New framework to study unmodeled physics from gravitational wave data

Guillaume Dideron [1,2,*] Suvodip Mukherjee [3,†] and Luis Lehner [2,‡]

[1]*Department of Physics and Astronomy, University of Waterloo,*
200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
[2]*Perimeter Institute for Theoretical Physics,*
31 Caroline Street North, Waterloo, Ontario N2L 2Y5, Canada
[3]*Department of Astronomy and Astrophysics, Tata Institute of Fundamental Research,*
1, Homi Bhabha Road, Colaba, Mumbai 400005, India

A confident discovery of physics beyond what has been consistently modeled from gravitational wave (GW) data requires a technique that can distinguish between noise artifacts and unmodeled signatures while also shedding light on the underlying physics. We propose a new data analysis method, SCoRe (Structured Correlated Residual), to search for unmodeled physics in GW data, which addresses both of these aspects. The method searches for structure in the cross-correlation power spectrum of residual strains of pairs of GW detectors by projecting this power spectrum onto a frequency-dependent template. The template may be model independent or model dependent and is constructed based on the properties of the GW source parameters. The projection of the residual strain enables distinction between noise artifacts and any true signal while capturing possible dependence on the GW source parameters. Our method is constructed within a Bayesian framework, and we demonstrate its application on a model-independent toy example and for a model motivated by an effective field theory of gravity. The method developed here will be useful for searching for a wide variety of new physics and yet-to-be-modeled known physics in GW data accessible from the current network of LIGO-Virgo-KAGRA detectors, as well as from future earth- and space-based GW detectors such as A+, LISA, Cosmic Explorer, and the Einstein Telescope.

I. INTRODUCTION

The direct detection of gravitational waves (GWs) by the LIGO-Virgo Collaboration [1] and the discovery of nearly 100 GW sources since then [2–4] has opened a new avenue to study fundamental physics using compact binaries in relativistic regimes. Binary neutron stars (BNSs), neutron star-black holes (NSBHs), and binary black holes (BBHs) provide unprecedented opportunities to explore the nature of gravity over a vast range of mass scales and cosmological distances. Among the many intriguing questions that can be addressed with GWs is whether data from the LIGO-Virgo-KAGRA (LVK) and future detectors can reveal deviations from our current Standard Model (SM) for GWs. This model includes (nonexhaustively) our best theory of gravity—General Relativity (GR), black holes and/or neutron stars describing compact objects, and the extent to which relevant physical features can be consistently accounted for in our waveform templates (for example, how well certain ranges of parameters, such as

high eccentricity or high mass ratios, can be captured by numerical template banks), along with our understanding of the noise in the GW detection facilities. We will refer to any deviation from this combination of assumed knowledge as a beyond modeled (BM) signature. For each event, the standard model is used to produce a best-fit waveform and, crucially, to interpret the physical content of the signal. Current tests for BM signatures examine potential deviations from this best-fit waveform.

Null tests are arguably the cleanest approach to testing the current SM. They determine whether signals are consistent with the expectations of the model. However, on their own, they do not inform us of the physical meaning of a given deviation. To that end, other tests have been devised that target parametrized deviations in the post-Newtonian (and post-Einsteinian) description of the inspiraling behavior and/or quasinormal modes of the newly formed black hole that might result from a merger (see e.g. [5,6] and references cited therein). With these tests, such deviations can, in principle, be closely connected with specific effects from particular theories—or at least provide physical information on the main impact of deviations from GR [7]. On the other hand, agnostic tests examine the residual (between the consistent SM signal and

*gdideron@perimeterinstitute.ca
†suvodip@tifr.res.in
‡llehner@perimeterinstitute.ca

the observed data) and aim to qualify the behavior of any departure (see e.g. [5,8]).[1] This approach can accommodate yet-to-be-modeled physics. At this point, we find it worthwhile to highlight an obvious fact; eventually, for sufficiently high SNR, all models (even within the SM) will show limitations due to systematics, such as physics still to be computed, yet unknown physical ingredients to be accounted for, or intrinsic waveform modeling errors. This is particularly true for future GW detectors. Their sensitivity will allow thousands of events to be combined to test for the presence of a BM signal. Even small systematic errors in waveform models can then accumulate and cause the misidentification of a BM deviation as shown in [10]. The problem is amplified when events overlap, which may frequently occur in future detectors. Thus, devising suitable strategies to analyze residuals will become increasingly important.

To date, available observations obtained with the LVK network have already made it possible to put constraints on BM phenomenology. For instance, by constraining beyond-GR gravity and potential effects associated with exotic compact objects. For the former, some constraints have been derived on deviations with particular physical consequences. For example, constraints have been derived on the coupling parameters of certain scalar-tensor theories in neutron star binaries [11,12], the graviton mass [5], and the number of possible space-time dimensions [13]. Others have been derived on deviation coefficients that may arise in many different beyond GR theories; for example on the post-Newtonian expansion coefficients [14,15], on propagation effects [16–19], on additional modes of polarizations, and on some potential signatures of black hole mimickers [20–23].[2] Constraints for all of the effects mentioned above have been presented by the LVK Collaboration [5,44].

Notably, no compelling evidence has so far been found for deviations from GR. However, as the sensitivity of GW detectors improves [45–47], new ground- [47] and space-based [48] detectors become available, and more events are detected, increasingly deeper tests will be possible. One key challenge in identifying a BM signature from the GW data is instrumental noise, especially in light of current evidence suggesting any deviation is likely to be subtle. Indeed, even in the scenario where the statistical properties of a detector are well characterized, it is difficult to

confidently assign the origin of a small BM signature as astrophysical and not from a known (or unknown) instrumental property. This is because the noise property of a detector can be far from a Gaussian distribution and it can also have nonstationary behavior [49–54]. In particular, glitches and other uncontrolled sources of noise make searching for BM signatures challenging [55]. Distinguishing the BM signature from detector noise is only the first challenge—as then one would also like to understand the physics of such signature and its dependence on the GW source properties.

Motivated by the above considerations, we propose a new method called *Structured Correlated Residual* (SCoRe) power spectrum. SCoRe searches for BM signatures by measuring the correlated residual power spectrum (CRPS) between multiple pairs of GW detectors. This method is designed such that it is not susceptible to contribution from uncorrelated noise between a pair of GW detectors and can search for any (un)modeled signals.

SCoRe applies cross-correlation, a technique used in searches for stochastic GW background [56,57] to the residual data obtained after subtracting a best-fit model of the signal from the strain data. This best-fit model is constructed using the values allowed by the posterior on the GW source parameters inferred from the detector network using the SM. The method then searches for a specific type of frequency dependence in the CRPS which can be unmodeled and driven by *chirplike* behavior, or model specific. The cross-correlation technique ensures that uncorrelated sources of noise will not contribute to the mean of the CRPS. Meanwhile, projection to a suitable chosen set of functions can help pick features missing from the SM for follow-up analysis on their physical meaning. This is described in detail in Sec. III. It is important to note that any source of correlated noise, such as the noise that arises due to the Schumann resonance [58–61], will contaminate the cross-correlation signal. We do not consider correlated noise in this work. Future studies will investigate the impact of correlated noise on SCoRe and how it can be mitigated.

In this work, we consider the best-fit model of the signal to be the maximum likelihood GR waveform obtained by parameter estimation of the network strain. In the presence of a BM signature, this estimator may have a "stealth bias". This is when a GR waveform with different source parameters can—even partially—account for the signature [62,63]. In the worst scenario, a different set of GR parameters may be able to perfectly reproduce the BM signature. It would then be impossible for any residual test to distinguish such BM signatures from the SM. In [64], the effect of stealth biases on the measurement of a particular BM model has been studied. The model consisted in numerical relativity (NR) waveforms produced through an order-reduced strategy designed to capture, to a certain extent, dynamical Chern-Simons gravity with source parameters similar to a detected signal, GW150914 [1].

---

[1]Rather than looking at the difference between the data and the best-fit waveform template, agnostic tests can also add deviations with general forms directly to the waveform templates, such as in [9], where splines are used to fit the deviations.

[2]Frameworks for obtaining these constraints are developed in [24] for the mass of the graviton, in [25–28] and [29–31] for propagation effects, and for exotic compact objects by analyzing the postmerger ringdown [32–38]. Some descriptions of possible exotic compact objects include [39–42]. The use of detectors to search for non-GR polarization modes was described as far back as 1973 [43].

Although the residual strain could partially be reproduced by GR waveforms, it was found that at least part of it could not be accounted for by the GR source parameters correcting for the BM signature, a consequence of deviations lying off the GR waveforms manifold. We note in passing that it is not guaranteed that all BM models have an orthogonal component to the SM manifold.

In fact, there exist beyond GR theories allowing for signals purely degenerate with GR waveforms for some source parameters and not for others. For example, theories that predict larger deviations at higher-curvature scales may not lead to measurable differences with GR for larger compact object masses. Another well-known example is spin-induced moment [65–68], which by definition is an effect proportional to the spin of the object. We do not explore such degenerate scenarios in this work.

This work is organized as follows. In Sec. II, we describe the motivation behind this new technique. In Sec. III and Sec. IV, we discuss the formalism of the work and its application to a simple toy example. In Sec. V, we show the application of this method to a specific theoretical model. Finally, we discuss the conclusions and future outlooks in Sec. VI.

In this work, we develop the mathematical framework of this new technique SCoRe and tested its performance for Gaussian stationary noise. In future work, we will consider its application on a population of GW sources detected by the LVK Collaboration and will include nonstationary and non-Gaussian noise.

II. MOTIVATION BEHIND SCORE

The search for BM physics using GWs is challenging due to multiple reasons, including both theoretical and experimental limitations. On one side, we are limited by the number of theoretical waveforms needed to test for the vast range of potentially BM scenarios (either for missing physics within GR, or sufficiently generic and complete waveforms beyond it). On the other side, we are limited by our understanding of the instrument noise properties when differentiating any BM signature from noise artifacts. Thus, often one fits the data with available waveform models to infer the best-fit source parameters and searches for any coherent residual signal [5,69–72]. However, the limitation remains in how well a framework can identify a BM signature and what the appropriate metric to quantify a deviation is. An example of this is the uneven performance of specific tests with different events (e.g. [8]).

A few salient features expected from any BM signature(s) in the GW data are (i) BM signatures will be common in the data of multiple detectors, but uncorrelated detector noise will not, (ii) a large class (if not all) of BM signatures can be understood as additional loss/gain in the orbital energy of a coalescing binary (e.g. [7,73–75]) when compared to the SM (GR, and sources involving black holes and/or

neutron stars), and (iii) BM signatures can depend on the GW source properties.

In this work, we propose a new technique to search for BM signatures in GW data that aims to address the issues mentioned above. This method has three key characteristics, each exploiting and/or targeting the three of BM signatures listed previously: (i) the residual GW data obtained after subtracting a modeled GW signal for a given waveform and source parameters are cross-correlated within pairs of detectors to help distinguish signal from noise; (ii) the CRPS is projected on a physically motivated template. This template is either composed of the derivatives of the orbital frequency up to order n (for a model-independent search), or derived from a particular physical model (for a model-dependent search) to help characterize the signal; and (iii) the projection is estimated as a function of the GW best-fit source parameters from the signal after marginalizing its uncertainties.

The first step helps in mitigating the contribution from uncorrelated noise to the residual signal by cross-correlating a total of $N_{\text{det}}(N_{\text{det}} - 1)/2$ pairs of detectors, where N_{det} is the number of detectors. Cross-correlation is a useful tool technique to search for weak signals from noisy GW data and is often used to search for stochastic GW background signals [56,57]. It is a time averaging over the product of the strains in two different detectors. Because the noise at the two detectors is not correlated, its mean average is zero, and its standard deviation decreases as the timescale of integration is increased. A price is paid however for this noise reduction: features shorter than the averaging timescale will be lost. The second step aids to look for a *physics-driven* BM signal in either a model-dependent or a model-independent way. Projecting onto template functions also further reduces noise by filtering particular, model-motivated structures in the signal. We suggest templates that are naturally expressible in terms of the orbital frequency. The final step takes into account possible dependencies of the BM signal on the GW source parameters and can also mitigate uncertainties related to the source parameters.

Although the idea of using template functions in frequency space is similar to the frames used in the BayesWave framework [76,77] (which can also be used as template functions in SCoRe), our method does not seek to reproduce the residual signal with a set of functions. This is unlike current residual tests that use BayesWave [5,72]. Instead, we measure the degree to which a predetermined finite set of functions can explain the residual strain. This makes SCoRe morphology-dependent, but the morphology of the chosen template functions is arbitrary and can be nonspecific. For example, one could reproduce the deviations parameters used in parametrized tests [6,72,78,79], but choosing an agnostic template set using SCoRe is also possible, as we will explain in Sec. III B.

By using cross-correlation and projection onto a template, we hope to reduce noise well below that of individual

detectors. With this noise reduction, SCoRe may be able to unearth minute BM signatures, which would otherwise remain undetected by only considering the strain of individual detectors. By looking at the structure of the CRPS, the method can help constrain potential deviations with a vast range of possible forms. This also means that the search for potential deviations can be made tighter as their form is guided with theoretical work. Furthermore, the method can also probe for correlations between the BM signature and the GW source parameters. A schematic diagram illustrating the basic principles of this technique is depicted in Fig. 1.

