Detecting non-Gaussian gravitational wave backgrounds: A unified framework

Riccardo Buscicchio,[1,2,3,*] Anirban Ain,[4,5] Matteo Ballelli,[4,5] Giancarlo Cella,[4] and Barbara Patricelli[4,5]

[1]*Dipartimento di Fisica "G. Occhialini," Universitá degli Studi di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy*

[2]*INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy*

[3]*Institute for Gravitational Wave Astronomy and School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, United Kingdom*

[4]*INFN Sezione Pisa, Largo B. Pontecorvo 3, I-56127 Pisa, Italy*

[5]*Università di Pisa, Dipartimento di Fisica "E. Fermi," Largo B. Pontecorvo 3, I-56127 Pisa, Italy*

(Received 14 September 2022; accepted 19 December 2022; published 24 March 2023)

We describe a novel approach to the detection and parameter estimation of a non-Gaussian stochastic background of gravitational waves. The method is based on the determination of relevant statistical parameters using importance sampling. We show that it is possible to improve the Gaussian detection statistics by simulating realizations of the expected signal for a given model. While computationally expensive, our method improves the detection performance, leveraging the prior knowledge on the expected signal, and can be used in a natural way to extract physical information about the background. We present the basic principles of our approach, characterize the detection statistic performances in a simplified context, and discuss possible applications to the detection of some astrophysical foregrounds. We argue that the proposed approach, complementarily to the ones available in literature might be used to detect suitable astrophysical foregrounds by currently operating and future gravitational wave detectors.

DOI: 10.1103/PhysRevD.107.063027

I. INTRODUCTION

Over the past seven years, the Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) [1] and Advanced Virgo [2] have collected data and released, together with the KAGRA Collaboration [3], three incremental catalogs of gravitational-wave (GW) detections, amounting to a total of 90 confident events [4]. In addition, the LIGO, Virgo, and KAGRA Collaborations (LVKC) have performed a population study on a subset of 76 of them [5]. Further upgraded second generation interferometers [6] and third generation GW detectors, such as the Einstein Telescope (ET) [7], will possibly become operational during the next decade: these experiments promise to be sensitive enough to observe both the cosmological and astrophysical stochastic gravitational wave background (SGWB). In addition, when the Large Interferometer Space Antenna (LISA) [8] becomes operational, our sensitivity to astrophysical GW transients will broaden to lower frequencies and new source categories.

The superposition from various unresolved astrophysical and cosmological sources generates a SGWB. Searches for such a stochastic background have been performed on available data: no evidence for a SGWB has been found; nonetheless upper limits on its cosmological energy density have been placed [5,9]. Among the sources that may contribute to the SGWB: core-collapse supernovae [10–12]; neutron stars [13–15]; compact binary coalescences [16–19]; binary white dwarfs [20]; cosmic strings [21–23]; and gravitational waves produced during inflation [24–26] or by primordial black holes [27]. A detection of the cosmological SGWB would give very important constraints on the earliest epochs of the Universe, while the detection of an astrophysical SGWB would provide key information about the sources generating it, e.g., the merger rate of compact binary systems, the star formation history [28,29], or the occurrence of gravitational-wave lensing [30,31].

Searches for SGWBs typically assume that the background is Gaussian, based on the central limit theorem (see, e.g., Refs. [32,33]). However, if the rate of events generating the background is not sufficiently high compared to their duration or frequency bandwidth, a non-Gaussian background is expected, characterized by discontinuous or intermittent signals. For instance, predictions based on population modeling suggest that, for many realistic astrophysical models, there may not be enough overlapping sources, resulting in the formation of such a non-Gaussian background (see, e.g., Refs. [34,35]). Furthermore, it has been shown that the background from cosmic strings could be dominated by a non-Gaussian contribution arising from the closest sources [22].

*riccardo.buscicchio@unimib.it

In the past decades several methods to search for non-Gaussian SGWB have been proposed. For instance, the authors of Ref. [36] derived an algorithm suitable for the detection of a non-Gaussian component in a SGWB observed by two colocated and coaligned detectors with white Gaussian noise. Later, the author of Ref. [37] introduced a maximum likelihood estimator to be used in a more realistic case of a network of spatially separated interferometers with colored, non-Gaussian noise. Most recently, the authors of Ref. [38] devised a Bayesian search strategy for a background of unresolved binaries. Other approaches have also been explored (see, e.g., Refs. [39–41]), constructing alternative parametrizations for SGWB non-Gaussianities. In the context of LISA, various pipelines for the detection and characterization of an astrophysical SGWB have been developed [42–44], parametrizing a certain level on non-Gaussianities in the signal model. The expected level has also been assessed for confusion noise arising from extreme mass ratio inspirals (EMRIs) and galactic binary white dwarfs [45].

In this paper, we explore a novel approach for a detection of non-Gaussian SGWBs–inherently complementary to the ones available in literature [38]—using a detailed stochastic model of the underlying signal population. The paper is organized as follows: in Sec. II we discuss the basic principles for the detection of a SGWB, and we give examples of application for the case of an isotropic background; after a discussion of the Neyman-Pearson detection statistic (DS) in a frequentist context (Sec. II A) we show how a Bayesian analysis of a non-Gaussian stochastic background can be implemented (Sec. II B); in Sec. III we discuss a simplified model for a non-Gaussian stochastic background, with the purpose of estimating the improvement in detection performance of the proposed approach; in Sec. IV we give details about the application to a more realistic case, namely an isotropic stochastic background of astrophysical origin; we show how this can be represented by a (generalized) point process (Sec. IV A), and give details about the stochastic sampling procedure required by the inference method (Sec. IV B); finally, in Sec. V we draw some conclusions pointing at possible future developments, in particular toward applications to the nonisotropic case; in the Appendix we provide detailed proofs of results shown in the main text. Some are available in literature (see, e.g., Ref. [46] and references therein); nonetheless we choose to reproduce them to ensure consistency of notation across the text.

II. THE STATISTICAL PROBLEM

In this paper, observations are written as the sum of signal and noise; however, in our case it is more convenient to write the data collected by a network of detectors in a slightly different form, namely

$$s_i^{\mathcal{A}} = g_i^{\mathcal{A}} + h_i^{\mathcal{A}} + n_i^{\mathcal{A}}, \tag{1}$$

where g is a Gaussian part of the stochastic signal and h a non-Gaussian one, while n is the noise of the detectors. We assume statistical independence among the three components g, h, and n. We assume also that they have zero mean. A nonzero average is observationally irrelevant and can be removed in the time domain. In the frequency domain it could model a spurious nonstochastic contamination that should be removed before the analysis.

In this paper we always assume an additive and Gaussian noise, although an extension is possible to account for transient nonstationarities arising from the noise. Capital indices label the detector while lowercase ones enumerate generically the data series: we will specialize it if needed by explicitly writing our expressions in time or frequency domain. It is worth emphasizing that the decomposition in Eq. (1) is not unique: one can always add and subtract an arbitrary Gaussian contribution to g and from h. This is a feature arising from the inherent modeling freedom for h in Eq. (1), and it is not related to the noise properties.

Under our hypotheses the noise is described by a multivariate Gaussian probability distribution that we can write as

$$p_n[n_i^{\mathcal{A}}] = \mathcal{N}_n \exp\left(-\frac{1}{2}\mathcal{W}_n(n,n)\right), \tag{2}$$

where $\mathcal{N}_n$ is a normalization constant and for future convenience we defined the scalar product over detectors and data indices

$$\mathcal{W}_x(u,v) \equiv \sum_{\mathcal{A},\mathcal{B}} \sum_{i,j} [\mathbb{C}_{xx}^{-1}]_{ij}^{\mathcal{A}\mathcal{B}} u_i^{\mathcal{A}} v_j^{\mathcal{B}}, \tag{3}$$

$$[\mathbb{C}_{xy}]_{ij}^{\mathcal{A}\mathcal{B}} \equiv \langle x_i^{\mathcal{A}} y_j^{\mathcal{B}} \rangle. \tag{4}$$

Hereafter, following Einstein's convention on repeated indices, we drop the summation symbols over data and detectors indices. $\mathbb{C}_{nn}$ is the noise cross-correlation array, so $\mathcal{W}_n$ in Eq. (2) is the Wiener match between u and v with respect to the noise n [47]. Explicitly we can write

$$\mathcal{N}_x = \exp\left(-\frac{1}{2}\mathrm{Tr}\,\ln 2\pi\mathbb{C}_x\right), \tag{5}$$

$$[\mathbb{C}_{xy}]_{ij}^{\mathcal{A}\mathcal{B}} \approx [\check{\mathbb{C}}_{xy}]_{ij}^{\mathcal{A}\mathcal{B}} = \overline{x_i^{\mathcal{A}} y_j^{\mathcal{B}}}, \tag{6}$$

where the trace is performed over detectors and data indices. For simplicity, in autocorrelations $\mathbb{C}_{xx}$ we drop a redundant index, therefore denoting them $\mathbb{C}_x$.

In Eq. (6) and in what follows we often replace the true cross-correlations $\mathbb{C}_{xy}$—a theoretical expectation value defined through the model and frequently unmeasurable—with estimators obtained from the data. We label them with an overhead check. Correspondingly, averaging over the data indices is denoted with an overline,

$$[\check{\mathbb{C}}_{xy}]_{ij}^{\mathcal{AB}} = \overline{x_i^{\mathcal{A}} y_j^{\mathcal{B}}}. \tag{7}$$

Because of statistical fluctuations, the uncertainty on $\check{\mathbb{C}}_{xy}$ can be improved by averaging over chunks of data: as we show in Eq. (24), this comes at the cost of a reduced probability of detection.

