravitational-wave strain data are not strictly Gaussian, and thus the aforementioned interpretation is only approximately true. We can also see this from Fig. 3 that the background distribution for our detection statistic, the log Bayes factor of IMRE versus IMR, is different for the case of Gaussian noise and real noise in O1, with a

noticeable tail to the right in the case of O1.

When interpreting the O1 results, they stated that a log Bayes factor having a value less than 1 does “not worth more than a bare mention” [19]. The use of a nomenclature to interpret (log) Bayes factor is sub-optimal because the scale to interpret the Bayesian evidence is not universally applicable. As mentioned in Kass et al. [37], for forensic evidence to be “conclusive” in court trials, the Bayes factor needs to be at least 10 times larger than what was originally suggested by Jeffrey, which suggests that the scale is not universally applicable to different situations. A rigorous justification that the scale proposed in Kass et al. [37] is appropriate was not given in that paper, and its authors merely mentioned, as quoted here: “From our own experience, these categories seem to furnish appropriate guidelines”. With a Gaussian noise model when performing the parameter estimation, the effects due to the non-Gaussianity of the noise can only be properly accounted by sampling the background distribution of the log Bayes factor in real data, which was done in our work as described in Sec. III D 1.

When generating templates of gravitational-wave echoes for parameter estimation, Nielsen et al. chose to fix the parameters that govern the inspiral-merger-ringdown part of the waveform such as component masses and the luminosity distance of the source of the signal. However, these inferred parameters have non-negligible uncertainties. By allowing the IMR parameters to vary during parameter estimation as we do in our work, we can marginalize over these parameters in model selection properly and hence getting a more accurate value for the log Bayes factor, instead of replacing the joint posterior distribution, that carries information such as correlation between parameters, with a product of Dirac delta functions.

3. Model-agnostic method proposed by Tsang et al.

In Tsang et al. [18], they also adopted the Bayesian approach and used log Bayes factor as the detection statistic. However, their search methodology is morphology-agnostic, meaning that no detailed knowledge of the waveform of gravitational-wave echoes is needed prior to a search. Their method is modified based on a search pipeline used for searching gravitational-wave bursts *BayesWave*, which is suitable for searching gravitational-wave signals that are un-modeled or poorly modeled. This is exactly the current status of the modeling of gravitational-wave echoes emitted from exotic compact objects, where there is no consensus that a particular waveform model can accurately model echoes. However, it is exactly because their search methodology requires no prior knowledge on the waveform of gravitational-wave echoes that echoes need to be loud in order to be detected by their search. As in our proposed search methodology, we can make use of the knowledge on the waveform to extract weaker echoes buried in noise and the estimation of physical parameters of echoes can be done easily.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have demonstrated that our proposed search methodology using Bayesian model selection between the presence of echoes and their absence can identify and estimate the parameters of an inspiral-merger-ringdown-echo (IMRE) signal buried in both Gaussian noise and real noise in the first observing run (O1) of Advanced Laser Interferometer Gravitational-wave Observatory (LIGO). In the validation test, the recovered echo parameters were both close to the true value and having narrow posterior probability distributions. We demonstrated that we can use a Bayesian model selection to test the existence of echoes in simulated data, and report the statistical significance of the detection. By performing many analyses on simulated data with gravitational-wave echoes injected, we also found that the search was able to identify gravitational-wave echoes in simulated Gaussian noise about 82.3% of the time and in O1 real noise about 61.1% of the time with $\geq 5\sigma$ significance.

In the short term, analyses will be performed on all of the gravitational-wave events so far to search for echoes in real gravitational-wave data. In the future, we can repeat the analysis with different parameterized gravitational-wave echoes waveform model that are more physical to provide more realistic evidence on the existence of echoes from exotic compact objects. When we understand the physics of exotic compact objects better in a sense that we can come up with physical waveform models of the echoes from different types of exotic compact objects, the methodology proposed in this paper can be readily modified to test the nature of exotic compact objects, using sub-hypotheses of \mathcal{H}_1 such as $\mathcal{H}_{\text{Gravastar}}$ and $\mathcal{H}_{\text{Fuzzball}}$, that is

$$\mathcal{H}_1 = \mathcal{H}_{\text{Gravastar}} \vee \mathcal{H}_{\text{Fuzzball}} \vee \dots,$$

so that we can learn even more about the properties and structure of exotic compact objects.

ACKNOWLEDGMENTS

The authors acknowledge the generous support from the National Science Foundation in the United States. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation and operates under cooperative agreement PHY-0757058. Virgo is funded by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale della Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by Polish and Hungarian institutes. RKLL and TGFL would also like to gratefully acknowledge the support from the Croucher Foundation in Hong Kong. The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong (Project No. CUHK 24304317) and the Direct Grant for Research from the Research Committee of the Chinese University of Hong Kong. The authors acknowledge the use of IUCAA LDG cluster Sarathi for the computational/numerical work.

The authors are also grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459. The authors would like to thank Ajith Parameswaran for reviewing this paper during the LSC internal review. RKLL would also like to thank Zachary Mark, Rory Smith, Peter T. H. Pang, Ignacio Magana, Alex Nielsen, Ofek Birnholtz and Yanbei Chen for the fruitful conversations with the first author. Figure 2 was generated using the Python package `corner.py` [38]. This paper carries LIGO Document Number LIGO-P1800319.