III. FORMALISM FOR SCORE

We describe below the formalism of SCoRe, which consists of four parts: (A) Cross-correlation; (B) Choice of a residual template; (C) Projection on a template; and (D) Inference using a Bayesian framework. We can write a linear model of the GW data $d^x(t)$ in a detector denoted by x as

$$d^x(t) = s^x(t) + n^x(t), \quad (1)$$

where $s^x(t) = \sum_i F_i^x(t) h_i(t)$ denotes the observed GW signal as seen by a detector, written in terms of the actual signal $h_i(t)$ for the polarization state "i", and the corresponding detector response function $F_i^x(t)$. The noise in detector x is $n^x(t)$. Whether a signal is detected in a particular detector depends on the noise in the detector and the detector response function $F_i^x(t)$ to the source position. In this work, we consider that a GW signal detected with a SM template in multiple detectors with a matched filtering SNR ρ greater than a minimum threshold network SNR ρ_* is classified as an event.[3] Performing parameter estimation over the strains in the detector network gives a posterior over the source parameters θ. By definition, the maximum likelihood estimate (MLE), θ_{MLE}, corresponds to the minimum residual strain amplitude in comparison to the variance in the detector noise, as θ_{MLE} should maximize the likelihood.

We propose to perform the test of BM physics on the residual data which is constructed using the MLE parameters inferred using a network of detectors, θ_{MLE} as

$$\Delta d^x_{\theta_{MLE}}(t) = s(t) - s_{BF}(t) + n^x(t),$$
$$= \Delta s(t) + n^x(t), \quad (2)$$

where $s_{BF}(t)$ is the SM waveform for θ_{MLE}. The MLE parameters need not be the true parameters of the source parameters—they only resemble the data best given a

---

[3] See Refs. [2–4] for a description of the actual pipelines used by the LVK to identify candidates.

waveform model and the assumption of the standard model. Since such parameters will be generically described with a distribution, we introduce in Sec. III D a Bayesian framework to marginalize their (posterior) distribution.

A. Cross-correlation of the residual signal

One of the key aspects of a BM signature in the residual data is that it is likely to be correlated in all the detectors. However, it will be observed differently in each detector depending on the detector response function. Noise between multiple GW detectors from unrelated sources will be random. To identify the origin of the residual as either astrophysical (a BM signature) or due to detector noise, we propose to do a cross-correlation of the residual signal between the data from two detectors. Given the output $d^x(t)$ and $d^{x'}(t)$ of two detectors x and x', the general cross-correlation $Y(\tau)$ of the two signals over a timescale τ is defined as

$$Y(\tau) \equiv \frac{1}{\tau} \int_{t-\tau/2}^{t+\tau/2} dt_1 \int_{t-\tau/2}^{t+\tau/2} dt_2 d^x(t_1) d^{x'}(t_2) Q(t_1, t_2), \quad (3)$$

where $Q(t_1, t_2)$ is a filter function. The choice of τ should balance two effects: correlated noise is reduced better as τ increases, but averaging washes away signal features with timescale shorter than τ. In the limit where τ spans the length of the signal, the cross-correlation becomes the total residual power observed. There are natural choices of τ, which rely on maximizing SNR while ensuring the features of the signal are preserved. They will depend on the source parameters, most importantly the chirp mass, which sets the timescale of the orbital frequency evolution, and the choice of template function on which to project the CRPS. We discuss the methodology behind choosing τ with a specific example in Sec. IV B.

The definition (3) has been used to search for stochastic GW signal [56,57]. There, the strain in multiple detectors is used to search for a common, periodic signal [80]. In stochastic GW background searches, the phase of the signal is not expected to be resolved. This makes choosing the optimal filter function nontrivial. By contrast, we can measure the phase of a well-detected event and measure Δt, the delay time separating the two detectors within a timing uncertainty. The optimal filter function, which gives the highest SNR, will correspond to the signals in the detectors overlapping in time. It is therefore $Q(t_1, t_2) = \delta(|t_1 - t_2| - \Delta t)$ (assuming $\tau \gg \Delta t$). We can then define the cross-correlation with maximum SNR with angular brackets:

$$\langle d^x(t) d^{x'}(t + \Delta t) \rangle \equiv \frac{1}{\tau} \int_{t-\tau/2}^{t+\tau/2} dt' d^x(t') d^{x'}(t' + \Delta t), \quad (4)$$

Deviation from standard model

Residual strain in detector x

Residual strain in detector x'

Detector cross-correlation

Physically motivated template

Projection of cross-correlation onto template
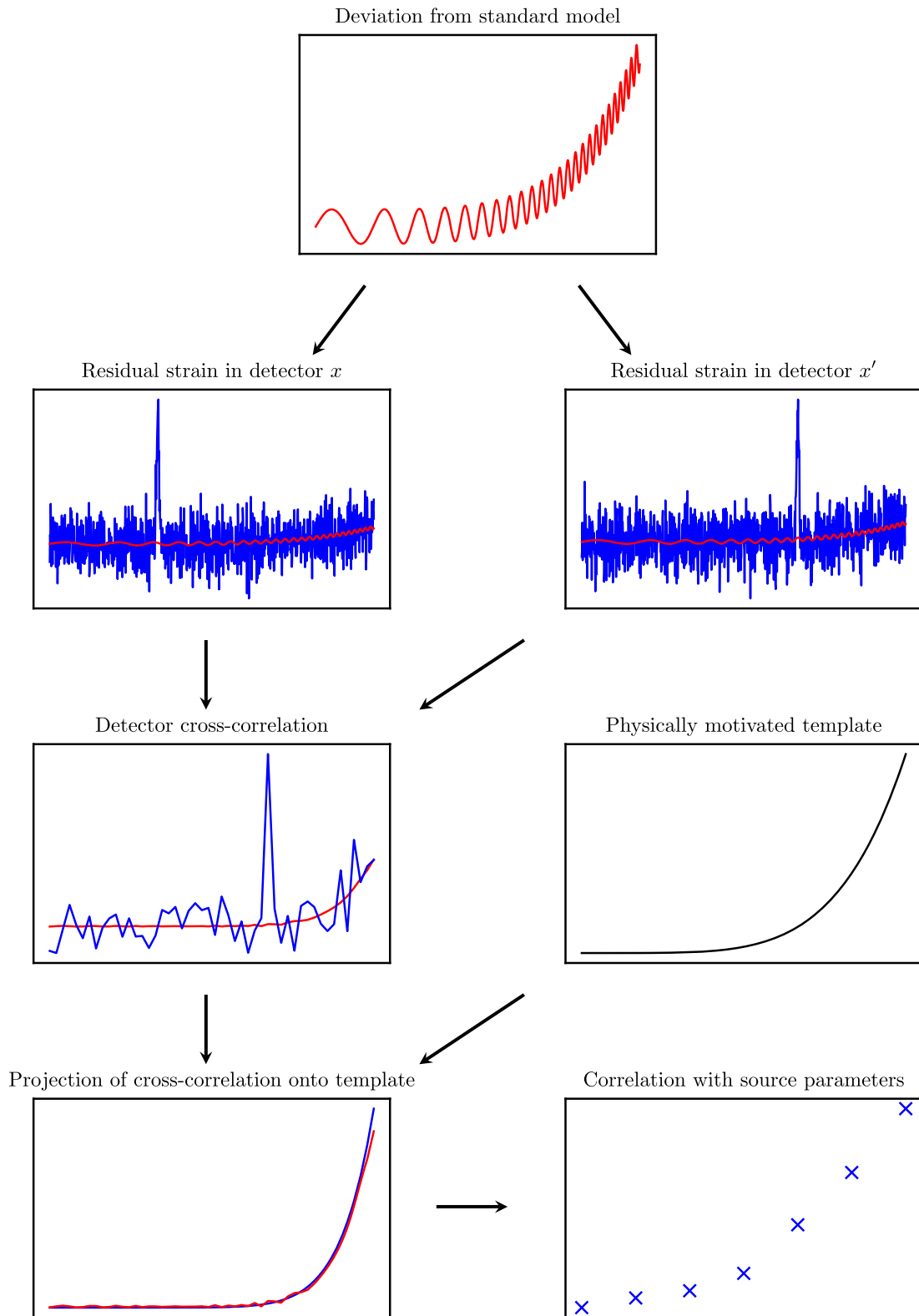
Correlation with source parameters

FIG. 1.   Outline of the SCoRe method. When a BM signature (in red) is measured by a pair of GW detectors, each detector will add onto it a different noise. The resulting residual strains, computed by subtracting the best-fit SM waveform from the data, are plotted in blue. Noise that is uncorrelated between detectors will be reduced when taking the cross-correlation between the two detectors. Higher-power transient noise and correlated noise may remain, but can be mitigated by projecting CRPS onto some physically motivated template. Most BM signature models have a dependence on the source parameters. The projection coefficients measured over events of different source parameters may reveal such correlation.

and, more specifically, the cross-correlation of the residual strains:

$$
\begin{aligned}
D^{xx'}(t) &\equiv \langle \Delta d^x_{\theta_{\mathrm{MLE}}}(t) \Delta d^{x'}_{\theta_{\mathrm{MLE}}}(t + \Delta t) \rangle, \\
&= \langle \Delta s^x_{\theta_{\mathrm{MLE}}}(t) \Delta s^{x'}_{\theta_{\mathrm{MLE}}}(t + \Delta t) \rangle + \langle n^x(t) n^{x'}(t + \Delta t) \rangle.
\end{aligned}
\tag{5}
$$

The value of Δt will depend on the source parameters such as the sky position and also on the position of the two GW detectors x and x' between which the signal is correlated. This can be inferred for individual pairs of detectors, and need not be assumed in the analysis. The detection of the correlated signal $D(t)$ depends on the overlap of the signal $\Delta s^x_{\theta_{\mathrm{MLE}}}(t)$ between the individual detectors, which depends on the response functions of the detectors $F^x_i(\mathrm{t})$. We can marginalize over the response functions and extrinsic source parameters, as explained in Sec. III D, so that we may account for different residual signatures that will be induced by the same BM deviation in different detectors. But, if, for a given antenna pattern, a theory predicts no measurable residual signature in multiple detectors, then the cross-correlation will not yield anything.

Using the cross-correlation definition given in Eq. (4), we can model the measured cross-correlated signal as

$$
D^{xx'}(t) = S^{xx'}_{\theta_{\mathrm{MLE}}}(t) + \mathcal{N}^{xx'}_c(t),
\tag{6}
$$

where the first term denotes the correlated signal between the two detectors and the second term denotes the correlated residual noise between the two detectors. We usually assume that the noise between detectors is uncorrelated. In reality, some noise, like magnetic noise, can be correlated. Though these types of noise are unlikely to be dominant, they will play a role at some level in the mean, and their impact on the signal should be estimated when this technique is applied to the data.

If the noise is uncorrelated between the two detectors, then the second term in Eq. (6) vanishes and only the first term is nonzero. As a result, we can write

$$
\begin{aligned}
D^{xx'}(t) &= \langle \Delta s^x_{\theta_{\mathrm{MLE}}}(t) \Delta s^{x'}_{\theta_{\mathrm{MLE}}}(t + \Delta t) \rangle \\
&\quad + \langle n^x(t) n^{x'}(t + \Delta t) \rangle \delta_{xx'}.
\end{aligned}
\tag{7}
$$

Alternatively, one can work in the time-frequency domain using the short-time Fourier transforms of the strains $s^x(t)$ and $s^{x'}(t)$ and substituting in Eq. (4). The timescale used to compute the short-time Fourier transform should not be greater than the timescale of cross-correlation τ. In the remaining analysis, for concreteness, we will work in the time domain.

B. Choice of residual templates

The second "key-physics"-driven aspects that we want to explore in CRPS signals is that there might be more energy injection/extraction from a binary system in BM scenarios than in the SM templates. This difference in power released results, in particular, in a change in the orbital frequency evolution $f(t)$. Previous residual tests, such as the ones using BayesWave [76,77], have proposed to decompose residual signals in terms of frames that have a natural expression in the frequency domain (Morlet-Gabor sine-Gaussian wavelets and "chirplets" [76]). Along similar arguments, we propose to decompose the CRPS signal into a template of n orbital frequency-dependent functions $Z_i(f(t))$, which can capture any gain/loss in the (radiated) energy of the binary source after taking into account the dependence on the source parameters

$$
S^{xx'}_{\theta_{\mathrm{MLE}}}(t) = \sum_{i=1}^{i=n} \alpha_i(\theta_{\mathrm{MLE}}, t) Z_i(f(t)),
\tag{8}
$$

where the (real) coefficients α_i may depend on the source parameters and time. These coefficients vanish when no BM signature is present in the data. Either side of the equation is valued in strain squared and will depend on the integration timescale τ. The only requirement on the set of templates $Z_i(f(t))$ is that its elements should not be parallel to one another. They need not be either normalized or orthogonal. For example, one could choose wavelets or chirplets. Notice that one can also model the BM signature in terms of the residual strain of the GW signal rather than the power. As one still uses the CRPS to extract posterior information in either approach, they will yield the same information. We describe the strain modeling in Appendix A.

The specific form of the template functions $Z_i(f(t))$ can be motivated by a particular type of BM signature if such knowledge is available. As an alternative to choosing a specific, model-dependent set of template functions, we present a choice of template that we argue allows for an unmodeled BM search. The unmodeled template is constructed from the following observation; BM effects modify the rate of energy dissipation. This in turn impacts the rate at which orbital frequency increases. We thus expect the derivatives of the orbital frequency of the system to be good tracers of a BM signature.