We model the stochastic signal described by g as Gaussian with a probability distribution analogous to Eq. (2), namely

$$p_g[g_i^{\mathcal{A}}] = \mathcal{N}_g \exp\left(-\frac{1}{2}\mathcal{W}_g(g,g)\right). \tag{8}$$

However, we make no statistical hypothesis about the remainder h, which can be described by a generic probability distribution $p_h[h_i^{\mathcal{A}}]$. Then we can write the probability distribution for the observed signal as a convolution between p_n, p_g, and p_h, namely

$$\begin{aligned} p_s[s] &= \int_h \int_g p_s[s|h,g] p_h[h] p_g[g] \\ &= \mathcal{N}_n \mathcal{N}_g \int_h \int_g p_h[h] e^{-\frac{1}{2}\mathcal{W}_n(s-h-g,s-h-g)-\frac{1}{2}\mathcal{W}_g(g,g)}. \end{aligned} \tag{9}$$

The Gaussian integral over g can be performed explicitly. By virtue of Woodbury's identity (see the Appendix for details), we observe that

$$\mathcal{W}_{n+g}(u,v) = \mathcal{W}_n(u,v) - \mathcal{G}(u,v) \tag{10}$$

or equivalently

$$\mathbb{C}_n^{-1} - \mathbb{C}_n^{-1}(\mathbb{C}_g^{-1} + \mathbb{C}_n^{-1})^{-1}\mathbb{C}_n^{-1} = \mathbb{C}_{n+g}^{-1}, \tag{11}$$

where we have defined for future convenience

$$\mathcal{G}(u,v) \equiv \mathbb{G}_{ij}^{\mathcal{AB}} u_i^{\mathcal{A}} v_j^{\mathcal{B}}, \tag{12}$$

$$\mathbb{G} \equiv \mathbb{C}_n^{-1}(\mathbb{C}_n^{-1} + \mathbb{C}_g^{-1})^{-1}\mathbb{C}_n^{-1}, \tag{13}$$

which go to zero when $g = 0$. Using Eq. (11) the integral further simplifies to

$$p_s[s] = \mathcal{N}_{n+g} \int_h p_h[h] e^{-\frac{1}{2}\mathcal{W}_{n+g}(s-h,s-h)}. \tag{14}$$

The key point is that we can rewrite p_s as

$$p_s[s] = \mathcal{N}_{n+g} \langle e^{-\frac{1}{2}\mathcal{W}_{n+g}(s-h,s-h)} \rangle, \tag{15}$$

where the expectation value $\langle \cdots \rangle$ is evaluated over an ensemble of realizations for the non-Gaussian part h of the SGWB. Note that this expectation value is evaluated at fixed data s, which is considered here an independent variable.

While it is difficult to write an explicit expression for p_h in the non-Gaussian case, realizations of a stochastic background h can be simulated. This opens up the possibility of evaluating p_s and connected quantities related to DS and parameter estimation procedures.

A. The frequentist approach

As a first example we show an expression for the optimal Neyman-Pearson DS [48], under the hypothesis of a known background and a known noise. The two hypotheses to be tested are

$\mathcal{H}_1$: presence of a known stochastic background, with a given Gaussian part g and a given non-Gaussian one h.

$\mathcal{H}_0$: absence of the background, $g = h = 0$, which means $s = n$.

The DS is defined by the test statistic $\hat{Y}(s) > \lambda$ where

$$\hat{Y}(s) \equiv \log \frac{p_s[s|\mathcal{H}_1]}{p_s[s|\mathcal{H}_0]} - \log \frac{\mathcal{N}_{n+g}}{\mathcal{N}_n} \tag{16}$$

$$= \frac{1}{2}\mathcal{G}(s,s) + \log \langle e^{-\frac{1}{2}\mathcal{W}_{n+g}(h,h)} e^{\mathcal{W}_{n+g}(s,h)} \rangle. \tag{17}$$

We subtracted from the standard definition of $\hat{Y}(s)$ a data independent constant, whose effect can be compensated by a redefinition of the relation between the threshold λ and the false alarm probability [49]. Note that the average in Eq. (17) is evaluated under the $\mathcal{H}_1$ hypothesis.

1. Gaussian case

We discuss shortly the particular case of a Gaussian background, as this clarifies some aspects relevant in the following sections. If the background is Gaussian, we can assume, without loss of generality, that $h = 0$ and the optimal statistic is given by

$$\begin{aligned} \hat{Y}(s) &= \frac{1}{2}\mathcal{G}(s,s) \\ &\simeq \frac{1}{2}[\mathbb{C}_n^{-1}\mathbb{C}_g\mathbb{C}_n^{-1}]_{ij}^{\mathcal{AB}} s_i^{\mathcal{A}} s_j^{\mathcal{B}} + \mathcal{O}(\|\mathbb{C}_g\mathbb{C}_n^{-1}\|^2), \end{aligned} \tag{18}$$

where we expanded Eq. (12) to lowest order, under the hypothesis that the SGWB power spectrum is much smaller than every detector's noise spectrum. While the frequentist approach makes direct use of such an assumption, the corresponding Bayesian approach in Sec. II B does not assume it, hence making it suitable in other contexts. As $\hat{Y}$ is an approximately Gaussian variable, we are comparing two Gaussian distributions with given means and variances.

Having access to only estimators of noise and signal spectra ensemble averages, we use them to replace correlations in the test statistics. Consequently, the average of $\hat{Y}$

(i.e., the optimal statistics using estimators for noise correlations) under $\mathcal{H}_0$ is given by

$$\mu_{\mathcal{H}_0} = \frac{1}{2}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}]_{ij}^{\mathcal{AB}}\langle n_i^{\mathcal{A}} n_j^{\mathcal{B}}\rangle \tag{19}$$

$$= \frac{1}{2}\mathrm{Tr}(\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_n), \tag{20}$$

where the trace is performed over detector and data indices. In a similar way we find, under the hypothesis $\mathcal{H}_1$,

$$\mu_{\mathcal{H}_1} = \mu_{\mathcal{H}_0} + \frac{1}{2}\mathrm{Tr}(\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_g), \tag{21}$$

and the variances are given by

$$\sigma^2_{\mathcal{H}_0} = \frac{1}{2}\mathrm{Tr}[(\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_n)^2], \tag{22}$$

$$\sigma^2_{\mathcal{H}_1} \simeq \sigma^2_{\mathcal{H}_0} + \mathrm{Tr}(\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_n\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_g), \tag{23}$$

where once again we included only the first correction for $\sigma^2_{\mathcal{H}_1}$ in the small signal approximation. The receiver operating characteristic (ROC) of the DS reads

$$P_{\mathrm{D}} = \frac{1}{2}\mathrm{erfc}\left(\frac{\sigma_{\mathcal{H}_0}}{\sigma_{\mathcal{H}_1}}\mathrm{erfc}^{-1}(2P_{\mathrm{FA}}) - \frac{\mu_{\mathcal{H}_1} - \mu_{\mathcal{H}_0}}{\sigma_{\mathcal{H}_1}\sqrt{2}}\right), \tag{24}$$

where P_{D} (P_{FA}) is the detection (false-alarm) probability and erfc is the complementary error function.

Note that, because of the two traces over the N data points $(\mu_{\mathcal{H}_1} - \mu_{\mathcal{H}_0})/\sigma_{\mathcal{H}_1} \propto \sqrt{N}$, so the detection probability improves with the square root of the measurement time. On the contrary, the first term affects only mildly the detector performance as it remains constant while more data points are accumulated. For this reason, Eq. (24) is often rewritten with trivial definitions for the "offset" o and "deflection coefficient" d as follows:

$$P_{\mathrm{D}} = \frac{1}{2}\mathrm{erfc}\left(o - \sqrt{d^2}\right). \tag{25}$$

However, the approach just illustrated is not always viable: to attain a detection we need to know $\mu_{\mathcal{H}_0}$ with an error of the order of the ratio between the signal's and the noise's power spectra. This is because we need to know $\mu_{\mathcal{H}_1} - \mu_{\mathcal{H}_0}$ with the same precision. This cannot be done experimentally (we cannot switch off the coupling of the detectors to the SGWB), and it is not realistic to estimate theoretically the noise budget of a detector with such precision.

Usually this issue is solved by the additional assumption that noises across different detectors are uncorrelated, namely the matrix $\mathbb{C}_n^{\mathcal{AB}}$ is diagonal in detector's indices. Consequently, the noise dominated terms along the diagonal $\mathcal{A} = \mathcal{B}$ can be eliminated by defining a "diagonal–free" statistic $\hat{Y}_G(s)$ by removing in the sum of Eq. (18) all terms dominated by the noise

$$\hat{Y}_G(s) \equiv \sum_{\mathcal{A}\neq\mathcal{B}} \frac{1}{2}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}]_{ij}^{\mathcal{AB}} s_i^{\mathcal{A}} s_j^{\mathcal{B}}. \tag{26}$$

We get a new average $\mu_{\mathcal{H}_{0,\mathrm{G}}} = 0$, and the detector becomes robust with respect to errors in the noise model. The new means and variances are given by

$$\mu_{\mathcal{H}_{0,\mathrm{G}}} = 0, \tag{27}$$

$$\mu_{\mathcal{H}_{1,\mathrm{G}}} = \frac{1}{2}\mathrm{Tr}(\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_g), \tag{28}$$

$$\sigma^2_{\mathcal{H}_{0,\mathrm{G}}} = \frac{1}{2}\mathrm{Tr}((\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_n)^2), \tag{29}$$

$$\sigma^2_{\mathcal{H}_{1,\mathrm{G}}} \simeq \sigma^2_{\mathcal{H}_{0,\mathrm{G}}} + \mathrm{Tr}(\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_n\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_g), \tag{30}$$

where we label diagonal-free correlation matrices (with no implicit summation over detector indices)

$$[\mathbb{C}]^{\mathcal{AB}} = [\mathbb{C}]^{\mathcal{AB}}(1 - \delta^{\mathcal{AB}}). \tag{31}$$

Notably $\mu_{\mathcal{H}_{1,\mathrm{G}}} - \mu_{\mathcal{H}_{0,\mathrm{G}}} < \mu_{\mathcal{H}_1} - \mu_{\mathcal{H}_0}$ based on Eq. (24): the additional robustness introduced affects the deflection coefficient, i.e., its asymptotic performances.