- [1] LIGO Scientific Collaboration, J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso, and et al., *Classical and Quantum Gravity* **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].
- [2] F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, N. Allemandou, A. Allocca, J. Amarni, P. Astone, G. Balestri, G. Ballardin, and et al., *Classical and Quantum Gravity* **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].
- [3] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Phys. Rev. Lett.* **116**, 061102 (2016).
- [4] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Phys. Rev. Lett.* **116**, 241103 (2016).
- [5] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaboration), *Phys. Rev. Lett.* **118**, 221101 (2017).
- [6] The LIGO Scientific Collaboration, the Virgo Collaboration, B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, and et al., *ArXiv e-prints* (2017), arXiv:1711.05578 [astro-ph.HE].
- [7] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, and et al., *Physical Review Letters* **119**, 141101 (2017), arXiv:1709.09660 [gr-qc].
- [8] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, and et al., *Physical Review Letters* **119**, 161101 (2017), arXiv:1710.05832 [gr-qc].
- [9] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al., *Physical Review X* **6**, 041015 (2016), arXiv:1606.04856 [gr-qc].
- [10] V. Cardoso, E. Franzin, and P. Pani, *Physical Review Letters* **116**, 171101 (2016), arXiv:1602.07309 [gr-qc].
- [11] V. Cardoso, S. Hopper, C. F. B. Macedo, C. Palenzuela, and P. Pani, *Physical Review D* **94**, 084031 (2016), arXiv:1608.08637 [gr-qc].
- [12] J. Abedi, H. Dykaar, and N. Afshordi, *Phys. Rev. D* **96**, 082004 (2017), arXiv:1612.00266 [gr-qc].
- [13] G. Ashton, O. Birnholtz, M. Cabero, C. Capano, T. Dent, B. Krishnan, G. D. Meadors, A. B. Nielsen, A. Nitz, and J. Westerweck, *ArXiv e-prints* (2016), arXiv:1612.05625 [gr-qc].
- [14] J. Abedi, H. Dykaar, and N. Afshordi, *ArXiv e-prints* (2017), arXiv:1701.03485 [gr-qc].
- [15] A. Maselli, S. H. Völkel, and K. D. Kokkotas, *Phys. Rev. D* **96**, 064045 (2017), arXiv:1708.02217 [gr-qc].
- [16] R. S. Conklin, B. Holdom, and J. Ren, *ArXiv e-prints* (2017), arXiv:1712.06517 [gr-qc].
- [17] J. Westerweck, A. B. Nielsen, O. Fischer-Birnholtz, M. Cabero, C. Capano, T. Dent, B. Krishnan, G. Meadors, and A. H. Nitz, *Phys. Rev. D* **97**, 124037 (2018), arXiv:1712.09966 [gr-qc].
- [18] K. W. Tsang, M. Rollier, A. Ghosh, A. Samajdar, M. Agathos, K. Chatziioannou, V. Cardoso, G. Khanna, and C. Van Den Broeck, *Phys. Rev. D* **98**, 024023 (2018), arXiv:1804.04877 [gr-qc].
- [19] A. B. Nielsen, C. D. Capano, and J. Westerweck, *ArXiv e-prints* (2018), arXiv:1811.04904 [gr-qc].
- [20] M. Maggiore, *Gravitational Waves: Volume 1: Theory and Experiments* (Oxford University Press, 2007).
- [21] J. Skilling, *AIP Conference Proceedings* **735**, 395 (2004), <https://aip.scitation.org/doi/pdf/10.1063/1.1835238>.
- [22] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Ayloott, 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, *Phys. Rev. D* **91**, 042003 (2015), arXiv:1409.7215 [gr-qc].
- [23] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms* (Cambridge University Press, New York, NY, USA, 2002).
- [24] Z. Mark, A. Zimmerman, S. M. Du, and Y. Chen, *Phys. Rev. D* **96**, 084002 (2017).
- [25] H. Nakano, N. Sago, H. Tagoshi, and T. Tanaka, *Progress of Theoretical and Experimental Physics* **2017**, 071E01 (2017), arXiv:1704.07175 [gr-qc].
- [26] M. R. Correia and V. Cardoso, *Phys. Rev. D* **97**, 084030 (2018), arXiv:1802.07735 [gr-qc].
- [27] Q. Wang and N. Afshordi, *Phys. Rev. D* **97**, 124044 (2018), arXiv:1803.02845 [gr-qc].
- [28] A. Testa and P. Pani, *Phys. Rev. D* **98**, 044018 (2018), arXiv:1806.04253 [gr-qc].
- [29] Gravitational Wave Open Science Center, “The O1 Data Release,” .
- [30] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, in *Journal of Physics Conference Series*, *Journal of Physics Conference Series*, Vol. 610 (2015) p. 012021, arXiv:1410.4839 [gr-qc].
- [31] LIGO Scientific Collaboration, “LIGO Algorithm Library,” .
- [32] 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].
- [33] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044006 (2016), arXiv:1508.07250 [gr-qc].
- [34] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044007 (2016), arXiv:1508.07253 [gr-qc].
- [35] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al., *Physical Review Letters* **116**, 241102 (2016), arXiv:1602.03840 [gr-qc].
- [36] T. Dent and C. Pankow, “Astrophysical Rates of Gravitational-Wave Compact Binary Sources in O3,” (2018), OpenLVEM Town Hall Meeting, Amsterdam, Netherlands.
- [37] R. E. Kass and A. E. Raftery, *Journal of the American Statistical Association* **90**, 773 (1995), <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1995.10476572>.
- [38] D. Foreman-Mackey, *The Journal of Open Source Software* **24** (2016), 10.21105/joss.00024.