More specifically, to capture the proportional change in time of the orbital frequency, we use the derivatives of $\dot{f}/f$ as our template functions

$$
Z_{i>0}(t) = \left\langle \left(\frac{d^i \ln f(t)}{dt^i} \right)^2 \right\rangle,
\tag{9}
$$

while $Z_0(t)$ is constant, defined to capture any BM signature with constant power. The template functions take

values at bin positions. If the template functions are defined as functions of continuous time, one can simply take the angular bracket, as defined in Eq. (4), to define the expected CRPS functions.

We note that for inferring the values of α_i from data, it is convenient to orthogonalize and normalize the template functions, e.g. using the Gram-Schmidt process, if possible. If the physical interpretation of the coefficients is easier in the original form, then one can always transform the functions and their coefficients back.

C. Projection on a template

The projection of the CRPS of the data onto the chosen template encodes how well the model matches the signal. In this subsection, we define the projection coefficients and the associated uncertainty. The projection coefficient $\Gamma_i(D)$ of the CRPS signal $D^{xx'}(t)$ onto the template Z_i is

$$\Gamma_i(D) = \int_{t_s}^{t_e} dt D^{xx'}(t) Z_i(f(t)), \qquad (10)$$

(t_s, t_e are the start/end of the signal in the data). Then, the requirement that the template functions should not be parallel is

$$Z_i - \Gamma_i(Z_j)\tilde{Z}_i \neq 0 \; \forall \; \{j \neq i\} \in n, \qquad (11)$$

where $\tilde{Z}_i$ is normalized, so $\Gamma_i(\tilde{Z}_i) = 1$. The templates can be made orthonormal, that is $\Gamma_i(Z_i) = 1$ and $\Gamma_i(Z_j) = \delta_{ij}$, if required for the inference of the signal. If an orthonormal set is chosen, then the projection coefficient $\Gamma_i(S)$ will be equal to α_i. This is possible as long as the template functions are not linearly dependent. In the rest of this work, we will assume that the Z_i used are orthonormal and hence the value of α_i is measured as the mean value of the projection of the data onto the template functions.

When measuring the CRPS and the coefficients, we will need to know the associated uncertainty. In this work, we compute the standard deviation in the limit of small-signal (zero mean value). In the limit where the cross-correlation scale goes to zero, $\tau \to 0$, the noise on the cross-correlation estimator can be obtained as

$$K_n^2 \equiv \langle (\Delta d^x(t)\Delta d^{x'}(t + \Delta t))^2 \rangle - D(t)^2. \qquad (12)$$

Assuming that the noise in each detector is Gaussian, stationary, and uncorrelated, we can simplify the above expression as[4]

$$\begin{aligned} K_n^2 &\equiv \langle n^x(t)n^x(t) \rangle \langle n^{x'}(t)n^{x'}(t) \rangle, \\ &= N_x^2 N_{x'}^2, \end{aligned} \qquad (13)$$

---

[4]The angular brackets are defined in Eq. (4).

where in the second line we have defined the noise autocorrelation of each detector as N_x^2.

If there are no BM signatures present in the data, then the value of the coefficients α_i should be consistent with zero. However, in the case of a nonzero signal, we can write the estimator $\hat{\alpha}_i$ from the data. Assuming α_i does not depend on time, then we can write it in terms of the CRPS $D(t)$, the template functions $Z_i(f(t))$, and the noise on the CRPS, K_n, as

$$\hat{\alpha}_i^{xx'}(\theta_{\mathrm{MLE}}) = \int_{t_s}^{t_e} dt W_i(t) D^{xx'}(t) Z_i(f(t)),$$

$$\text{where } W_i(t) \equiv \frac{K_n^{-2}(t)}{\int_{t_s}^{t_e} K_n^{-2}(t) Z_i^2(f(t)) dt}. \qquad (14)$$

In the above equation, $W_i(t)$ are weights which serve two purposes: (i) they ensure that the equation is normalized, and (ii) they account for any variation in the noise properties in the detector by inverse noise weighting. This helps in reducing the uncertainty in the estimator. For the situation of uncorrelated stationary Gaussian noise considered in this analysis, the values of $W_i(t)$ will be constant and will simply normalize $Z_i(t)$. However if the noise power spectral density (PSD) across a data segment shows variation with time $N_x^2(t)$ (nonstationarity), then $W_i(t)$ will not be a constant. If the $Z_i(f(t))$ are not orthogonal, $\hat{\alpha}_i$ will generally be nonzero even when $D(t) = Z_j(t)$, for $j \neq i$.

Finally, it is of interest to estimate the signal-to-noise ratio (SNR) for measuring the parameters $\hat{\alpha}$. For N_{GW} sources with GW similar source parameters and with N_{det} GW detectors, we can write the combined SNR, ρ_i, as

$$\rho_i = \bar{\alpha}_i(\theta_{\mathrm{MLE}}) \times \left[\sum_{x=1,x'>x}^{N_{\mathrm{det}}} \sum_{j}^{N_{\mathrm{GW}}} (K_{\alpha_i}^{xx'})^{-2} \right]^{1/2}, \qquad (15)$$

where, $\bar{\alpha}_i(\theta_{\mathrm{MLE}})$ is the mean value of the inferred signal for all the detectors pairs and sources, and $(K_{\alpha_i}^{xx'})^2$ is the noise on the parameter α_i, defined as

$$(K_{\alpha_i}^{xx'})^2 = \int_{t_s}^{t_e} dt (W(t) K_n Z_i(t))^2. \qquad (16)$$

From the expression given in Eq. (15), we can conclude that the total SNR ρ to measure a deviation α_i scales as $\sqrt{N_{\mathrm{GW}}}$ and $\sqrt{N_{\mathrm{det}}(N_{\mathrm{det}} - 1)/2}$. Finally it is important to clarify that the value of α can explicitly depend on the value of the source parameters $\{\theta_{\mathrm{MLE}}\}$, and in that case so will the number of GW sources $N_{\mathrm{GW}}(\theta_{\mathrm{MLE}})$ with the best-fit source parameter θ_{MLE} that needs to be combined to obtain the SNR.

D. Inference using the Bayesian framework

In this section we formulate a hierarchical Bayesian framework to measure the presence of any BM signature in the observed data by marginalizing over the uncertainties associated with the GW source parameters. Let us denote $\mathcal{M}_{\mathrm{SM}}$ as the SM, and Δ_M as a set of parameters which captures deviations from this model. Then, we can write the posterior distribution $\mathcal{P}(\Delta_M|\{d^x(t)\})$ on the parameters Δ_M given the observed set of data $\{d^x(t)\}$ detected at detectors labelled x using Bayes' theorem [81] as

$$\mathcal{P}(\Delta_M|\{d^x(t)\}) \propto p(\{d^x(t)\}|\Delta_M)\Pi(\Delta_M), \quad (17)$$

where $p(\{d^x(t)\}|\Delta_M)$ denotes the likelihood and $\Pi(\Delta_M)$ denotes the prior. For a set of N_{obs} independent events of GW sources detected (denoted by $\mathcal{S}$) above a matched filtering SNR ρ_*, we can simplify the likelihood as

$$
\begin{aligned}
&p(\{d^x(t)\}|\Delta_M) \\
&= \prod_{x,x'>x=1}^{N_{\mathrm{det}}} \prod_{i=1}^{N_{\mathrm{obs}}} p(\{d^x,d^{x'}\}_i|\mathcal{S},\Delta_M) \\
&= \prod_{x,x'>x=1}^{N_{\mathrm{det}}} \prod_{i=1}^{N_{\mathrm{obs}}} \frac{p(\mathcal{S}|\{d^x,d^{x'}\}_i,\Delta_M)p(\{d^x,d^{x'}\}_i|\Delta_M)}{p(\mathcal{S}|\Delta_M)},
\end{aligned}
\quad (18)
$$

where in the third line, the first term denotes the probability of detecting events in the data given a small variation in the model Δ_M, and the second term $p(\{d^x,d^{x'}\}_i|\Delta_M)$ is the likelihood. We can write the likelihood including the GW source parameters θ with a prior $\Pi(\theta)$, and the dependence of any deviation on the GW source parameter, $p(\Delta_M|\theta)$, as

$$p(\{d^x,d^{x'}\}_i|\Delta_M) = \int d\theta\, p(\{d^x,d^{x'}\}_i|\theta,\Delta_M)p(\Delta_M|\theta)\Pi(\theta). \quad (19)$$

The term in the denominator in Eq. (18) is the evidence, which we can write in terms of the all possible events which are detectable above a matched filtering SNR ρ_* as

$$
\begin{aligned}
&p(\mathcal{S}|\Delta_M) \\
&= \int_{\rho \geq \rho_*} d\{d^x(t)\} \int d\theta\, p(\{d^x(t)\}|\theta,\Delta_M)p(\Delta_M|\theta)\Pi(\theta).
\end{aligned}
\quad (20)
$$

If the detector noise is not changing with time, then the term $p(\{d^x(t)\}|\theta,\Delta_M)$ will not vary with time and only depend on the detectability of an event above ρ_*, which we usually denote by $p_{\mathrm{det}}(\theta,\Delta_M)$. If we assume, any signal with a matched filtering SNR above ρ_* is detected, then Eq. (18) further simplifies. The term $\int p(\mathcal{S}|\{d^x,d^{x'}\}_i,\Delta_M)$ becomes unity. Putting all of this together, we can write the posterior on the parameters Δ_M as

$$
\begin{aligned}
&\mathcal{P}(\Delta_M|\{d^x(t)\}) \\
&= \Pi(\Delta_M) \prod_{x,y>x=1}^{N_{\mathrm{det}}} \prod_{i=1}^{N_{\mathrm{obs}}} \frac{\int p(\{d^x,d^{x'}\}_i|\theta,\Delta_M)p(\Delta_M|\theta)\Pi(\theta)d\theta}{p(\mathcal{S}|\Delta_M)}.
\end{aligned}
\quad (21)
$$

This is the most general hierarchical Bayesian framework to search for a BM signature marginalizing over the source parameters uncertainties. In this analysis, we perform the Bayesian estimation on the residual signal obtained around a set of best-fit model parameters θ_{BF} assuming $\mathcal{M}_{\mathrm{SM}}$ as the SM. This allows us to define the deviation Δ_M as a function of θ_{BF} so we may check any correlation between BM signature and source parameters. Searches around the best-fit can also help in reducing the computational cost in performing the Bayesian analysis. However, the analysis pipeline can be easily modified to include the full Bayesian framework. The above equation around a best-fit value can be simplified into

$$\mathcal{P}(\alpha|\{d(t)\}) = \Pi(\alpha) \prod_{x=1,y>x}^{N_{\mathrm{det}}} \prod_{i=1}^{N_{\mathrm{obs}}} \frac{\int p(\{d^x,d^{x'}\}_i|s(\theta_{\mathrm{MLE}}),\alpha)p(\alpha|\theta_{\mathrm{MLE}})p(\theta_{\mathrm{MLE}})d\theta_{\mathrm{MLE}}}{p(\mathcal{S}|\alpha)}, \quad (22)$$

where, $s(\theta_{\mathrm{BF}})$ is the GW signal for the set of best-fit parameters, $p(\theta_{\mathrm{BF}})$ is the posterior on the best-fit parameters obtained from the data assuming the SM, and α is the set of template coefficients for Δ_M. The likelihood $p(\{d^x,d^{x'}\}_i|s(\theta_{\mathrm{BF}}),\alpha)$ can be written in terms of the CRPS as

$$p(\{d^x,d^{x'}\}_i|s(\theta_{\mathrm{BF}}),\alpha) = p(D(t)|\alpha). \quad (23)$$

If we further assume the likelihood to be Gaussian and there is no correlated noise, then it becomes

$$p(\{d^x,d^{x'}\}_i|s(\theta_{\mathrm{BF}}),\alpha) \propto \exp\left(-\int dt \frac{(D(t)-S_{\theta_{\mathrm{MLE}}}(\alpha))^2}{2K_n^2(t)}\right), \quad (24)$$

where $D(t)$ can be calculated using Eq. (7) and the noise covariance matrix can be calculated using Eq. (12). In the

presence of correlated noise $\mathcal{N}_c$ it can be included in the noise covariance matrix.

Though this method can detect the presence of any kind of BM signature in the GW data, model-independent searches may fail to completely capture the structure of the deviation if the chosen template only partially overlaps with the BM signature. However, using the Bayesian framework proposed here, one can do a Bayesian model comparison for different BM scenarios and perform a tuned search for the model with higher Bayesian evidence.

IV. APPLICATION OF SCORE ON A TOY EXAMPLE FOR AN UNMODELED SEARCH

We now illustrate the SCoRe method by using the unmodeled power template to recover a toy model injection from simulated data. We describe the toy model, then discuss the appropriate choice of cross-correlation time-scale. Using the SCoRe framework, we finally recover the injected value in the simulated data and perform the Bayesian analysis described in Sec. III D.