2. Non-Gaussian case

In the more general case of a non-Gaussian model, we rewrite Eq. (16) as

$$\hat{Y}(s) = \frac{1}{2}[\check{\mathbb{C}}_g]_{ij}^{\mathcal{AB}}\mathfrak{s}_i^{\mathcal{A}}\mathfrak{s}_j^{\mathcal{B}} + \chi_h\sum_{n=1}^{\infty}\frac{1}{n!}\check{\Gamma}_{i_1\cdots i_n}^{\mathcal{A}_1\cdots\mathcal{A}_n}\mathfrak{s}_{i_1}^{\mathcal{A}_1}\cdots\mathfrak{s}_{i_n}^{\mathcal{A}_n} \tag{32}$$

with

$$\chi_h = \langle e^{-\frac{1}{2}\mathcal{W}_{n+g}(h,h)}\rangle. \tag{33}$$

Here $\mathfrak{s}_i^{\mathcal{A}} = [\check{\mathbb{C}}_n^{-1}]_{ij}^{\mathcal{AB}}s_j^{\mathcal{B}}$ is a "double whitened" signal, and $\check{\Gamma}_{i_1\cdots i_n}^{\mathcal{A}_1\cdots\mathcal{A}_n}$ are estimators of the connected moments for an h distributed according to

$$p'_h[h] = \chi_h^{-1}e^{-\frac{1}{2}\mathcal{W}_{n+g}(h,h)}p_h[h] \tag{34}$$

and are fully contracted over a suitable number of signals $\mathfrak{s}_i^{\mathcal{A}}$, which we denote with a subscript $\{\mathcal{A}, i\}$.

Now we can evaluate the expectation value of $\hat{Y}(s)$ under the hypothesis $\mathcal{H}_0$. We find

$$\mu_{\mathcal{H}_0} = \frac{1}{2}\mathrm{Tr}(\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_n) + \chi_h \sum_{n=1}^{\infty}\frac{1}{n!}\check{\Gamma}^{\mathcal{A}_1\cdots\mathcal{A}_n}_{i_1\cdots i_n}\mathbb{N}^{\mathcal{A}_1\cdots\mathcal{A}_n}_{i_1\cdots i_n}, \tag{35}$$

where $\mathbb{N}^{\mathcal{A}_1\cdots\mathcal{A}_n}_{i_1\cdots i_n}$ is the nth order moment of the double whitened noise. As the noise is Gaussian, by virtue of the Isserlis theorem [50] it can be written as a sum over all pairings of products of second order moments, and using the symmetry of the connected moments Γ we get

$$\mu_{\mathcal{H}_0} = \frac{1}{2}\mathrm{Tr}(\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_n) + \chi_h \sum_{n=1}^{\infty}\frac{1}{(2n)!!}\check{\Gamma}^{\mathcal{A}_1\cdots\mathcal{A}_{2n}}_{i_1\cdots i_{2n}}\prod_{k=1}^{n}[\check{\mathbb{C}}_n^{-1}\mathbb{C}_n\check{\mathbb{C}}_n^{-1}]^{\mathcal{A}_{(2k-1)}\mathcal{A}_{2k}}_{i_{(2k-1)}i_{2k}}. \tag{36}$$

As for the Gaussian case, $\mu_{\mathcal{H}_0}$ depends on an estimate of the real spectral covariance of the noise, which is not sufficiently under control. We can set to zero the first term in Eq. (36) by using the same approach discussed for the Gaussian case, but this is not enough to eliminate the second. To obtain a robust detector we define the new statistic

$$\mathring{Y}(s) \equiv \hat{Y}(s) - \hat{Y}(\mathring{s}). \tag{37}$$

Here $\mathring{s}$ are the observed data, transformed in such a way that $\mathring{s}^{\mathcal{A}}_i$ satisfies the following:

$$\langle \mathring{s}^{\mathcal{A}}_i \mathring{s}^{\mathcal{B}}_j \rangle = \delta^{\mathcal{A}\mathcal{B}}[\mathbb{C}_s]^{\mathcal{A}\mathcal{B}}_{ij}, \tag{38}$$

$$\langle \mathring{s}^{\mathcal{A}}_i s^{\mathcal{A}}_j \rangle = 0. \tag{39}$$

This can be done by introducing appropriate and large enough shifts among detectors' data in the time domain [51], randomizing the phases in frequency domain, or scrambling data chunks, such that the original series of each detector and the new ones are statistically independent [therefore implying Eq. (39)], and the cross-correlations across detectors are removed [i.e., Eq. (38)]. Henceforth, we will denote $\mathring{s}$ and $\mathring{Y}(s)$—the latter not to be confused with $\hat{Y}(\mathring{s})$, a statistics insensitive by construction to the GW signal—as "scrambled data" and "scrambled detection statistic," respectively. We defer a detailed characterization of the statistical subtleties of this procedure in a realistic scenario to future study.

Under hypothesis $\mathcal{H}_0$ the correlations are computed on noise-only data; therefore they are already diagonal in the detector's indices, so $\langle\hat{Y}(s)\rangle = \langle\hat{Y}(\mathring{s})\rangle$, and we get

$$\mathring{\mu}_{\mathcal{H}_0} \equiv \langle\mathring{Y}(s)\rangle = 0. \tag{40}$$

Taking the expectation value under the hypothesis $\mathcal{H}_1$ we find

$$\mathring{\mu}_{\mathcal{H}_1} = \frac{1}{2}\mathrm{Tr}(\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_g) + \frac{1}{2}\mathrm{Tr}(\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_h) + \chi_h \sum_{n=1}^{\infty}\frac{1}{n!}\check{\Gamma}^{\mathcal{A}_1\cdots\mathcal{A}_n}_{i_1\cdots i_n}(\mathbb{S}^{\mathcal{A}_1\cdots\mathcal{A}_n}_{i_1\cdots i_n} - \mathring{\mathbb{S}}^{\mathcal{A}_1\cdots\mathcal{A}_n}_{i_1\cdots i_n}), \tag{41}$$

where $\mathbb{S}$ are the momenta of the signal $\mathfrak{s}$ and $\mathring{\mathbb{S}}$ the momenta of $\mathring{\mathfrak{s}}$.

When $h = 0$ only the first term is non-null, reproducing the Gaussian result (on scrambled data). From the sum we see that additional contributions arise in the general case. These are expected to improve the DS performances and open up the possibility of a stricter characterization of the SGWB statistical properties.

B. The Bayesian approach

The detection and parameter estimation proposed in Sec. II A can be equivalently formulated in a Bayesian context observing that Eq. (15) is the unnormalized probability distribution of the observed data conditioned on a given model, i.e., the likelihood. The probability distribution of a model given some observed data (i.e., the posterior) is obtained through Bayes theorem, as

$$p(\mathcal{M}|s) \propto \mathcal{N}_{n+g}\int_h e^{-\frac{1}{2}\mathcal{W}_{n+g}(s-h,s-h)}p_h[h|\mathcal{M}]\pi(\mathcal{M}), \tag{42}$$

where $\mathcal{M}$ and $\pi(\mathcal{M})$ are the model and its prior distribution, up to a model independent normalization constant.

The posterior can be estimated with importance sampling using a Monte Carlo Markov chain (MCMC), by generating a sequence of $\mathcal{M}$'s with the probability distribution defined by Eq. (42). In principle each MCMC step would require a nontrivial integration to be performed, and this can also be obtained with a nested importance sampling. This is a nontrivial task, as the estimation of the integral in Eq. (42) has statistical errors roughly proportional to $1/\sqrt{N_s}$, where N_s is the number of the evaluation steps. A trade-off between the amount of knowledge on the probability distribution and the computational cost of the procedure would be required.

A better procedure can be devised by focusing on the integrand of Eq. (42), i.e., the joint posterior on model and non-Gaussian realization h,

$$p(h,\mathcal{M}|s) \propto \mathcal{N}_{n+g}e^{-\frac{1}{2}\mathcal{W}_{n+g}(s-h,s-h)}p_h[h|\mathcal{M}]\pi(\mathcal{M}), \tag{43}$$

and a single sequence of $\mathcal{M}$'s and h's can be generated at the same time with MCMC techniques. This does not solve the computational cost issue, but makes evident that in

principle the model estimation can be improved at will with a large enough number of MCMC steps. Each step can be performed along the lines of the Metropolis-Hastings algorithm as follows:

Step 1: Starting from a model $\mathcal{M}_k$, a new one $\mathcal{M}_{k+1}$ is generated with transition distribution $T(\mathcal{M}_{k+1}|\mathcal{M}_k)$.

Step 2: A realization h_{k+1} is generated accordingly with the distribution $p_h[h_{k+1}|\mathcal{M}_{k+1}]$. In Sec. IV B we provide a well-defined procedure for this purpose.

Step 3: The value of

$$\mathcal{I}_{k+1} = \mathcal{N}_{n+g} e^{-\frac{1}{2}\mathcal{W}_{n+g}(s-h_{k+1}, s-h_{k+1})} \tag{44}$$

is compared with the one evaluated at the previous step, and the new model is accepted with probability

$$\min\left\{1, \frac{\mathcal{I}_{k+1}}{\mathcal{I}_k}\frac{T(\mathcal{M}_{k+1}|\mathcal{M}_k)}{T(\mathcal{M}_k|\mathcal{M}_{k+1})}\right\}. \tag{45}$$

Otherwise, the process is repeated.

Using this approach the prior probability $\pi(\mathcal{M})$ is not considered and can be used later to obtain the posterior. Depending on the framework chosen, the model $\mathcal{M}$ can be constructed to explore a fixed-dimension parametric family of distributions, or it can be tailored to explore models with different dimensions using reversible jump MCMC methods [52]. In addition, $\pi(\mathcal{M})$ can be incorporated simply with the redefinition

$$\mathcal{I}_{k+1} = \mathcal{N}_{n+g} e^{-\frac{1}{2}\mathcal{W}_{n+g}(s-h_{k+1}, s-h_{k+1})} \pi(\mathcal{M}_{k+1}). \tag{46}$$

This can be an advantage in some specific cases when the prior is informative and optimization of the convergence rate is required.