A. Generating mock data

As a toy model on which to apply this template, we create simulated mock data by adding Gaussian noise realizations n^x over a simulated strain $s(t)$ as

$$d^x(t) = s(t) + n^x(t). \tag{25}$$

The strain is $s(t) \equiv s_{SM}(t) + \Delta s_{BM}$; it includes a standard model signal $s_{SM}(t)$ and a BM signal Δs_{BM}. The BM signal is modeled as $\Delta s_{BM}(t) = b \frac{d \ln f_{SM}(t)}{dt}$, with b as a free parameter that controls the strength of the BM signal. The standard model waveform is the same for all events. It is computed using LALSuite [82] called through PyCBC [83] and the IMRPhenomD approximant [84,85] for an equal mass, nonspinning, circular BBH merger, at a luminosity distance of 100 Mpc and the individual masses are both set to $m = 5M_\odot$. A sampling rate of 16384 Hz is used throughout this work. Both cross- and plus-polarizations are used to derive f_{SM}. The plus-polarization is used as the standard model signal s_{SM}. We assume Gaussian, stationary noise and the same sensitivity for all detectors. Different events in each of the two distinct detectors are then simulated by drawing realizations n^x of the expected O4 noise sensitivity [86].[5]

We project the CRPS onto the unmodeled template given in (9). In the rest of the work, we implement the angular bracket average of Eq. (4) as a binned mean.[6] Out of the template terms given in Eq. (9), we only use the linear

---

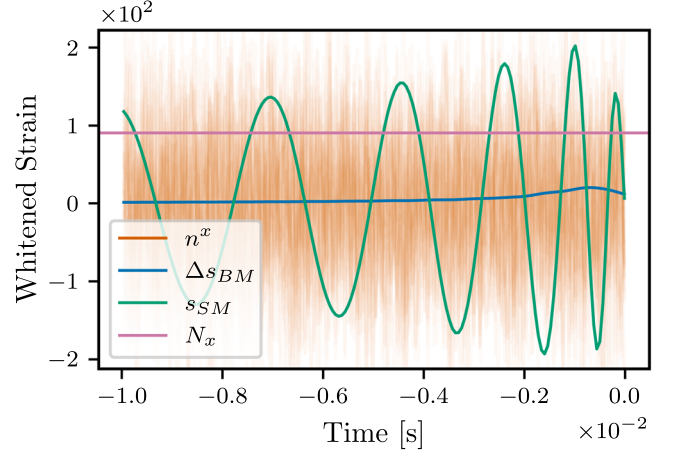[6]One could also use a running mean, instead of a binned mean value.

FIG. 2. Example toy model data. We use the PyCBC software package to generate s_{SM}, the waveform for a circular, non-spinning, equal mass BBH, with individual masses both equal to $5M_\odot$. This is plotted in green. We add onto it a BM signature, Δs_{BM} (blue line), that is proportional to the change in the orbital frequency logarithm ($\tilde{\alpha}_1 = 0.05$). It is normalized so that, for $\tilde{\alpha}_1 = 1$, the maximum amplitude reaches to the noise auto-correlation, N_x (purple line). Realizations of Gaussian, stationary noise are added to $s_{SM} + \Delta s_{BM}$ to obtain different events. Some examples of noise realizations are plotted in orange.

term Z_1.[7] The toy model with the BM signal Δs_{BM} is constructed such that it perfectly overlaps with the template Z_1 to show how the recovery can happen for the best-case scenario. We expect that the template coefficient α_1 can recover the injected BM signal Δs_{BM} completely. Certainly, in reality different BM theories will only have a partial projection on the template, unless one performs a model-dependent search. In the future, we will explore different choices of template to illustrate how they can project onto different BM theories. Theoretical efforts to produce waveforms in some such theories have been presented in e.g. [74,87] and so have their impact data analysis in modeled and unmodeled searches e.g. [9,88].

We generate N_{GW} pairs of strain data as in Eq. (25) and attribute each of these strains to one of the two detectors x or x'. Examples of each of the components used to construct these data are shown in Fig. 2. We assume that the best-fit parameters are known exactly, and that they are those used to generate $s_{SM}(t)$. We therefore subtract s_{SM} from each of these strains to obtain $\Delta d^{x,x'}$. In real data, the error from estimating θ_{MLE} needs to be accounted for both in the residual strains and in the template and we need to perform the marginalization described in Sec. III D.

Each pair of $\Delta d^{x,x'}$ is then cross-correlated as in Eq. (5) to obtain the CRPS, $D_i^{x,x'}$, where i labels events. We obtain N_{GW} residual time series that, for real data, would each be

---

[7]Since we are only using one template function, our template is orthogonal by construction.
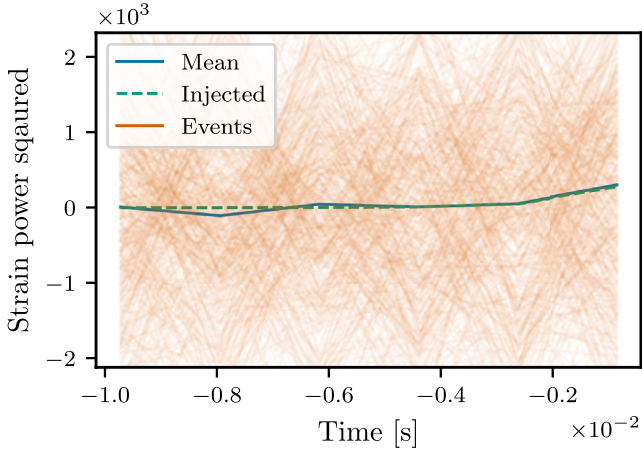
FIG. 3. Residual data and injected BM signature cross-correlated over $\tau = 1.28 \times 10^{-3}$ s. The blue line is the mean of the CRPS over 500 events, plotted in orange. It corresponds to the MLE value of $\tilde{\alpha}_1$ when projected onto the template. The dashed green line is the cross-correlation of Δs_{BM}.

associated with a GW event. Examples of the CRPS are shown by orange lines in Fig. 3. An estimator $\hat{\alpha}_1$ is obtained for each of these time series by projecting onto the template according to Eq. (14). It is convenient to define the dimensionless quantities

$$\tilde{Z}_1 = P^{-1}Z_1, \quad \text{such that} \quad \int_{t_s}^{t_e} \tilde{Z}_1^2 dt = 1,$$

$$\tilde{\alpha}_1 = \frac{\alpha_1}{\alpha_1^0}, \qquad \alpha_1^0 = \frac{PN_x^2}{\max\left(\frac{d\ln f}{dt}\right)^2}, \qquad (26)$$

where α_1^0 is defined so that, when the maximum amplitude of Δs_{BM} is N_x, the projection of its cross-correlation onto $\tilde{Z}_1$ is $\tilde{\alpha}_1 = 1$. The normalization constant, P, is a measure of the mean power per bin in the template function. During this procedure, the timescale τ over which the residual strains are cross-correlated is a free parameter. We discuss the choice of τ in the next subsection.

B. Choice of τ

When τ is finite, the noise on the cross-correlation estimator, in the limit of a small signal is

$$K_D^2 = \frac{1}{n_{\text{window}}} K_n^2, \qquad (27)$$

where n_{window} is the number of data points measured in the timescale τ. The advantage of taking the mean over a timescale τ is that $K_D \propto 1/\tau$, so, in the limit of large τ, the cross-correlation tends to be the residual signal (and correlated noise). On the other hand, as mentioned, some information is lost in the averaging over the timescale τ. This is illustrated in Fig. 4. The cross-correlation template
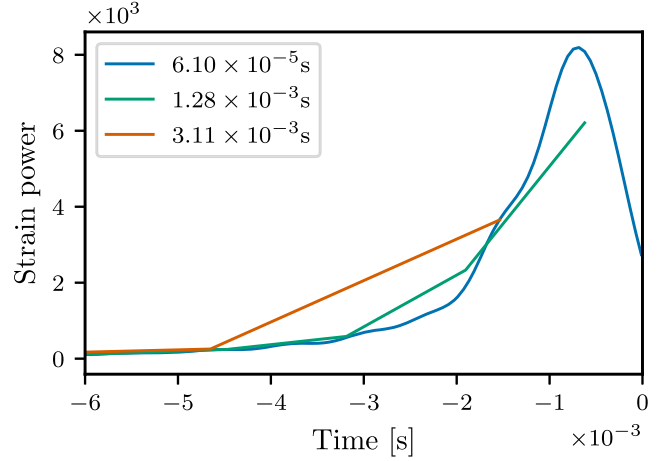
associated with a GW event. Examples of the CRPS are shown by orange lines in Fig. 3. An estimator $\hat{\alpha}_1$ is obtained for each of these time series by projecting onto the template.

correlation template $\tilde{Z}_1$ for different cross-correlation timescales τ. The smallest timescale allowed by the sampling rate, $\tau = 6.10 \times 10^{-5}$ s, has features that are washed away as the timescale is increased. With $\tau \in [0.8, 1.6] \times 10^{-3}$ s, no new physical features appear and information remains the same as τ is increased. For timescales longer than 2.1×10^{-3} s, all the power is effectively concentrated in a single bin.

$\tilde{Z}_1$ is computed from s_{SM} with different values of τ. As τ is increased, the shape of $\tilde{Z}_1$ is qualitatively changed. By integrating over τ, we lose the ability to differentiate between BM models that give different template predictions below this timescale.

For a specific template, two choices of τ have equivalent signal content if no new physical behavior arises at an intermediate timescale. If they are equivalent, the average power per bin, P, will be the same. This quantity is plotted in Fig. 5 as a function of τ. For this example, at timescales
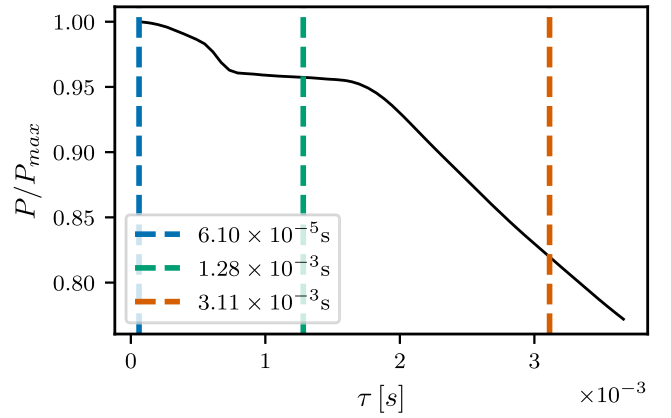

FIG. 4. Cross-correlation template $\tilde{Z}_1$ for different cross-correlation timescales τ. The smallest timescale allowed by the sampling rate, $\tau = 6.10 \times 10^{-5}$ s, has features that are washed away as the timescale is increased. With $\tau \in [0.8, 1.6] \times 10^{-3}$ s, no new physical features appear and information remains the same as τ is increased. For timescales longer than 2.1×10^{-3} s, all the power is effectively concentrated in a single bin.

FIG. 5. Template normalization factor (mean power per bin P) as τ is increased. The template with $\tau = 6.10 \times 10^{-5}$ s contains a pulse. As the pulse is washed away by the averaging, the mean power per bin decreases. Between $0.8 \times 10^{-3} s < \tau < 1.6 \times 10^{-3} s$, averaging over a greater τ does not change the shape of the template and the mean power remains constant. At scales greater than 2.1×10^{-3} s a single bin contains the majority of the power and the mean power scales as $\tau^{-1/2}$.

shorter than 0.8×10^{-3} s, the mean power decreases as the "pulse" feature of the template seen in Fig. 4 is integrated out. For $\tau \in [0.8, 1.6] \times 10^{-3}$ s, the mean power remains constant as the template functions resemble power laws and a smaller τ does not add information. An example of a template with a timescale in this range is shown by the green line in Fig. 4 (obtained by averaging over $\tau = 1.28 \times 10^{-3}$ s). When averaging over timescales larger than $\tau \sim 2.1 \times 10^{-3}$ s, all the power of the BM signature is concentrated in a single bin. Past this point, the mean power per bin, P scales as $(\tau)^{-1/2}$.

This exact evolution sequence of the mean power per bin P as a function of τ is specific to our chosen template (9) and the standard model waveform used for the computation. However, the criteria that can be used to choose τ will be the same in all cases.

The first natural choice of τ corresponds to the maximum SNR, which is achieved when most of the power is concentrated in as small a bin as possible. This is determined by the τ where the mean power per bin, P, starts scaling as $\tau^{-1/2}$. This may not, however, correspond to the highest Bayes factor in favor of a BM.

If information about the structure of the BM signal is preferred over a higher SNR, then τ should be accordingly chosen. In our toy model, for our specific choice of template and SM waveform, to conserve information about the "chirp" part of the template, the τ value needs to be set to the rightmost point on the horizontal section of the P shown in Fig. 5, $\tau = 1.8 \times 10^{-3}$ s. In this case, it is not possible to preserve information about the pulse shape while still taking the cross-correlation, as one would need to set τ to the sampling interval. It may yet be possible to keep some benefits of taking the cross-correlation while not averaging over the pulse shape by letting the averaging timescale vary with time. For example, we may set τ to 1 over the sampling rate from $t = -10^{-3}$ s to $t = 0$, and set $\tau \in (0.8, 1, 6) \times 10^{-3}$ s for $t \in (-5, -1) \times 10^{-3}$ s to maximize noise reduction over the "chirp". Finally, we could average the rest of the time domain, where the CRPS is dominated by noise, into one bin $[t = (t_s, -5) \times 10^{-3}$ s].

In this section, we consider only a time-constant τ, which we set to the SNR-maximizing value, $\tau = 1.8 \times 10^{-3}$ s. The cross-correlations plotted in Fig. 3 are computed over this timescale. This choice of τ is specific to this template function and best-fit source parameters, as the variation of the mean power P with τ will change for other choices of template and source parameters. Since the templates depend on the orbital frequency, we expect the choice of τ to vary (scale) with chirp mass for all choices of templates.

C. Bayesian inference of the injected signal

In this section, we use the Bayesian framework presented in Sec. III D to combine events into one measurement of $\tilde{\alpha}_1$.

We assume the source parameters are perfectly known to be $\tilde{\theta}_{\mathrm{BF}}$. Therefore, $p(\tilde{\theta}_{\mathrm{BF}}) = \delta(\theta_{\mathrm{MLE}} - \tilde{\theta}_{\mathrm{BF}})$, where $\delta(x)$ is the Dirac delta function. This assumption is made to show how this new method will work in the best-case scenarios. In reality, each estimated parameter will have an error, and one can marginalize the uncertainty as presented in Sec. III D. Furthermore, since we assume α_1 is unique for the given source parameters, $p(\alpha_1|\tilde{\theta}_{\mathrm{BF}}) = 1$. The posterior on $\tilde{\alpha}_1$ is then

$$\mathcal{P}(\tilde{\alpha}_1|\{d(t)\}) \propto \Pi(\tilde{\alpha}_1) \exp\left(-\frac{1}{2\tilde{K}_{\tilde{\alpha}}^2}(\tilde{\alpha}_1 - \tilde{\alpha}_{1,\mathrm{MLE}})^2\right), \quad (28)$$

where $\tilde{\alpha}_{1,\mathrm{MLE}} = 1/(N_{\mathrm{GW}}\alpha_1^0)\sum_i^{N_{\mathrm{GW}}}\int_{t_s}^{t_e}\tilde{Z}_1 D_i dt$, and

$$\tilde{K}_{\tilde{\alpha}_1}^2 = \frac{K_{\alpha_1}^2}{(a_1^0)^2 N_{\mathrm{GW}}}. \quad (29)$$

The likelihood is thus a Gaussian distribution with maximum likelihood estimator (MLE) $\tilde{\alpha}_{1,\mathrm{MLE}}$ and variance $\tilde{K}_{\tilde{\alpha}_1}^2$. With our assumption, the MLE is the mean over events of the estimator $\hat{\alpha}_1$. Since the projection is linear, it corresponds to the projection of the mean CRPS, shown in blue in Fig. 3.

If a signal is correlated between detectors, its cross-correlation $S^{xx'}(t)$ must be positive. Our template function is positive, therefore so must be α_1 and $\tilde{\alpha}_1$. We choose a flat prior on $\tilde{\alpha}_1$ ranging from 0 to $\tilde{\alpha}_{1,\mathrm{max}}$, so that $\Pi(\tilde{\alpha}_1) = (\tilde{\alpha}_{1,\mathrm{max}})^{-1}$. The posterior obtained from the events in Fig. 3, which have injection $\tilde{\alpha}_1 = 0.05$, is shown is shown in Fig. 6. The prior used is set with $\tilde{\alpha}_{1,\mathrm{max}} = 1$.

With this posterior, we can compare the standard model hypothesis with the hypothesis of a BM signature by computing $\mathcal{B}_{\mathrm{SM}}^{\mathrm{BM}}$, the Bayes factor in favor of a BM signature being present as opposed to just the SM.
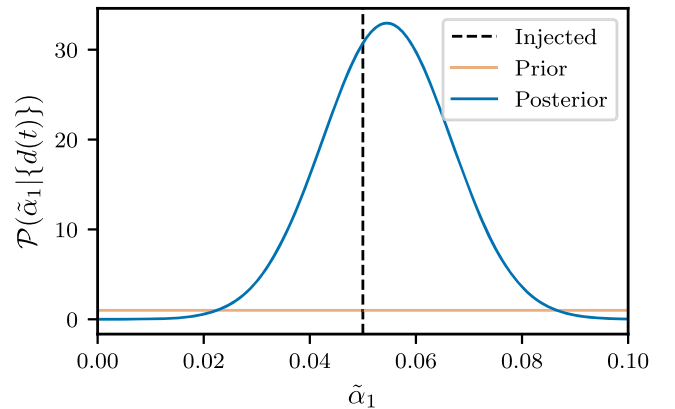

FIG. 6. Posterior on $\tilde{\alpha}_1$ for $N_{\mathrm{GW}} = 500$, when the template function Z_1 is injected in Gaussian, stationary noise. We used a flat prior from 0 to 1, giving a Bayes factor of 773. The error on $\tilde{\alpha}$ is $\tilde{K}_{\tilde{\alpha}}$, the SNR for this measurement is 4.5.
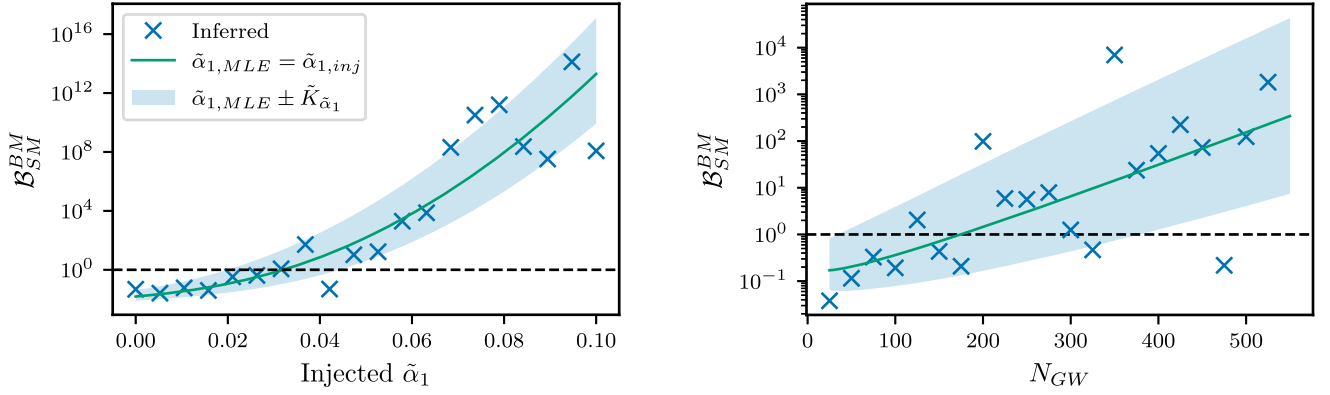
FIG. 7. Bayes factor $\mathcal{B}_{SM}^{BM}$ in favor of a BM hypothesis over the standard model hypothesis *left* for different values of $\tilde{\alpha}_{1,inj}$ ($N_{GW} = 500$), *right* for different N_{GW} ($\tilde{\alpha}_{1,inj} = 5 \times 10^{-2}$). The solid green line indicates the case of perfect accuracy as a guide. The blue region covers one $\tilde{K}_\alpha$ deviations in the measured value of $\tilde{\alpha}_{1,MLE}$. The standard model waveform used is that of an equal mass BBH with individual mass $5M_\odot$.

Since the standard model is recovered when the coefficient α_i is zero, we compute $\mathcal{B}_{SM}^{BM}$ as the Savage-Dickey density

$$\mathcal{B}_{SM}^{BM} = \frac{\Pi(0)}{P(\tilde{\alpha}_1 = 0|\{d(t)\})}, \tag{30}$$

which can be computed directly in our toy model, since the analytical form of the posterior in terms of $\tilde{\alpha}_{1,MLE}$, K_D^2, and N_{GW} is known. In fact, the expression for $\mathcal{B}_{SM}^{BM}$ can be written explicitly in terms of error functions

$$\mathcal{B}_{SM}^{BM} = \sqrt{\frac{\pi}{2}} \frac{\tilde{K}_{\tilde{\alpha}_1}}{\tilde{\alpha}_{1,max}} \exp\left(-\frac{1}{2}\left(\frac{\tilde{\alpha}_{1,MLE}}{\tilde{K}_{\tilde{\alpha}_1}}\right)^2\right)$$
$$\times \left(\text{erf}\left(\frac{\tilde{\alpha}_{1,MLE}}{\sqrt{2}\tilde{K}_{\tilde{\alpha}_1}}\right) - \text{erf}\left(\frac{\tilde{\alpha}_{MLE} - \tilde{\alpha}_{1,max}}{\sqrt{2}\tilde{K}_{\tilde{\alpha}_1}}\right)\right), \tag{31}$$

where $c = (1/2)\tilde{K}_{\tilde{\alpha}_1}^2$ and $\text{erf}(x) = (2/\sqrt{\pi})\int_0^x \exp(-y^2)dy$. This can be computed numerically. We plot the dependence on N_{GW} and on the injected value of $\tilde{\alpha}_1$ in green on Fig. 7, assuming a perfectly accurate measurement ($\tilde{\alpha}_{1,MLE} = \tilde{\alpha}_{1,inj}$). In the left plot, the number of events is kept fixed at $N_{GW} = 500$ as the injected value of $\tilde{\alpha}_1$ is varied, while on the right plot, N_{GW} is varied as $\tilde{\alpha}_{1,inj}$ is kept fixed. We also plot values of $\tilde{\alpha}_{1,MLE}$ measured from realizations of the toy-model data with crosses.

The value of $\Pi(0)$ can significantly change the recovered value of the Bayes factor. This is a manifestation of Occam's penalty [89]: a wider prior compared to the likelihood indicates a more constraining underlying hypothesis. The prior we choose has $\tilde{\alpha}_{1,max} = 1$. This maximum prior is quite large ($1 \gg \tilde{\alpha}_1 + \tilde{\Sigma}_{\tilde{\alpha}_1}$), and reducing it can increase the Bayes factor obtained without truncating most of the area under the likelihood. However, we choose it on physical grounds, as we assume any yet undetected BM signature has to be smaller than the noise variance.

V. APPLICATION OF SCORE ON A BEYOND GR MODEL

A. Toy-model EFT and numerical setup

To explore our method in a realistic scenario—including inspiral, merger and ringdown—we construct a waveform model that captures key effects expected from an effective field theory (EFT) extension of General Relativity under the assumption that the coupling scale is of the order (or below) that of the BH scale.[8] The model accounts for: (i) higher curvature corrections in the inspiraling regime can be captured at leading order by tidal effects, (ii) the merger transitions to a postmerger regime described by quasinormal modes for the black hole described by such a theory and, importantly, that corrections scale with the mass of the system appropriately. To fix ideas, and choose a convenient scaling so its dependence is sufficiently marked, we envision our system described by a Lagrangian $\mathcal{L} = R + l^4\mathcal{L}_{(6)}$, where $\mathcal{L}_{(6)}$ indicates the higher-order corrections to the curvature, scaling in this example as M^{-4} (see e.g. [91–94]). For modeling the merger, we follow [95] and construct an interpolating signal between inspiral and ringdown guided by the behavior of null geodesics corresponding to the final black hole. The main characteristics (mass and spin) of such a final black hole can be estimated using a strategy similar to that given in [96,97] as long as the general solution for a rotating black hole in a given theory is known. For the inspiral regime, we use a simple effective one-body EOB model [98] with tidal interactions as in [99] (see also [100]) transitioning to the final black hole at the innermost stable circular orbit (ISCO) frequency

---

[8]If such a scale is larger, as considered in [90], $\mathcal{O}(1)$ corrections to the GR gravitational Lagrangian are induced, rendering unclear what describes the regime with smaller scales (thus including merger and ringdown), as they would lie outside the EFT regime.

of the final black hole. For concreteness, we focus on the equal mass, nonrotating, binary black holes with no eccentricity.[9] Importantly, we note that for the theories described by the above Lagrangian, closed-form solutions are only known for slowly (or non)rotating scenarios while the merger outcome should yield a black hole with spin $a/M \approx 0.67$ with a small correction that also scales as M^{-4}. Without such a solution, the specific value of such correction can not be computed yet. However, we note that this correction is quite smaller than the correction to the horizon area (e.g. [92]) and so black hole compaction is more affected than associated quasinormal modes. Last, gravitational radiation from this theory can be captured to leading order by the correction to the quadrupole of the system; its associated effect, in turn, can be accounted for in the stationary phase approximation to the waveform as discussed in [90]. Combining the above information, we "phenomenologically" account for dynamical departures in the theory by extrapolating from the slowly rotating case with the following assignments:

$$C = \frac{1}{2} - \frac{5}{16}\epsilon, \qquad k^T = \frac{1008}{25}\epsilon, \tag{32}$$

$$\delta\omega_{\mathrm{QNM}}^R / {}^{\mathrm{GR}}\omega_{\mathrm{QNM}}^R = 1 + 0.45\epsilon, \tag{33}$$

$$\delta\omega_{\mathrm{QNM}}^I / {}^{\mathrm{GR}}\omega_{\mathrm{QNM}}^I = 1 - 2.75\epsilon, \tag{34}$$

with C the compaction of each black hole, k^T the tidal Love number, $\delta\omega$ the correction to QNMs and where $\epsilon = (M_*/M_T)^4$ and M_* a base scale which we take to be $= M_\odot = 1.5$ km. That is, the EFT is such that it assumes corrections to General Relativity, become $\mathcal{O}(1)$ at such length, thereby affecting the inspiral only at a subtle level.

B. Choice of template

We perform the same analysis as in Sec. IV on the NR waveforms motivated by the EFT. The strain signal $s(t)$ is now the full numerical waveform obtained with the source at $d_l = 100$ Mpc. We assume that the standard waveform corresponding to the best-fit parameters, $s_{\mathrm{SM}}(t)$, is the NR waveform obtained with the same source parameters in GR. These correspond to equal mass, nonspinning, quasicircular black hole binaries with individual masses given by $m/M_\odot = 1, 2$ and 5.