In a generalized approach, Steps 2 and 3 are modified as follows:

Step 2′: A set of N_s sequences $h_{k+1,i}$ are generated accordingly with the distribution $p_h[h_{k+1,i}|\mathcal{M}_{k+1}]$.

Step 3′: The value of

$$\mathcal{I}_{k+1} \equiv \frac{1}{N_s}\sum_i \mathcal{N}_{n+g} e^{-\frac{1}{2}\mathcal{W}_{n+g}(s-h_{k+1,i}, s-h_{k+1,i})} \tag{47}$$

is compared with the one evaluated at the previous step, and the new model is accepted or rejected with the same rule described in Eq. (45).

The likelihood in Eq. (42) is obtained in the limit $N_s \to \infty$, and we can see N_s as a free parameter to tune.

The MCMC sampler favors models with a low value of $\mathcal{W}_{n+g}(s-h, s-h)$, and this can be interpreted as follows: new models are accepted at each MCMC step when they perform better at removing the non-Gaussian part of the signal. When such a part is weak, we can expect that a decision based on a single sequence h could be dominated by statistical fluctuations, so averaging over a large value of N_s provides additional robustness to the algorithm. Note that the convergence of the sampler is guaranteed for each value of N_s.

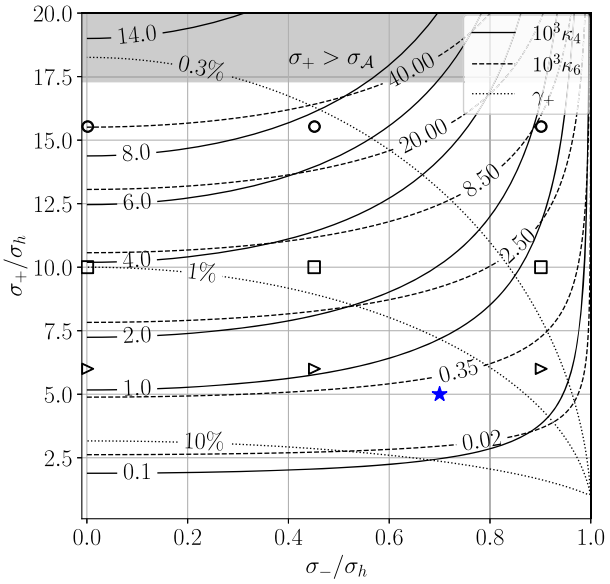

FIG. 1. Cumulant values for the mixture toy model introduced in Sec. III over the full parameter space. Solid and dashed lines denote mixtures with equal fourth- and sixth-order cumulants, k_4 and k_6, respectively. Dotted black lines denote models with equal mixture weights γ_+. The shaded gray region denotes models with the brightest of the two components, γ_+ greater than the noise level of a single detector $\sigma_{\mathcal{A}}$. Any given two black lines intersect only once, hence providing an alternative representation of the full parameter space. Circles (squares, triangles) denote a discrete set of models with various levels of non-Gaussianity. Their DS is characterized in greater detail, with results and signal realizations shown in top (middle, bottom) panels of Fig. 2. The blue star denotes an additional model, exhibiting significant correlations between σ_+ and σ_-. We use this model to characterize the performance of a Bayesian parameter estimation, as described in Sec. III B. As shown by the posterior in Fig. 4, parametrizing the mixture model through its cumulants helps naturally decorrelate them.

As we shall see in Sec. III (see Figs. 1, 3, and 4), the natural parametrization of the non-Gaussian components in terms of higher-order cumulants, along the lines of the Edgeworth or Gram-Charlier A expansions [45,53], appears to also fit in a Bayesian context as it provides parameters inherently decorrelated upon inference. In Sec. IV A we show how this is also a very convenient parametrization for SGWBs characterized by the incoherent superposition of multiple independent signals, with a significant reduction of the computational cost to perform importance sampling. This is a crucial need of our proposed algorithm: Eq. (15) is a sort of "Wiener filter" with stochastic templates. If their sample space is complicated to explore, e.g., when the duty cycle [54] of the background is low, a large number of evaluations might be needed to ensure the algorithm convergence.

III. A TOY MODEL EXAMPLE

Let us consider the very simplified model

$$s_i^{\mathcal{A}} = h_i^{\mathcal{A}} + n_i^{\mathcal{A}}. \tag{48}$$

The noise is modeled by uncorrelated Gaussian variables $n_i^{\mathcal{A}}$ with

$$\langle n_i^{\mathcal{A}} \rangle = 0, \tag{49}$$

$$\langle n_i^{\mathcal{A}} n_j^{\mathcal{B}} \rangle = \sigma_{\mathcal{A}}^2 \delta^{\mathcal{A}\mathcal{B}} \delta_{ij}. \tag{50}$$

We set $h_i^{\mathcal{A}} = h_i$, where h_i are independent variables with probability distribution

$$p(h_i) = \gamma_+ \mathcal{N}(h_i; \sigma_+) + \gamma_- \mathcal{N}(h_i; \sigma_-), \tag{51}$$

$$\gamma_+ = \frac{\sigma_h^2 - \sigma_-^2}{\sigma_+^2 - \sigma_-^2}, \qquad \gamma_- = 1 - \gamma_+ = \frac{\sigma_+^2 - \sigma_h^2}{\sigma_+^2 - \sigma_-^2}. \tag{52}$$

Here $\mathcal{N}(x; \sigma_i)$ is a Gaussian distribution for x with zero mean and variance σ_i^2, and the parametrization of ordered variances $\sigma_+^2 > \sigma_h^2 > \sigma_-^2$ is chosen such that for any values of σ_+, σ_- the distribution variance is σ_h^2. The kurtosis is given by

$$\beta \equiv \left\langle \frac{h_i^4}{\sigma_h^4} \right\rangle = 3\left(\frac{\sigma_+^2}{\sigma_h^2} - \frac{\sigma_+^2}{\sigma_h^2}\frac{\sigma_-^2}{\sigma_h^2} + \frac{\sigma_-^2}{\sigma_h^2} \right), \tag{53}$$

which has a minimum of 3 when $\sigma_+ = \sigma_h$ or $\sigma_- = \sigma_h$ (Gaussian cases with $\gamma_{+,-} = 1$) and grows larger and larger with σ_+. The whole family of leptokurtic probabilities, parametrized by $\sigma_+, \sigma_-, \sigma_h$ can be equivalently explored by three nontrivial cumulants k_n, formally defined by the power expansion of the cumulant generating function K (see Appendix A 6 for more details)

$$K(t) = \log\langle e^{tX} \rangle, \tag{54}$$

$$k_n = \left. \frac{\partial^n K(t)}{\partial t^n} \right|_{t=0}, \quad n = 2, 4, 6. \tag{55}$$

For our toy model they are equal to

$$k_2 = \sigma_h^2, \tag{56}$$

$$k_4 = 3(\sigma_-^2 \sigma_h^2 + \sigma_+^2 \sigma_h^2 - \sigma_+^2 \sigma_-^2 - \sigma_h^4), \tag{57}$$

$$k_6 = 15(\sigma_+^2 - \sigma_h^2)(\sigma_-^2 - \sigma_h^2)(2\sigma_h^2 - \sigma_+^2 - \sigma_-^2). \tag{58}$$

In Fig. 1 we plot contours of constant cumulants k_4, k_6 as a function of the mixture parameters σ_+, σ_-, at a reference value of σ_h, alongside the mixture component weights, uniquely specified by γ_+. It is apparent that the nonlinear relation between k_i and $\sigma_\pm$ could affect significantly the stochastic sampling involved in the Bayesian analysis, while for a frequentist DS it serves only as an alternative parametrization.

Though very simple, this model is expected to capture some features of a realistic non-Gaussian background. For example, the particular case $\sigma_- = 0$ represent backgrounds with burstlike events which are so short that their structure cannot be resolved. One of them (and only one) can be present or not at a given time with a specific probability γ_+, and their amplitude has a Gaussian distribution with standard deviation σ_+, somewhat in the spirit of the analysis in [38]. As only a single event can contribute to the signal at a given time, statistical independence holds: $P(h(t_1), \ldots, h(t_k)) = \prod_k P(h(t_k))$. In a realistic scenario this is not true. Assuming the event waveform has a given shape u_i, the strain at a given time contains contributions from several events. In some peculiar cases it is possible to factorize the probability distribution by using a different domain to describe the signal (e.g., frequency for monochromatic waveforms) but this will be impossible in a generic setup, and the full machinery of point processes [55] described in Sec. IV A should instead be adopted.

A. Frequentist detection

The DS in Eq. (16) can be evaluated analytically for the chosen toy model. As the noise spectrum is white and the signal values across different data points are independent we have [see Eq. (A78) for a detailed proof]

$$\hat{Y}(s) = \sum_i \log \left\langle \exp\left[-\sum_{\mathcal{A}} \frac{h_i(h_i - 2s_i^{\mathcal{A}})}{2\sigma_{\mathcal{A}}^2} \right] \right\rangle. \tag{59}$$

The expectation value can be evaluated explicitly, obtaining

$$\hat{Y}(s) = \sum_i \hat{y}\,[u(s_i)] \tag{60}$$

with $\hat{y}$ a nontrivial function of a single data point

$$\hat{y}(u) = \log\left[\sum_{\alpha=+,-} \frac{\gamma_\alpha \sigma}{\sqrt{\sigma^2 + \sigma_\alpha^2}} \exp\left(\frac{\sigma_\alpha^2 u^2}{2(\sigma^2 + \sigma_\alpha^2)} \right) \right], \tag{61}$$

$$u(s_i) = \sigma \sum_{\mathcal{A}} \frac{s_i^{\mathcal{A}}}{\sigma_{\mathcal{A}}^2}, \tag{62}$$

$$\frac{1}{\sigma^2} \equiv \sum_{\mathcal{A}} \frac{1}{\sigma_{\mathcal{A}}^2}. \tag{63}$$

When the number of data points is large, $\hat{Y}$ becomes a Gaussian variable according to the central limit theorem, so mean and variance suffice to characterize the detection performances.