In Sec. IV, the BM signature introduced was directly proportional to the template function, Z_1. As described in Sec. III, more realistic residual signals such as the one obtained from the EFT-inspired waveform will require a possibly infinite number of higher-order terms to be accurately described. The first order (normalized) template

---

[9]As noted in [101] eccentricity can introduce a systematic bias in the identification of departures from GR—a potential issue that should be contemplated carefully in the analysis.

function $\tilde{Z}_1$ obtained from the GR waveforms using τ equal to the sampling interval is plotted for different m with dotted lines in Fig. 8. The power of the residual strains is plotted with full blue lines. Although the template functions do not track the short-scale time evolution, they do capture the time interval over which residual power is present.

To choose the timescale over which to cross-correlate, we note that the power templates have two peaks. Ideally, one would integrate over the smaller feature seen in the template—in this case, the narrower of the peaks in power. This would be the feature seen around −0.2 to 0 milliseconds for $m = 1M_\odot$, −0.4 to 0 milliseconds for $m = 2M_\odot$, and from −0.6 to 0.2 milliseconds for $M = 5M_\odot$. However, for $m = 1M_\odot$ and $2M_\odot$, at the maximum currently available sampling rate (16384 Hz), the residual power signals only span 2–5 sample points. We therefore choose to lose precision in favor of a higher SNR; averaging over both peaks (up to the value of τ at which the mean power per bin P scales as $\tau^{-1/2}$), which gives the highest SNR. This corresponds to integrating from around −0.25 to 0.25 milliseconds and from −0.5 to 0.5 milliseconds for $m = 1M_\odot$ and $2M_\odot$, respectively. For $m = 5M_\odot$, the larger peak is smoother. As a result, integrating over both peaks (for example, from −1 to 1 milliseconds) does not give a larger SNR. For all masses, lower values of τ lead to a decrease in SNR.

The values of τ chosen for each mass are shown in Table I. The coefficients obtained by projecting the residual power onto the template with the aforementioned values of τ are also shown in the table.

C. Recovery of injected signal

The values of $\tilde{\alpha}_1$ corresponding to the injected residual power are plotted with green crosses in Fig. 9. The values recovered from toy data with $N_{\mathrm{GW}} = 500$ are plotted with blue discs and error bars. For $m = 2$ and $5M_\odot$, the recovered values are consistent with zero in comparison to the standard deviation $\tilde{K}_{\tilde{\alpha}_1}$, as indicated by a downward arrow on the error bars.

In this particular BM model, the power of the residual signal decreases with mass as m^{-8}. Regardless of the template chosen, the projection of the residual power onto it will also scale as m^{-8}. This is the case for the injected values of $\tilde{\alpha}_1$. The dotted orange line corresponds to a fit though these points with gradient m^{-8}—the template functions correctly capture the scaling of the NR waveform residuals. The fact that the template functions correctly capture the timescale and the power scaling of the BM signature is not trivial, as they are agnostic to the model and were not informed about the nature of the EFT used to simulate the NR waveforms. This is one of the salient aspect of the method SCoRe. For $m = 2$ and $5M_\odot$, $\tilde{\alpha}_{1,\mathrm{inj}} \ll \tilde{K}_{\tilde{\alpha}}$. The Bayes factor for both these masses is inconsequential, and we cannot recover the scaling with mass.
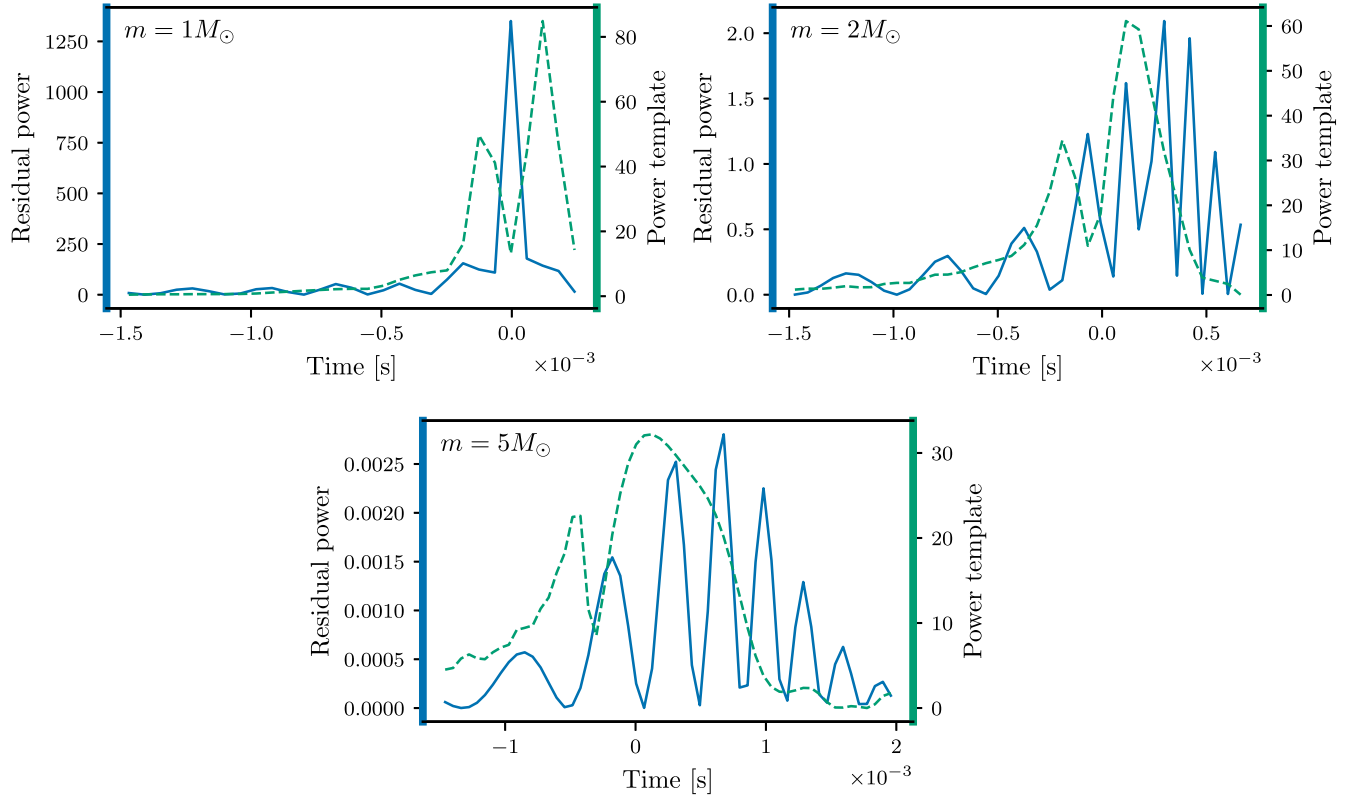
FIG. 8. Residual power and template functions. The residual signals are obtained from subtracting the GR waveforms from the EFT waveforms with the same source parameters. They are plotted with blue, solid lines (left axis). The template functions $\tilde{Z}_1$ are computed from the GR signal using a value of τ equal to the sampling interval. They are plotted with dotted green lines (right axis). The template in Eq. (9) is unmodeled, and is not expected to match actual signal at finite order, but it does capture the time interval over which the most power is present.

We have assumed that the only term scaling with the source parameter we are investigating, which is mass in this example, is due to the BM signature, so that the residual strain scales with a power k, where k captures the dependence of the BM signal, on the source parameters θ as $\Delta d \propto \theta^k$. For the EFT model considered here, $k = -4$

TABLE I. Projection coefficients of the injected BM signal and the recovered maximum likelihood estimation (MLE) of the NR EFT waveforms for 500 events. The values of τ are chosen so that the SNR is maximized and the smallest scale feature captured. They depend on the best-fit waveform. The drop in Bayes factor from $m = 1M_\odot$ to $m = 2, 5M_\odot$ is due to the mean recovered value of $\tilde{\alpha}_1$ moving within error range of $\tilde{\alpha}_1 = 0$. The FAR is computed as the probability of obtaining a value equal or larger than the $\tilde{\alpha}_1$ measured when there is no signal in the data (see Appendix B) for a Gaussian stationary noise.

$\frac{m}{M_\odot}$	τ (s)	$\tilde{\alpha}_{1,\text{inj}}$	$\tilde{\alpha}_{1,\text{MLE}}$	SNR	$\mathcal{B}_{\text{SM}}^{\text{BM}}$	FAR
1	4.27×10^{-4}	7.20×10^{-2}	1.02×10^{-1}	3.05	8.69	1.6×10^{-3}
2	1.22×10^{-3}	2.22×10^{-4}	1.66×10^{-2}	0.60	0.06	2.7×10^{-1}
5	1.22×10^{-3}	1.78×10^{-7}	5.47×10^{-3}	0.46	0.02	3.2×10^{-1}
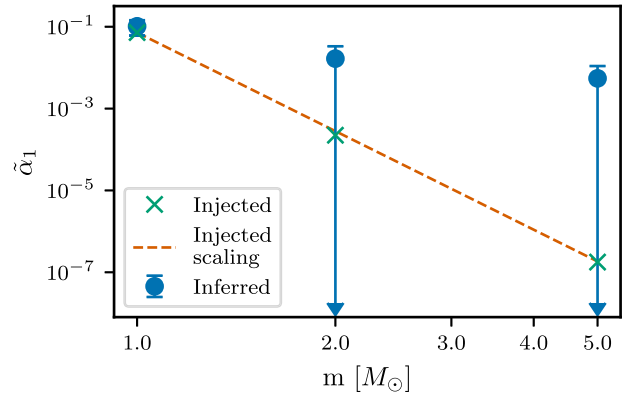
FIG. 9. Injected and recovered values of $\tilde{\alpha}_1$. The projections of the actual residual signal on Z_1 are shown with green crosses. They correctly scale as m^{-8}, shown by the orange dotted line. The blue discs are values measured from toy data constructed with 500 events. The error $\tilde{K}_{\tilde{\alpha}}$, shown by the error bars is several orders of magnitude larger than the injected values for $m = 2$ and $5M_\odot$. The measurements for these masses are consistent with zero, as indicated by the downward porting arrows, and the scaling with m cannot be recovered.

and θ is the source mass. In reality, the residual strain may contain contributions from unaccounted BM physics, waveform systematics, or a stealth bias, for example. These contributions may scale with the same source parameter as the BM signature, but with a different dependence, such as $\theta^{k'}$. Then depending on the value of k' relative to k and the range over which the value of θ is observed, one may still fail to recover the scaling with k even with large network SNR and $\tilde{K}_{\tilde{\alpha}} \ll \tilde{\alpha}_{1,\text{inj}}$. In the case of a stealth bias, we would naïvely assume $k' = -4$. This is because, as the magnitude of the BM signature changes, so will the magnitude of its projection onto the SM manifold. More complicated dependence on the source parameters of the BM signature may yet arise, as noted in [64].

VI. CONCLUSION AND FUTURE OUTLOOK

In this work we propose a new method, SCoRe, to search for BM signatures in GW data by projecting the cross-correlated residual power spectrum signal on a theory-independent or theory-dependent template, which can capture a large class of BM scenarios and its dependence on the GW source parameters.

As one of the key features of BM signatures is that they will be universal in all the detectors, cross-correlating between different pairs helps mitigate contamination from uncorrelated noise artifacts in the BM signal. Correlated and strong transient noise, as well as any unmodeled BM physics, will not be reduced by cross-correlation. To account for this, the second step of the SCoRe method is to project the cross-correlated residual power onto a template capturing a specific frequency dependence. This frequency dependence can be tuned for both modeled and unmodeled searches. An expected facet of BM signatures, however, is that they can be understood as a change in the orbital energy radiated by the coalescing binary. As an alternative to modeled searches, we therefore also propose to search for dependence on the derivatives of the logarithm of orbital frequency as a tracker of energy radiated per orbit.

Many BM signatures are expected to vary as a function of the source parameters. The final step of the SCoRe method is to check for correlation between the projection coefficients measured and the values of the best-fit parameters used to compute the residuals. We incorporated a hierarchical Bayesian framework for the marginalization of source parameters (around a best-fit value) of multiple events of the cross-correlation signal obtained by combining all possible pairs of GW detectors.

For the unmodeled searches, the template may not be able to completely capture the deviation from the standard model, however, it can indicate signatures of possible deviations from the standard model. To understand which BM scenarios best explain the data, one can do a Bayesian model comparison. This can be incorporated into the framework proposed in this work.

Following that a model-dependent search using the template choice which is tuned for the BM scenario which shows maximum Bayesian evidence in the unmodeled search can be performed.

To illustrate our method, we have applied it to a toy model and also to waveforms motivated by an EFT of gravity [91–94]. In both cases, the injected residual power was mainly concentrated in the last few milliseconds (before the peak of the waveform) of a signal and the noise was Gaussian, stationary, and colored according to the expected O4 PSD.

For the toy model, we showed how well different signal strengths could be recovered, as measured by the Bayes factor in favor of the presence of a BM signature, for different levels of power injected and number of events combined. The events used all had the same source parameters. Crucially, they were all at a luminosity distance d_l of 100 Mpc. In reality, sources are expected to be distributed over a large range of luminosity distances, in which case the residual power will scale as d_l^{-2}. In addition, since the properties of the source and their astrophysical population will critically impact the Bayes factors recovered, we will work next on taking into account astrophysical populations to obtain more realistic estimates on how well the method can infer the presence and structure of BM signatures.

The illustration of the method on the EFT-motivated NR waveforms constrained the BM signatures around zero and could not detect the injected signatures for the masses giving the smallest residual power ($m = 2, 5 M_\odot$). However, the model-independent template we suggest for agnostic searches was successful in capturing the time range over which most of the residual power was emitted. It could also correctly infer the minute BM signal for $m = M_\odot$. As this search is performed using a model-independent template, it could not reproduce the exact structure of the residual signal, as expected. In this scenario, we suggest that, once one has identified a promising result using the model-independent search, the templates should be refined. This can either be done by using a model-dependent templates or by adding higher-order derivatives to the model-independent template functions.

This work has been to present the main lines of the SCoRe method. One of the main caveat of our results is that they assume Gaussian, stationary noise. Although these assumptions are used in the LVK analyses [2–4], we are working on applying this method on the GW data available to investigate the effect of realistic noise, and what the FAR for different templates may be. We will also consider, in future work, how transient noise (glitches) affects the recovery of BM signals from data.

We hope that SCoRe will not only be useful to search for signatures of new physics, but also a large class of scenarios within the standard model ranging from unmodeled GR

effects to deviation in the waveform due to waveform modeling systematics.

We have illustrated the method with toy models in this work and will, in future works, provide tests and predictions for its applications. We will explore the capability of this technique in recovering different scenarios and the corresponding appropriate template banks which can be useful to search for both modeled and unmodeled searches.

APPENDIX A: MODELING RESIDUAL STRAIN

In Sec. III B, we have discussed the template constructed to recover signal from the cross-correlation power spectrum. In this section, we will briefly describe the template that can be used to model any deviation on the strain. Instead of modeling the residual power, $\mathcal{S}_\theta(t)$, we can model the residual strain Δs_θ. In that case, the above equations written for $\mathcal{P}_\theta(t)$ will only get a minor modification. The decomposition of the signal in terms of template functions [Eq. (8)] becomes

$$\Delta s_{\theta_{\mathrm{MLE}}}(t) = \sum_{i=1}^{i=n} \beta_i(\theta_{\mathrm{MLE}})\zeta_i(f(t)), \qquad (A1)$$

where the coefficients βi and bases ζ_i give α_i and Z_i when cross-correlated using Eq. (4). In this case, both sides of the equation still depend on the timescale τ, but they take values in strain rather than strain power.

In analogy to Eq. (14), we can write the estimator for β as

$$\hat{\beta}_i(\theta_{\mathrm{MLE}}) = \int_{t_s}^{t_e} dt W_i(t) D(t)^{1/2} \zeta_i, \qquad (A2)$$

with $W_i(t) = K_n^{-1}(t)/\int_{t_s}^{t_e} K_n^{-1}(t)\zeta_i(t)^2 dt$. The corresponding SNR is

$$\rho_i^2 = \sum_{x=1,x'>x}^{N_{\mathrm{det}}} \sum_j^{N_{\mathrm{GW}}} \left(\frac{(\hat{\beta}_i^{xx'}(\theta_{\mathrm{MLE}}))^2}{(K_{\beta_i}^{xx'})^2}\right)_j, \qquad (A3)$$

where, $(K_{\beta_i}^{xx'})^2$ is the noise on the parameter β is defined as

$$(K_{\beta_i}^{xx'})^2 = \int_{t_s}^{t_e} dt(W_i(t)K_n^{\frac{1}{2}}\zeta_i(t))^2, \qquad (A4)$$

where there is now a square root on K_n.

Again for a Gaussian, stationary, uncorrelated noise, we will get a $\sqrt{N_{\mathrm{GW}}}$ enhancement in the SNR by combining N_{GW} events and number of detector pairs $\sqrt{N_{\mathrm{det}}(N_{\mathrm{det}}-1)/2}$.

APPENDIX B: FALSE ALARM RATE

The false alarm rate (FAR) is a measure of the probability of detecting a signal when it is absent from the data. In the
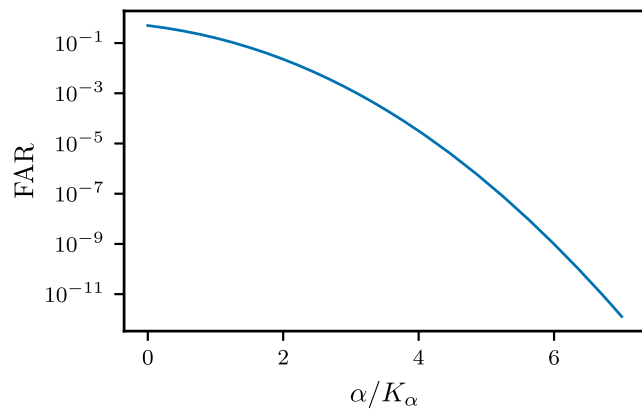
FIG. 10. FAR on the measurement of α for Gaussian, stationary noise. We define the FAR for a value of the projection coefficient α as the probability of obtaining a measurement equal or greater to α from data without a signal. Since our noise is Gaussian and stationary, this is given by the survival function of a Gaussian distribution. Here, we plot the survival function for the normal distribution, which gives the FAR as a function of the ratio of α to the error on α.

LVK catalogs [2–4], the FARs reported are computed using real data and using different methods depending on the pipeline used. In this work, we define the FAR for a given value of α to be the probability of obtaining a value of α equal or greater when there is no signal in the data. As we are considering Gaussian, stationary noise, the (frequentist) distribution of α measurements from noise is a Gaussian distribution centred around zero, with variance given by Eq. (16). The FAR is then the survival function of this Gaussian evaluated at α,

$$\mathrm{FAR}(\alpha) = 1 - \int_{-\infty}^{\alpha} d\alpha \frac{1}{K_{\alpha}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\alpha}{K_{\alpha}}\right)^{2}\right)$$

The FAR for any pair of measured α and K_{α} can be obtained by substituting α/K_{α} in the survival function for the normal distribution, plotted in Fig. 10. The same can be done with $\tilde{\alpha}_{1}/\tilde{K}_{\alpha}$.
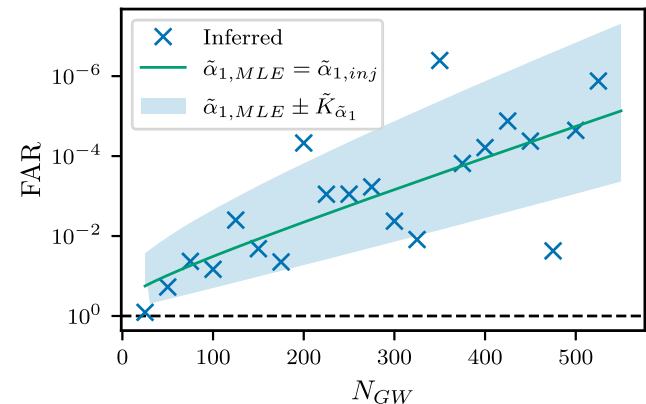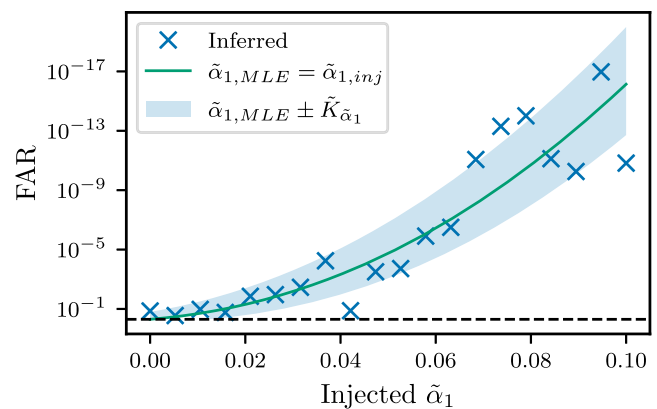

FIG. 11. FAR, or the probability of measuring a projection coefficient equal or greater to $\tilde{\alpha}_1$. The y-axis scale has been inverted to show that the FAR scales in the same way as $\mathcal{B}_{\mathrm{SM}}^{\mathrm{BM}}$, the Bayes factor in favor of a BM hypothesis, shown in Fig. 7. The dashed line is a visual aid indicating a FAR of 0.5.

The FAR scales with the injected projected coefficient and the number of events in the same way as the Bayes factor, as shown in Fig. 11. In future work, we will compute the FAR by considering available GW data and how often an event that is not believed to contain any BM signal can lead to a nonzero measurement of the projection coefficients.

[1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Observation of Gravitational Waves from a Binary Black Hole Merger, Phys. Rev. Lett. **116**, 061102 (2016).
[2] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs, Phys. Rev. X **9**, 031040 (2019).
[3] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-2: Compact Binary Coalescences Observed
by LIGO and Virgo During the First Half of the Third Observing Run, Phys. Rev. X **11**, 021053 (2021).
[4] B. P. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, arXiv:2111.03606.
[5] B. P. Abbott *et al.* (LIGO Scientific, Virgo and KAGRA Collaborations), Tests of general relativity with GWTC-3, arXiv:2112.06861 [Phys. Rev. D (to be published)].

[6] M. Agathos, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, J. Veitch, and S. Vitale, TIGER: A data analysis pipeline for testing the strong-field dynamics of general relativity with gravitational wave signals from coalescing compact binaries, Phys. Rev. D **89,** 082001 (2014).

[7] N. Yunes and F. Pretorius, Fundamental theoretical bias in gravitational wave astrophysics and the parameterized post-Einsteinian framework, Phys. Rev. D **80,** 122003 (2009).

[8] N. K. Johnson-McDaniel, A. Ghosh, S. Ghonge, M. Saleem, N. V. Krishnendu, and J. A. Clark, Investigating the relation between gravitational wave tests of general relativity, Phys. Rev. D **105,** 044020 (2022).

[9] B. Edelman, F. J. Rivera-Paleo, J. D. Merritt, B. Farr, Z. Doctor, J. Brink, W. M. Farr, J. Gair, J. S. Key, J. McIver, and A. B. Nielsen, Constraining unmodeled physics with compact binary mergers from GWTC-1, Phys. Rev. D **103,** 042004 (2021).

[10] Q. Hu and J. Veitch, Accumulating errors in tests of general relativity with the Einstein Telescope: Overlapping signals and inaccurate waveforms, Astrophys. J. **945,** 103 (2023).

[11] Z. Lyu, N. Jiang, and K. Yagi, Constraints on Einstein-dilation-Gauss-Bonnet gravity from Black hole-neutron star gravitational wave events, Phys. Rev. D **105,** 064001 (2022).

[12] R. Niu, X. Zhang, B. Wang, and W. Zhao, Constraining scalar-tensor theories using neutron star–black hole gravitational wave events, Astrophys. J. **921,** 149 (2021).

[13] K. Pardo, M. Fishbach, D. E. Holz, and D. N. Spergel, Limits on the number of spacetime dimensions from GW170817, J. Cosmol. Astropart. Phys. 07 (2018) 048.

[14] R. Nair, S. Perkins, H. O. Silva, and N. Yunes, Fundamental Physics Implications for Higher-Curvature Theories from Binary Black Hole Signals in the LIGO-Virgo Catalog GWTC-1, Phys. Rev. Lett. **123,** 191101 (2019).

[15] N. Yunes, K. Yagi, and F. Pretorius, Theoretical physics implications of the binary black-hole mergers GW150914 and GW151226, Phys. Rev. D **94,** 084002 (2016).

[16] V. A. Kostelecký and M. Mewes, Testing local Lorentz invariance with gravitational waves, Phys. Lett. B **757,** 510 (2016).

[17] S. Mastrogiovanni, D. A. Steer, and M. Barsuglia, Probing modified gravity theories and cosmology using gravitational-waves and associated electromagnetic counterparts, Phys. Rev. D **102,** 044009 (2020).

[18] M. Okounkova, W. M. Farr, M. Isi, and L. C. Stein, Constraining gravitational wave amplitude birefringence and Chern-Simons gravity with GWTC-2, Phys. Rev. D **106,** 044067 (2022).

[19] L. Shao, Combined search for anisotropic birefringence in the gravitational-wave transient catalog GWTC-1, Phys. Rev. D **101,** 104019 (2020).

[20] J. Abedi, H. Dykaar, and N. Afshordi, Echoes from the abyss: Tentative evidence for Planck-scale structure at black hole horizons, Phys. Rev. D **96,** 082004 (2017).

[21] G. Ashton, O. Birnholtz, M. Cabero, C. Capano, T. Dent, B. Krishnan, G. D. Meadors, A. B. Nielsen, A. Nitz, and J. Westerweck, Comments on: "Echoes from the abyss: Evidence for Planck-scale structure at black hole horizons", arXiv:1612.05625.

[22] N. Uchikata, H. Nakano, T. Narikawa, N. Sago, H. Tagoshi, and T. Tanaka, Searching for black hole echoes from the LIGO-Virgo catalog GWTC-1, Phys. Rev. D **100,** 062006 (2019).

[23] J. Westerweck, A. B. Nielsen, O. Fischer-Birnholtz, M. Cabero, C. Capano, T. Dent, B. Krishnan, G. Meadors, and A. H. Nitz, Low significance of evidence for black hole echoes in gravitational wave data, Phys. Rev. D **97,** 124037 (2018).

[24] C. M. Will, Bounding the mass of the graviton using gravitational-wave observations of inspiralling compact binaries, Phys. Rev. D **57,** 2061 (1998).

[25] E. Belgacem, Y. Dirian, S. Foffa, and M. Maggiore, The gravitational-wave luminosity distance in modified gravity theories, Phys. Rev. D **97,** 104066 (2018).