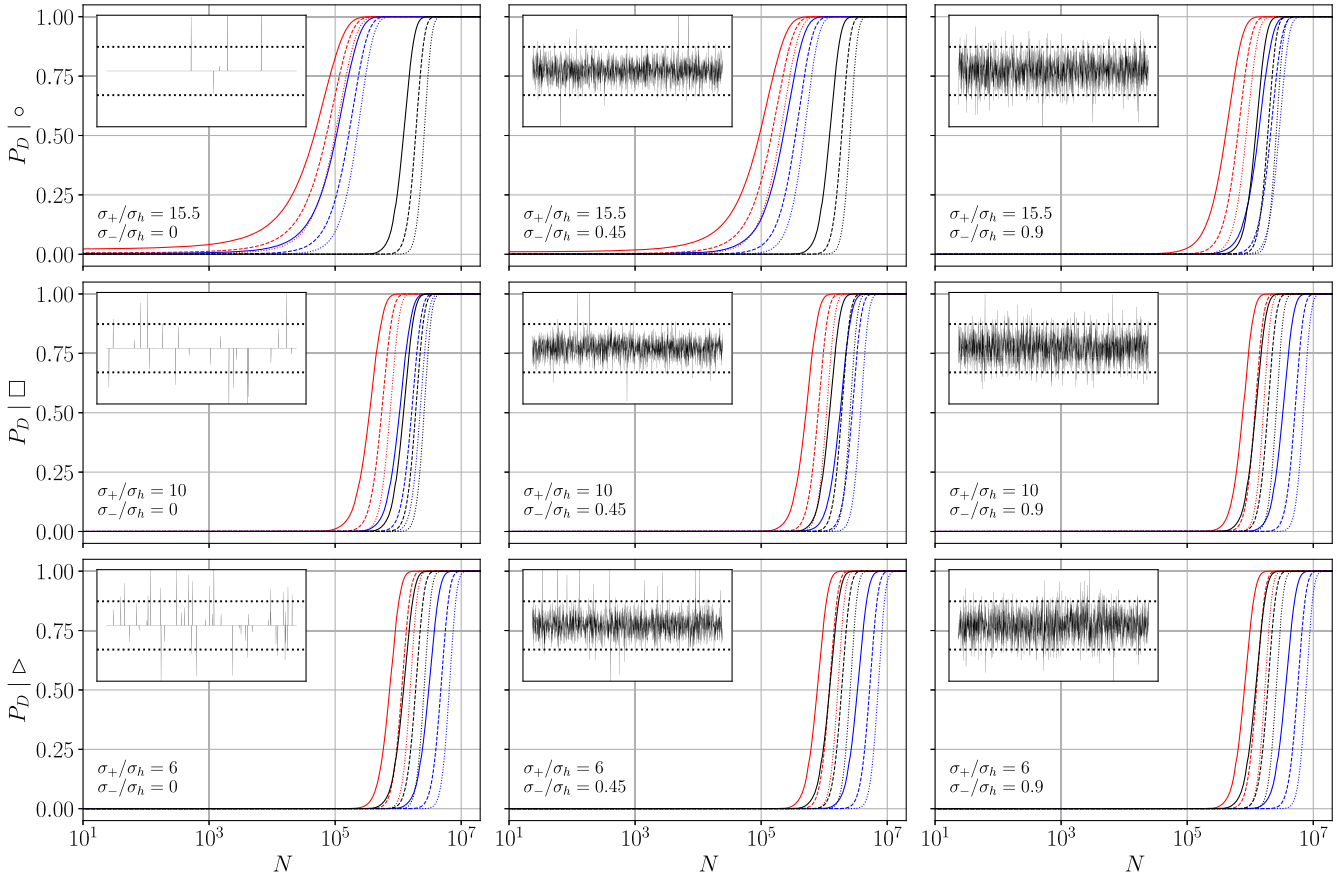

FIG. 2. Performances comparison between DSs for a selection of models across the parameter space in Fig. 1. Black (red, blue) lines denote the probability of detection P_D as a function of the number of data points N for the Gaussian ("optimal," non-Gaussian on scrambled data) DS, i.e., Y_G (Y, $\mathring{Y}$). Solid (dashed, dotted) lines correspond to a probability of false alarm $P_{FA} = 10^{-10}$ (10^{-15}, 10^{-20}). The level $\sigma_h = 0.1$ is kept constant for all models, resulting in an overall shift of the black curves. Values for σ_+/σ_h and σ_-/σ_h are specified in each plot. Top (middle, bottom) row corresponds to models identified with circles (squares, triangles) in Fig. 1 where higher order cumulant values can be recovered. Performances improve as the non-Gaussianity is enhanced, (top-left panel). The non-Gaussian DS on unscrambled data (red lines) outperforms the Gaussian one everywhere in the parameter space, and it performs similarly to it only for signals with small non-Gaussianity (bottom row, corresponding to triangles in Fig. 1). Data scrambling (blue lines) can suppress the advantage of the optimal non-Gaussian DS (red lines) if non-Gaussianity is not high enough. Upper left insets in each subplot show short signal realizations for the respective model in the absence of noise. For reference, detector noise levels $\pm\sigma_{\mathcal{A}}$ are shown as horizontal dashed black lines.

Under the hypothesis $\mathcal{H}_0$ the variable u is by definition normally distributed.

$$p(u) \overset{\mathcal{H}_0}{=} \mathcal{N}(u;1). \tag{64}$$

Under the hypothesis $\mathcal{H}_1$ the expectation value of u is still zero, but the variance gets an additive contribution from the signal. For unscrambled data, we get

$$p(u) \overset{\mathcal{H}_1}{=} \gamma_+ \mathcal{N}\left(u;\sqrt{1+\frac{\sigma_+^2}{\sigma^2}}\right) + \gamma_- \mathcal{N}\left(u;\sqrt{1+\frac{\sigma_-^2}{\sigma^2}}\right) \tag{65}$$

while the equivalent formula for scrambled data is discussed in Appendix A 4.

In both cases we rewrite Eq. (24),

$$P_{\mathrm{D}} = \frac{1}{2}\mathrm{erfc}\left(r_1 \mathrm{erfc}^{-1}(2P_{\mathrm{FA}}) - d_1\sqrt{\frac{N}{2}}\right), \tag{66}$$

where $r_1 = \sigma_{\mathcal{H}_0}/\sigma_{\mathcal{H}_1}$ and $d_1 = (\mu_{\mathcal{H}_1} - \mu_{\mathcal{H}_0})/\sigma_{\mathcal{H}_1}$ can be evaluated easily by numerical integration in the $N = 1$ case.

In Fig. 2 we show the performance of our DS for a discrete set of toy model parameters with various levels of non-Gaussianity. Circles, squares, and triangles identify sets of models with constant σ_+ and varying σ_-. We illustrate the detection probability P_D as a function of the number N of data points, alongside the respective signal

realizations. We do this for three reference false alarm probabilities and for both original and scrambled data.

For comparison, we also show the performance of a Gaussian diagonal-free DS, namely

$$\hat{Y}_G(s) = \sum_i \sum_{\mathcal{A}\neq\mathcal{B}} \frac{s_i^{\mathcal{A}} s_i^{\mathcal{B}}}{\sigma_{\mathcal{A}}^2 \sigma_{\mathcal{B}}^2}, \tag{67}$$

applied to the toy model data. The values of r_1 and d_1 for this particular case are evaluated in Appendix A 4.

As expected, the non-Gaussian DS (without scrambled data) outperforms the scrambled and Gaussian ones. However, as discussed previously the optimal, nonscrambled DS cannot be practically implemented. The relevant performances to look at are those of the scrambled one. It performs better than the Gaussian one for large enough values of k_4 and k_6 (see Fig. 1 where the set of parameters chosen for Fig. 2 is shown). The Gaussian DS being better than the scrambled one for small non-Gaussianity is not unexpected: when we evaluate the scrambled statistics $\mathring{Y}$ we subtract two sets of data [see Eq. (37)] in order to have zero average under $\mathcal{H}_0$. We pay a price for this, introducing additional fluctuations: the variance of the scrambled DS is the sum of the variances evaluated on normal and scrambled data. For small enough values of non-Gaussianity this price is larger than the gain obtained.

Further insight is obtained by introducing a measure of the improvement between $\mathring{Y}$ and $\hat{Y}_G$. A simple possibility is to solve Eq. (66) for N, obtaining $N = N(P_{FA}, P_D, r_1, d_1)$ for a given DS. We evaluate the ratio $N_G/\mathring{N}$ for fixed values of P_D and P_{FA} in the space of toy models' parameters. This is a measure of how much more data one needs to collect to achieve with the Gaussian DS performances similar to those of the scrambled one. The result is shown in Fig. 3. It is evident that a significant advantage can be obtained in the large non-Gaussianity regime. We plot our results for different numbers of detectors in the network, and we observe that large N_D gives improved performance of the scrambled statistics: this is expected because additional fluctuations introduced by the scrambled data do not scale with N_D.

It is clear that the scrambled data subtraction procedure is not optimal, and it is worth exploring alternative options. For example, the cumulant expansion in Eq. (32) could be used to define the generalization of a diagonal–free Gaussian DS, by removing terms not enough under control order by order, i.e., with nonzero expectation values under $\mathcal{H}_0$. This approach is useful especially in the mild non-Gaussian regime, where a truncation in the cumulant expansion is accurate enough. We leave this study to future investigation.

B. Bayesian parameter estimation

The study of the Bayesian procedure with the toy model is simplified by the independence of noise and signal at different times. Taking advantage of it we can write a recursive procedure which, given the posterior distribution for the model given k data, evaluates the posterior distribution when we add the $k+1$ measurement. As we show in Appendix A 5, the likelihood (and subsequently the posterior) can be obtained analytically from Eq. (15) and can be written as the product of likelihoods over individual data points. Explicitly, it reads

$$\mathcal{L}(s_i|\mathcal{M}) \propto \sum_{\alpha=+,-} \frac{\gamma_\alpha}{\sqrt{1+\frac{\sigma_\alpha^2}{\sigma^2}}} e^{-\frac{1}{2}\mathcal{Q}^\alpha_{\mathcal{A}\mathcal{B}} s_i^{\mathcal{A}} s_i^{\mathcal{B}} - \sum_{\mathcal{A}} \log\sqrt{2\pi}\sigma_{\mathcal{A}}}, \tag{68}$$

where $\mathcal{Q}^\alpha_{\mathcal{A}\mathcal{B}}$, proportional to the transverse projector in the detector space, is

$$\mathcal{Q}^\alpha_{\mathcal{A}\mathcal{B}} = \frac{1}{\sigma_{\mathcal{A}}^2 \sigma_{\mathcal{B}}^2}\left(\delta_{\mathcal{A}\mathcal{B}} - \frac{\sigma_\alpha^2}{\sigma^2+\sigma_\alpha^2}\frac{\sigma_{\mathcal{A}}^{-1}\sigma_{\mathcal{B}}^{-1}}{\sigma^{-2}}\right). \tag{69}$$

In Fig. 4 we show the results of an inference performed on a representative model (identified by a blue star in Fig. 1). We perform inference through stochastic nested sampling [56], using the software package CPNEST [57].

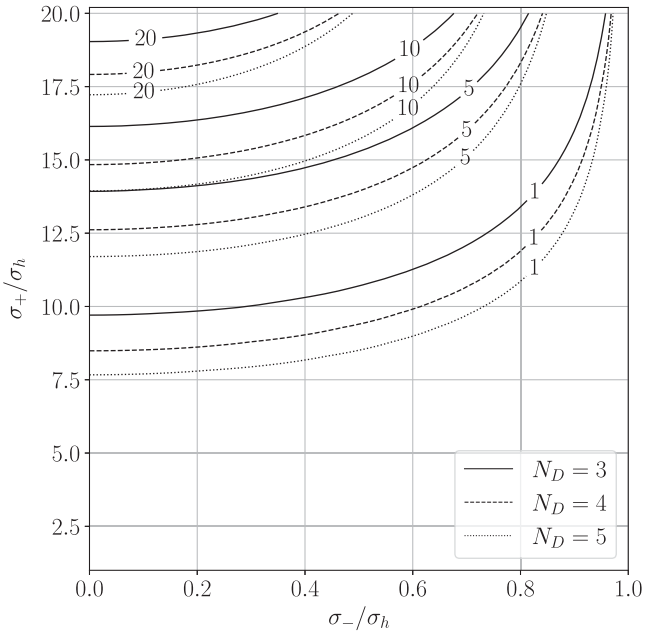

FIG. 3. Performance comparison between the Gaussian DS and the non-Gaussian one on scrambled data, shown across the toy model parameter space. We show contour levels of $N_G/\mathring{N}$, the number of data points required to achieve the same P_D at a fixed P_{FA}. Black, red, and blue solid lines denote contours for a configuration with three, four, and five detectors, respectively. In the high non-Gaussianity limit (top-left corner) the Gaussian DS requires as many as 20 times more data to achieve comparable performances to the non-Gaussian one on scrambled data. On the contrary, for mild non-Gaussianities the two DSs perform similarly. In comparison, increasing the number of detectors improves the non-Gaussian DS performances moderately.

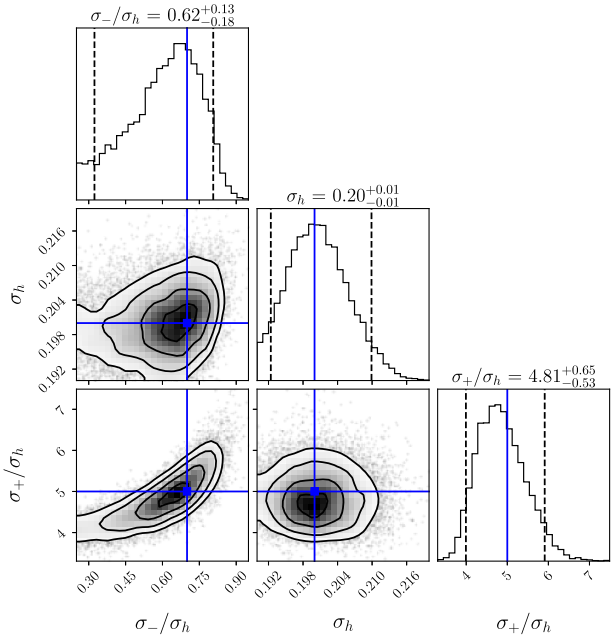

FIG. 4. Corner plot of the Bayesian posterior for the toy model analysis. The noise level is set to $\sigma_{\mathcal{A}} = \sqrt{3}$ with $N_D = 3$, is assumed known upon inference, and is additive to the signal in the data. The number of samples is set to $N = 4 \times 10^5$, which grants the likelihood significant constraining power on the model parameters within the chosen prior. The true signal parameters are shown with solid blue lines and correspond to the point in Fig. 1 labeled with a star symbol. Black dashed lines denote posterior 90% confidence intervals. Priors are uniform for all parameters, and relative ordering is enforced through hypertriangulation [58]. The nonlinear correlation observed in the bottom left subplot matches closely the levels of constant cumulants shown in Fig. 1, which suggest that the cumulant parametrization of the non-Gaussianities would be suitable for an efficient exploration of the parameter space. No predominance of a single cumulant can be identified in the posterior, as expected from contributions in Eq. (32).

We use uniform priors for $\sigma_{-,+}$ and σ_h, and enforce their mutual ordering through hypertriangulation [58]. The detector noise levels $\sigma_{\mathcal{A}}$ are fixed and assumed known. For ease of comparison with Fig. 1 we show posteriors and confidence intervals for the dimensionless parameters σ_+/σ_h and σ_+/σ_h.

Notably, the two-dimensional posterior for $\sigma_{-,+}$ has most of its support along regions of constant cumulants. This suggests that the cumulant parametrization of non-Gaussianities, beyond its naturalness in a statistical sense, is efficient at reducing correlations upon stochastic sampling of the parameter space.

IV. APPLICATION TO ASTROPHYSICAL BACKGROUNDS

An important example of an SGWB exhibiting non-Gaussianity is that of astrophysical origin [59]. The stochastic signal can be modeled as the result of many uncorrelated event superpositions, each event contributing with a well-defined waveform (a function of the source parameters, predictable only in a statistical sense). If there is a strong overlap between these contributions, in a sense that will be defined quantitatively below, the result is a Gaussian background. If this is not the case, non-Gaussian effects appear: the background is no more completely described by its power spectrum, and some additional modeling is required.

A. Point processes

We parametrize the incoherent superposition of multiple signals as a stochastic process

$$h_i^{\mathcal{A}} = \sum_{\sigma=1}^{N} u_i^{\mathcal{A}}(\theta_\sigma), \tag{70}$$

where N is a discrete random variable, describing the number of individual signals for a given realization of h. $u_i^{\mathcal{A}}$ are effective descriptions of gravitational wave signals, as observed by a given detector $\mathcal{A}$. The random *dots* $\{\theta_1, \ldots, \theta_N\}$ describe the intrinsic and extrinsic waveform properties. Such formalism allows one to implement dot distributions and correlations with a high degree of complexity (see [60] for detailed derivations and [61] for a broader introduction to the topic).

For the sake of exposition, we restrict to the time domain and we isolate from θ_σ a parameter τ_σ associated with the random arrangement of the individual signals with respect to the i index (e.g., time of arrivals). The remaining parameters will be referred to as $\hat{\theta}_\sigma$. The statistical generative model reads as follows:

$$N \sim p, \tag{71}$$

$$\{\tau_\sigma\}_{1,\ldots,N} | N \sim Q_N, \tag{72}$$

$$\{\hat{\theta}_\sigma\}_{1,\ldots,N} | N \sim P_N, \tag{73}$$

$$h^{\mathcal{A}}(t) = \sum_{\sigma=1}^{N} u^{\mathcal{A}}(t - \tau_\sigma; \hat{\theta}_\sigma), \tag{74}$$

$$\theta_\sigma = (\tau_\sigma, \hat{\theta}_\sigma). \tag{75}$$

We will consider here a specific case of this model, known in literature as *marked Campbell process*: independent identically distributed dots, characterized by a constant rate ρ for the time domain and a single distribution for $\hat{\theta}_\sigma \sim p_\theta$.