[26] E. Belgacem, Y. Dirian, S. Foffa, and M. Maggiore, Modified gravitational-wave propagation and standard sirens, Phys. Rev. D **98,** 023510 (2018).

[27] A. Nishizawa, Generalized framework for testing gravity with gravitational-wave propagation. I. Formulation, Phys. Rev. D **97,** 104037 (2018).

[28] S. Mukherjee, B. D. Wandelt, and J. Silk, Testing the general theory of relativity using gravitational wave propagation from dark standard sirens, Mon. Not. R. Astron. Soc. **502,** 1136 (2021).

[29] M. Mewes, Signals for Lorentz violation in gravitational waves, Phys. Rev. D **99,** 104062 (2019).

[30] S. Mirshekari, N. Yunes, and C. M. Will, Constraining Lorentz-violating, modified dispersion relations with gravitational waves, Phys. Rev. D **85,** 024041 (2012).

[31] K. O'Neal-Ault, Q. G. Bailey, T. Dumerchat, L. Haegel, and J. Tasson, Analysis of birefringence and dispersion effects from Spacetime-symmetry breaking in gravitational waves, Universe **7,** 380 (2021).

[32] E. Berti, V. Cardoso, and C. M. Will, Gravitational-wave spectroscopy of massive black holes with the space interferometer LISA, Phys. Rev. D **73,** 064030 (2006).

[33] V. Cardoso, E. Franzin, and P. Pani, Is the Gravitational-Wave Ringdown a Probe of the Event Horizon?, Phys. Rev. Lett. **116,** 171101 (2016).

[34] V. Cardoso, S. Hopper, C. F. B. Macedo, C. Palenzuela, and P. Pani, Gravitational-wave signatures of exotic compact objects and of quantum corrections at the horizon scale, Phys. Rev. D **94,** 084031 (2016).

[35] V. Cardoso and P. Pani, Tests for the existence of black holes through gravitational wave echoes, Nat. Astron. **1,** 586 (2017).

[36] S. Gossan, J. Veitch, and B. S. Sathyaprakash, Bayesian model selection for testing the no-hair theorem with black hole ringdowns, Phys. Rev. D **85,** 124056 (2012).

[37] J. Meidam, M. Agathos, C. Van Den Broeck, J. Veitch, and B. S. Sathyaprakash, Testing the no-hair theorem with black hole ringdowns using TIGER, Phys. Rev. D **90,** 064009 (2014).

[38] O. Dreyer, B. Kelly, B. Krishnan, L. S. Finn, D. Garrison, and R. Lopez-Aleman, Black-hole spectroscopy: Testing general relativity through gravitational-wave observations, Classical Quantum Gravity **21,** 787 (2004).

[39] P. O. Mazur and E. Mottola, Gravitational vacuum condensate stars, Proc. Natl. Acad. Sci. U.S.A. **101,** 9545 (2004).

[40] S. L. Liebling and C. Palenzuela, Dynamical boson stars, Living Rev. Relativity **15,** 6 (2012).

[41] G. F. Giudice, M. McCullough, and A. Urbano, Hunting for dark particles with gravitational waves, J. Cosmol. Astropart. Phys. 10 (2016) 001.

[42] U. Danielsson, L. Lehner, and F. Pretorius, Dynamics and observational signatures of shell-like black hole mimickers, Phys. Rev. D **104,** 124011 (2021).

[43] D. M. Eardley, D. L. Lee, and A. P. Lightman, Gravitational-wave observations as a tool for testing relativistic gravity, Phys. Rev. D **8,** 3308 (1973).

[44] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), Tests of General Relativity with GW170817, Phys. Rev. Lett. **123,** 011102 (2019).

[45] H. Yang, K. Yagi, J. Blackman, L. Lehner, V. Paschalidis, F. Pretorius, and N. Yunes, Black Hole Spectroscopy with Coherent Mode Stacking, Phys. Rev. Lett. **118,** 161101 (2017).

[46] A. Zimmerman, C.-J. Haster, and K. Chatziioannou, On combining information from multiple gravitational wave sources, Phys. Rev. D **99,** 124044 (2019).

[47] V. Kalogera et al., The next generation global gravitational wave observatory: The science book, arXiv:2111.06990.

[48] E. Barausse et al., Prospects for fundamental physics with LISA, Gen. Relativ. Gravit. **52,** 81 (2020).

[49] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914, Classical Quantum Gravity **33,** 134001 (2016).

[50] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals, Classical Quantum Gravity **37,** 055002 (2020).

[51] L. Blackburn et al., The LSC glitch group: Monitoring noise transients during the fifth LIGO science run, Classical Quantum Gravity **25,** 184004 (2008).

[52] K. Chatziioannou, C.-J. Haster, T. B. Littenberg, W. M. Farr, S. Ghonge, M. Millhouse, J. A. Clark, and N. Cornish, Noise spectral estimation methods and their impact on gravitational wave measurement of compact binary mergers, Phys. Rev. D **100,** 104004 (2019).

[53] R. Abbott et al. (LIGO Scientific and Virgo Collaborations), Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo, SoftwareX **13,** 100658 (2021).

[54] O. Edy, A. Lundgren, and L. K. Nuttall, Issues of mismodeling gravitational-wave data for parameter estimation, Phys. Rev. D **103,** 124061 (2021).

[55] J. Y. L. Kwok, R. K. L. Lo, A. J. Weinstein, and T. G. F. Li, Investigation of the effects of non-Gaussian noise transients and their mitigation in parameterized gravitational-wave tests of general relativity, Phys. Rev. D **105,** 024066 (2022).

[56] B. P. Abbott et al. (LIGO Scientific Collaboration), Analysis of first LIGO science data for stochastic gravitational waves, Phys. Rev. D **69,** 122004 (2004).

[57] B. Allen and J. D. Romano, Detecting a stochastic background of gravitational radiation: Signal processing strategies and sensitivities, Phys. Rev. D **59,** 102001 (1999).

[58] W. O. Schumann, über die Dämpfung der elektromagnetischen Eigenschwingungen des Systems Erde—Luft—Ionosphäre, Z. Naturforsch. A **7,** 250 (1952).

[59] P. Nguyen et al., Environmental noise in Advanced LIGO detectors, Classical Quantum Gravity **38,** 145001 (2021).

[60] E. Thrane, N. Christensen, and R. M. S. Schofield, Correlated magnetic noise in global networks of gravitational-wave detectors: Observations and implications, Phys. Rev. D **87,** 123009 (2013).

[61] E. Thrane, N. Christensen, R. M. S. Schofield, and A. Effler, Correlated noise in networks of gravitational-wave detectors: Subtraction and mitigation, Phys. Rev. D **90,** 023013 (2014).

[62] M. Vallisneri and N. Yunes, Stealth bias in gravitational-wave parameter estimation, Phys. Rev. D **87,** 102002 (2013).

[63] S. Vitale and W. Del Pozzo, How serious can the stealth bias be in gravitational wave parameter estimation?, Phys. Rev. D **89,** 022002 (2014).

[64] M. Okounkova, M. Isi, K. Chatziioannou, and W. M. Farr, Gravitational wave inference on a numerical-relativity simulation of a black hole merger beyond general relativity, Phys. Rev. D **107,** 024046 (2023).

[65] F. D. Ryan, Gravitational waves from the inspiral of a compact object into a massive, axisymmetric body with arbitrary multipole moments, Phys. Rev. D **52,** 5707 (1995).

[66] E. Poisson, Gravitational waves from inspiraling compact binaries: The quadrupole-moment term, Phys. Rev. D **57,** 5287 (1998).

[67] W. G. Laarakkers and E. Poisson, Quadrupole moments of rotating neutron stars, Astrophys. J. **512,** 282 (1999).

[68] N. V. Krishnendu, K. G. Arun, and C. K. Mishra, Testing the Binary Black Hole Nature of a Compact Binary Coalescence, Phys. Rev. Lett. **119,** 091101 (2017).

[69] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), Properties and astrophysical implications of the $150 M_\odot$ binary black hole merger GW190521, Astrophys. J. **900,** L13 (2020).

[70] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), Tests of General Relativity with GW150914, Phys. Rev. Lett. **116,** 221101 (2016).

[71] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), Tests of general relativity with the binary black hole signals from the LIGO-Virgo catalog GWTC-1, Phys. Rev. D **100,** 104036 (2019).

[72] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog, Phys. Rev. D **103,** 122002 (2021).

[73] T. Damour and G. Esposito-Farese, Tensor multiscalar theories of gravitation, Classical Quantum Gravity **9,** 2093 (1992).

[74] E. Barausse, C. Palenzuela, M. Ponce, and L. Lehner, Neutron-star mergers in scalar-tensor theories of gravity, Phys. Rev. D **87,** 081506 (2013).

[75] C. M. Will, The confrontation between general relativity and experiment, Living Rev. Relativity **17,** 4 (2014).

[76] N. J. Cornish and T. B. Littenberg, BayesWave: Bayesian inference for gravitational wave bursts and instrument glitches, Classical Quantum Gravity **32,** 135012 (2015).

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

[78] J. Meidam *et al.*, Parametrized tests of the strong-field dynamics of general relativity using gravitational wave signals from coalescing binary black holes: Fast likelihood calculations and sensitivity of the method, Phys. Rev. D **97,** 044033 (2018).

[79] A. K. Mehta, A. Buonanno, R. Cotesta, A. Ghosh, N. Sennett, and J. Steinhoff, Tests of general relativity with gravitational-wave observations using a flexible–theory-independent method, Phys. Rev. D **107,** 044020 (2023).

[80] S. Dhurandhar, B. Krishnan, H. Mukhopadhyay, and J. T. Whelan, Cross-correlation search for periodic gravitational waves, Phys. Rev. D **77,** 082001 (2008).

[81] T. Bayes and n. Price, LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S, Phil. Trans. R. Soc. London **53,** 370 (1763).

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

[83] A. Nitz *et al.*, Gwastro/pycbc: V2.0.2 release of PyCBC, Zenodo (2022).

[84] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal, Phys. Rev. D **93,** 044006 (2016).

[85] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era, Phys. Rev. D **93,** 044007 (2016).

[86] B. P. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, Living Rev. Relativity **23,** 3 (2020).

[87] L. Bernard, Dynamics of compact binary systems in scalar-tensor theories: Equations of motion to the third post-Newtonian order, Phys. Rev. D **98,** 044004 (2018).

[88] L. Sampson, N. Yunes, N. Cornish, M. Ponce, E. Barausse, A. Klein, C. Palenzuela, and L. Lehner, Projected constraints on scalarization with gravitational waves from neutron star binaries, Phys. Rev. D **90,** 124091 (2014).

[89] M. Isi, W. M. Farr, and K. Chatziioannou, Comparing Bayes factors and hierarchical inference for testing general relativity with gravitational waves, Phys. Rev. D **106,** 024048 (2022).

[90] N. Sennett, R. Brito, A. Buonanno, V. Gorbenko, and L. Senatore, Gravitational-wave constraints on an effective field-theory extension of general relativity, Phys. Rev. D **102,** 044056 (2020).

[91] C. de Rham, L. Heisenberg, and A. J. Tolley, Spin-2 fields and the weak gravity conjecture, Phys. Rev. D **100,** 104033 (2019).

[92] V. Cardoso, M. Kimura, A. Maselli, and L. Senatore, Black Holes in an Effective Field Theory Extension of General Relativity, Phys. Rev. Lett. **121,** 251105 (2018).

[93] P. A. Cano, K. Fransen, T. Hertog, and S. Maenaut, Gravitational ringing of rotating black holes in higher-derivative gravity, Phys. Rev. D **105,** 024064 (2022).

[94] R. Cayuso and L. Lehner, Nonlinear, noniterative treatment of EFT-motivated gravity, Phys. Rev. D **102,** 084008 (2020).

[95] S. T. McWilliams, Analytical Black-Hole Binary Merger Waveforms, Phys. Rev. Lett. **122,** 191102 (2019).

[96] A. Buonanno, L. E. Kidder, and L. Lehner, Estimating the final spin of a binary black hole coalescence, Phys. Rev. D **77,** 026004 (2008).

[97] M. Kesden, Can binary mergers produce maximally spinning black holes?, Phys. Rev. D **78,** 084030 (2008).

[98] A. Buonanno and T. Damour, Transition from inspiral to plunge in binary black hole coalescences, Phys. Rev. D **62,** 064015 (2000).

[99] K. Hotokezaka, K. Kyutoku, and M. Shibata, Exploring tidal effects of coalescing binary neutron stars in numerical relativity, Phys. Rev. D **87,** 044001 (2013).

[100] T. Damour, A. Nagar, and L. Villain, Measurability of the tidal polarizability of neutron stars in late-inspiral gravitational-wave signals, Phys. Rev. D **85,** 123007 (2012).

[101] S. A. Bhat, P. Saini, M. Favata, and K. G. Arun, Systematic bias on the inspiral-merger-ringdown consistency test due to neglect of orbital eccentricity, Phys. Rev. D **107,** 024009 (2023).

[102] C. R. Harris *et al.*, Array programming with NumPy, Nature (London) **585,** 357 (2020).

[103] P. Virtanen *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in PYTHON, Nat. Methods **17,** 261 (2020).

[104] J. D. Hunter, Matplotlib: A 2d graphics environment, Comput. Sci. Eng. **9,** 90 (2007).

[105] M. L. Waskom, SEABORN: Statistical data visualization, J. Open Source Software **6,** 3021 (2021).