For a realistic background, ρ will be the total rate of all the events that contribute to the signal. The assumption of independent dots means that the events are not correlated, which is generally true for an astrophysical background on the timescale of the experiment if we neglect very peculiar scenarios, e.g., lensing effects. It should be noted that the

formalism is flexible enough to be extended to such scenarios of correlated dots, by replacing ρ with a more complex set of Q_N, P_N [61]. The parameters $\hat{\theta}$ describe the event properties that we are interested in, e.g., their luminosity distance, their sky position, and the intrinsic source parameters. We will employ this machinery to evaluate the $h_i^{\mathcal{A}}$ cumulants

$$\Gamma^{\mathcal{A}_1\cdots\mathcal{A}_n}(t_1,\ldots,t_n) = \langle\langle h^{\mathcal{A}_1}(t_1)\cdots h^{\mathcal{A}_n}(t_n)\rangle\rangle, \quad (76)$$

and we can replace such an ensemble average, using $\langle u^{\mathcal{A}}\rangle_{\hat{\theta}} = 0$, with

$$\Gamma^{\mathcal{A}_1\cdots\mathcal{A}_n}(t_1,\ldots,t_n) = \rho\int\left\langle\prod_{k=1}^{n} u^{\mathcal{A}_k}(t-t_k;\hat{\theta})\right\rangle_{\hat{\theta}} dt. \quad (77)$$

The structure of this expression is quite straightforward to understand: contributions to the cumulants come only from the correlation of an event with itself, as seen by the chosen set of detectors. In principle, the procedure let us obtain a posterior probability distribution for the parameter's model, and upon suitable marginalization, for those of astrophysical interest: e.g., studying a background generated by coalescence events, the mass distribution as a function of redshift z. Remarkably, correlations are not trivial as a consequence of the expectation value taken over the parameters, which makes them nonfactorized. Therefore each cumulant contains nontrivial and independent information about the parameter distributions, and it contributes directly to the inference in Eq. (32). Moreover, it is worth highlighting an interesting scaling relation: scaling simultaneously the rate of events $\rho \to \rho' = \alpha\rho$ and their amplitude $u \to u' = \alpha^{-1/2}u$, cumulants of order n become proportional to $\alpha^{1-n/2}$, i.e., for $n > 2$ become negligible in the large ρ limit while for $n = 2$ they stay constant. This is a simple manifestation of the central limit theorem.

Finally, we stress that our approach uses a population-based construction of relevant cumulants: as a consequence, nonstationary noise contribution (i.e., glitches) can be absorbed in Eq. (70) as an additional population of signals [62]—with different coupling to the detectors—and integrated over in Eq. (77). This is a subject of ongoing study.

B. Importance sampling

The basic ingredient of the proposed approach is an efficient procedure to simulate a background with some target features. As we discussed in Sec. II B the building block is a procedure to generate a sample h with the correct probability $p_h[h_{k+1}|\mathcal{M}_{k+1}]$ conditioned to a model $\mathcal{M}_{k+1}$.

A general parametrization of a given model can be given in terms of the event rate in a given volume of the parameter space, measured in the observer frame. This can be written as

$$\mathcal{R}_0(\hat{\theta})d\hat{\theta}_1\cdots d\hat{\theta}_P. \quad (78)$$

The total rate of events will be given by

$$\rho = \int d\hat{\theta}_1\cdots\int d\hat{\theta}_P\mathcal{R}_0(\hat{\theta}) \quad (79)$$

and using it is possible to simulate dots in a given time interval. Notably, this rate can be very large, and it would be unfeasible to simulate in detail all the events. Instead, it is possible to introduce a threshold on events with negligible contribution to the background. Alternatively, one can include it as a Gaussian contributions to the model. This is, in fact, one of the two reasons for introducing $g_i^{\mathcal{A}}$ in Eq. (1), the second being to include other Gaussian components, e.g., of cosmological origin.

Once the dots are generated, we "decorate" them by choosing a family of suitable individual waveforms and associated parameters according to their distribution $\rho^{-1}\mathcal{R}_0(\hat{\theta})$. Finally the strain h_{ij} is generated, adding all contributions once projected onto each detector.

V. CONCLUSIONS AND PERSPECTIVES

In this paper, we propose a framework to construct detection statistics and perform Bayesian inference for non-Gaussian SGWBs.

The formalism is particularly suitable for stochastic backgrounds arising from the superposition of multiple overlapping sources. We discuss in detail superposition in the time domain, but the approach can be generalized to the frequency domain. We provide a recipe for computing the fundamental quantities required to perform our search in the realistic case of a SGWB of astrophysical origin. We do so by making use of marked Campbell processes. We provide detailed derivations for a number of quantities related to the characterization of DSs performances, which we explore on a subset of representative points on the parameter space.

In the first application to a very simplified toy model, comparative to the standard approach to detection of Gaussian SGWBs, we observe significantly improved performances, in terms of the number of samples (i.e., the observation time or the frequency band) required to reach a target detection significance. As expected, this is milder in the presence of smaller non-Gaussianities.

Our approach is inherently complementary to those available in literature, since it rigorously models the SGWB as a stochastic signal, whose properties arise from the superposition of individual signals: we leverage the knowledge about their distribution and make use of a natural language suited to the purpose, i.e., marked Campbell processes. We argue that the large flexibility attained in the data model through importance sampling motivates further studies on aspects crucial for a realistic application: (i) backgrounds with nontrivial overlap structure: a feature absent in our toy model, subject of ongoing study; (ii) superpositions of multiple backgrounds, as our framework offers a natural way to disentangle them; and

(iii) realistic noise models (nonstationary, correlated across detectors, non-Gaussian), to assess our approach performances compared to the ones in literature.

ACKNOWLEDGMENTS

R. B. thanks E. Buscicchio, F. Di Renzo, C. J. Moore, and G. Brocchi for useful conversations and stimulating observations. R. B. acknowledges support through the Italian Space Agency Grant *Phase A activity for LISA mission, Agreement No. 2017-29-H.0, CUP F62F17000290005*. *Software*: We acknowledge usage of *Mathematica* [63] and of the following PYTHON [64] packages for the analysis, postprocessing, and production of results throughout: CPNEST [57], MATPLOTLIB [65], NUMPY [66], and SCIPY [67].

APPENDIX: DETAILED PROOFS

We expand here on definitions, assumptions, and more detailed derivations for each expression in the paper, section by section. We will omit trivial steps that can be performed easily with most symbolic computation software. Moreover, we will omit full proofs, when a simplified version already contains the interesting concepts. This is frequently the case, e.g., for proofs given for single data point and/or single detector.

1. Definitions and assumptions

The signal is made of the superposition of

$$s_i^{\mathcal{A}} = g_i^{\mathcal{A}} + h_i^{\mathcal{A}} + n_i^{\mathcal{A}}. \tag{A1}$$

The noise and the Gaussian background are distributed as

$$p_n[n_i^{\mathcal{A}}] = \mathcal{N}_n \exp\left(-\frac{1}{2}\mathcal{W}_n(n,n)\right), \tag{A2}$$

$$p_g[g_i^{\mathcal{A}}] = \mathcal{N}_g \exp\left(-\frac{1}{2}\mathcal{W}_g(g,g)\right), \tag{A3}$$

where the quadratic form $\mathcal{W}$ is defined by the sum of scalar products:

$$\mathcal{W}_x(u,v) \equiv \sum_{\mathcal{A},\mathcal{B}} \mathcal{W}_x^{\mathcal{A}\mathcal{B}}(u,v), m \tag{A4}$$

$$\mathcal{W}_x^{\mathcal{A}\mathcal{B}}(u,v) = [\mathbb{C}_{xx}^{-1}]_{ij}^{\mathcal{A}\mathcal{B}} u_i^{\mathcal{A}} v_j^{\mathcal{B}}. \tag{A5}$$

The cross-correlation array defines the normalization and the inner structure of the quadratic form:

$$\mathcal{N}_x = \exp\left(-\frac{1}{2}\mathrm{Tr}\,\ln 2\pi\mathbb{C}_x\right), \tag{A6}$$

$$[\mathbb{C}_{xy}]_{ij}^{\mathcal{A}\mathcal{B}} = \langle x_i^{\mathcal{A}} y_j^{\mathcal{B}}\rangle, \tag{A7}$$

$$\mathbb{C}_x \equiv \mathbb{C}_{xx}. \tag{A8}$$

The trace is performed over detector and data indices, and implicit summation over repeated indices is assumed. We make no assumptions on the distribution of h, $p_h[h_i^{\mathcal{A}}]$.

2. The statistical problem

We first prove Eq. (10):

$$\mathcal{W}_{n+g}(u,v) = \mathcal{W}_n(u,v) - \mathcal{G}(u,v). \tag{A9}$$

This is a straightforward application of the Woodbury identity:

$$(A+UBV)^{-1} = A^{-1} - A^{-1}U(B^{-1}+VA^{-1}U)^{-1}VA^{-1}, \tag{A10}$$

with $U=I$, $V=I$, and $A+B=\mathbb{C}_{n+g}=\mathbb{C}_n+\mathbb{C}_g$, $A=\mathbb{C}_n$. We therefore obtain

$$\mathbb{C}_{n+g}^{-1} = \mathbb{C}_n^{-1} - \mathbb{C}_n^{-1}(\mathbb{C}_g^{-1}+\mathbb{C}_n^{-1})^{-1}\mathbb{C}_n^{-1}, \tag{A11}$$

hence Eq. (10). Now we can prove Eq. (14)

$$p_s[s] = \mathcal{N}_{n+g}\int_h p_h[h]e^{-\frac{1}{2}\mathcal{W}_{n+g}(s-h,s-h)}. \tag{A12}$$

The probability distribution of the data s is specified by the knowledge of its components, and by the conditional probability

$$p_s[s|h,g] = p_n[s-h-g] \tag{A13}$$

through Eq. (A13) we express Eq. (9) as

$$p_s[s] = \int_h\int_g p_s[s|h,g]p_h[h]p_g[g] \tag{A14}$$

$$= \mathcal{N}_n\mathcal{N}_g\int_h\int_g p_h[h]e^{-\frac{1}{2}(\mathcal{W}_n(s-h-g,s-h-g)-\mathcal{W}_g(g,g))}. \tag{A15}$$

The Gaussian integral on g can be performed explicitly:

$$p_s[s] = \int_h p_h[h]\int_g \mathcal{N}_n\mathcal{N}_g\exp\left[-\frac{1}{2}(s-h)^\top\mathbb{C}_n^{-1}(s-h)-\frac{1}{2}g^\top\mathbb{C}_n^{-1}g-\frac{1}{2}g^\top\mathbb{C}_g^{-1}g+(s-h)^\top\mathbb{C}_n^{-1}g\right] \tag{A16}$$

$$= \int_h p_h[h]\mathcal{N}_n\mathcal{N}_g\exp\left[-\frac{1}{2}(s-h)^\top\mathbb{C}_n^{-1}(s-h)\right]\int_g\exp\left[-\frac{1}{2}g^\top[\mathbb{C}_n^{-1}+\mathbb{C}_g^{-1}]g+(s-h)^\top\mathbb{C}_n^{-1}g\right], \tag{A17}$$

where $^\top$ denote transposing with respect to detectors and data indices. Defining

$$A \equiv \mathbb{C}_n^{-1}+\mathbb{C}_g^{-1}, \tag{A18}$$

$$v \equiv A^{-1}\mathbb{C}_n^{-1}(s-h), \tag{A19}$$

one gets

$$\int_h p_h[h]\mathcal{N}_n\mathcal{N}_g\exp\left[-\frac{1}{2}(s-h)^\top\mathbb{C}_n^{-1}(s-h)\right]\int_g\exp\left[-\frac{1}{2}g^\top Ag+v^\top Ag\right] = \tag{A20}$$

$$\int_h p_h[h]\mathcal{N}_n\mathcal{N}_g\exp\left[-\frac{1}{2}(s-h)^\top\mathbb{C}_n^{-1}(s-h)\right]\int_g\exp\left[-\frac{1}{2}(g-v)^\top A(g-v)+\frac{1}{2}v^\top Av\right]. \tag{A21}$$

Integrating over g's with fixed correlation matrix $\mathbb{C}_g$.

$$\int_h p_h[h]\mathcal{N}_n\mathcal{N}_g\exp\left[-\frac{1}{2}(s-h)^\top[\mathbb{C}_n^{-1}-\mathbb{C}_n^{-1}(\mathbb{C}_n^{-1}+\mathbb{C}_g^{-1})^{-1}\mathbb{C}_n^{-1}](s-h)\right]\int_g\exp\left[-\frac{1}{2}(g-v)^\top A(g-v)\right] = \tag{A22}$$

$$\int_h p_h[h] \frac{2\pi^{-\frac{K}{2}}\sqrt{\det(\mathbb{C}_n^{-1})}\sqrt{\det(\mathbb{C}_g^{-1})}}{\sqrt{\det(\mathbb{C}_n^{-1}+\mathbb{C}_g^{-1})}} \exp\left[-\frac{1}{2}(s-h)^\top[\mathbb{C}_n^{-1}-\mathbb{C}_n^{-1}(\mathbb{C}_n^{-1}+\mathbb{C}_g^{-1})^{-1}\mathbb{C}_n^{-1}](s-h)\right] \tag{A23}$$

with K equal to the product between the number of detectors and the number of data points. Using the Woodbury identity, Eq. (A15) becomes

$$p_s[s] = \frac{2\pi^{-\frac{K}{2}}}{\sqrt{\det(\mathbb{C}_g(\mathbb{C}_n^{-1}+\mathbb{C}_g^{-1})\mathbb{C}_n)}} \int_h p_h[h] \exp\left(-\frac{1}{2}(s-h)^\top \mathbb{C}_{n+g}^{-1}(s-h)\right) \tag{A24}$$

$$= \frac{2\pi^{-\frac{K}{2}}}{\sqrt{\det \mathbb{C}_{n+g}}} \int_h p_h[h] \exp\left(-\frac{1}{2}(s-h)^\top \mathbb{C}_{n+g}^{-1}(s-h)\right). \tag{A25}$$

By interpreting the integral as an average over realizations of h distributed according to $p_h[\cdot]$ we obtain Eq. (15).

3. The Neyman-Pearson detection statistic

We focus now on proving Eq. (17). From Eq. (16) we obtain using assumptions from respective hypotheses

$$\log\frac{p_s[s|\mathcal{H}_1]}{p_s[s|\mathcal{H}_0]} = \log\frac{\mathcal{N}_{n+g}}{\mathcal{N}_n} + \log\langle e^{-\frac{1}{2}\mathcal{W}_{n+g}(s,s)} e^{-\frac{1}{2}\mathcal{W}_{n+g}(h,h)} e^{\mathcal{W}_{n+g}(s,h)}\rangle_{\mathcal{H}_1} - \log\langle e^{-\frac{1}{2}\mathcal{W}_n(s,s)}\rangle_{\mathcal{H}_0} \tag{A26}$$

Averaging over h at fixed data s we obtain

$$\log\frac{p_s[s|\mathcal{H}_1]}{p_s[s|\mathcal{H}_0]} = \log\frac{\mathcal{N}_{n+g}}{\mathcal{N}_n} + \log\langle e^{-\frac{1}{2}\mathcal{W}_{n+g}(h,h)} e^{\mathcal{W}_{n+g}(s,h)}\rangle_{\mathcal{H}_1} - \frac{1}{2}(\mathcal{W}_{n+g}(s,s) - \mathcal{W}_n(s,s)) \tag{A27}$$

$$= \log\frac{\mathcal{N}_{n+g}}{\mathcal{N}_n} + \log\langle e^{-\frac{1}{2}\mathcal{W}_{n+g}(h,h)} e^{\mathcal{W}_{n+g}(s,h)}\rangle_{\mathcal{H}_1} + \frac{1}{2}\mathcal{G}(s,s), \tag{A28}$$

hence Eq. (17).

a. Gaussian case

The expansion of the DS reads as follows:

$$\hat{Y}(s) = \frac{1}{2}\mathbb{C}_n^{-1}(\mathbb{C}_n^{-1}+\mathbb{C}_g^{-1})^{-1}\mathbb{C}_n^{-1}$$

$$= \frac{1}{2}\mathbb{C}_n^{-1}(\mathbb{C}_g^{-1}(\mathbb{C}_g\mathbb{C}_n^{-1}+\mathbb{I}))^{-1}\mathbb{C}_n^{-1} \tag{A29}$$

$$= \frac{1}{2}\mathbb{C}_n^{-1}\mathbb{C}_g\mathbb{C}_n^{-1} + \mathcal{O}(\|\mathbb{C}_g\mathbb{C}_n^{-1}\|^2). \tag{A30}$$

As in the main text, we start from the DS in Eq. (18):

$$\hat{Y}(s) \simeq \left[\frac{1}{2}\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\right]_{ij}^{\mathcal{AB}} s_i^{\mathcal{A}} s_j^{\mathcal{B}} = \check{\mathbb{A}}_{ij}^{\mathcal{AB}} s_i^{\mathcal{A}} s_j^{\mathcal{B}}. \tag{A31}$$

We employ here estimates of $\mathbb{C}_n$, $\mathbb{C}_g$ labeled with a $\check{}$ symbol. They contain our prior knowledge about the noise and the Gaussian signal. The mean of $\hat{Y}$ under $\mathcal{H}_0$ reads

$$\mu_{\mathcal{H}_0} = \frac{1}{2}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}]_{ij}^{\mathcal{AB}}\langle n_i^{\mathcal{A}} n_j^{\mathcal{B}}\rangle = \mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n]. \tag{A32}$$

Similarly for the variance (we drop the detector indices because they follow the same contractions as the data indices)

$$\sigma^2_{\mathcal{H}_0} = \check{\mathbb{A}}^{\mathcal{AB}}_{ij}\check{\mathbb{A}}^{\mathcal{CD}}_{kl}\langle n^{\mathcal{A}}_i n^{\mathcal{B}}_j n^{\mathcal{C}}_k n^{\mathcal{D}}_l\rangle - (\mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n])^2 \tag{A33}$$

$$= \check{\mathbb{A}}_{ij}\check{\mathbb{A}}_{kl}([\mathbb{C}_n]_{ik}[\mathbb{C}_n]_{jk} + [\mathbb{C}_n]_{il}[\mathbb{C}_n]_{jk}) \tag{A34}$$

$$= \frac{1}{2}\mathrm{Tr}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_n\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_n]. \tag{A35}$$

Similarly under $\mathcal{H}_1$ (using in addition $\langle n^{\mathcal{A}}_i g^{\mathcal{B}}_j\rangle = 0$)

$$\mu_{\mathcal{H}_1} = \frac{1}{2}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}]^{\mathcal{AB}}_{ij}\langle(n+g)^{\mathcal{A}}_i(n+g)^{\mathcal{B}}_j\rangle \tag{A36}$$

$$= \mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n] + \mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_g] \tag{A37}$$

$$= \mu_{\mathcal{H}_0} + \frac{1}{2}\mathrm{Tr}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_g] \tag{A38}$$

and for the variance

$$\sigma^2_{\mathcal{H}_1} = \check{\mathbb{A}}^{\mathcal{AB}}_{ij}\check{\mathbb{A}}^{\mathcal{CD}}_{kl}\langle(n+g)^{\mathcal{A}}_i(n+g)^{\mathcal{B}}_j(n+g)^{\mathcal{C}}_k(n+g)^{\mathcal{D}}_l\rangle \tag{A39}$$

$$= \check{\mathbb{A}}^{\mathcal{AB}}_{ij}\check{\mathbb{A}}^{\mathcal{CD}}_{kl}[\langle n_i n_j n_k n_l\rangle + \langle n_i n_j g_k g_l\rangle + \langle n_i g_j g_k n_l\rangle + \langle n_i g_j n_k g_l\rangle + (\text{“}n\text{”} \leftrightarrow \text{“}g\text{”})] - \mu^2_{\mathcal{H}_1}. \tag{A40}$$

All terms with an odd number of n's cancel out after averaging (both g and n are multivariate Gaussians). The fourth order averages simplify through Isserlis theorem to

$$\begin{aligned}\sigma^2_{\mathcal{H}_1} = {} & \check{\mathbb{A}}_{ij}\check{\mathbb{A}}_{kl}[\langle n_i n_j\rangle\langle n_k n_l\rangle + \langle n_i n_k\rangle\langle n_j n_l\rangle + \langle n_i n_l\rangle\langle n_k n_j\rangle] \\ & + \check{\mathbb{A}}_{ij}\check{\mathbb{A}}_{kl}[\langle n_i n_j\rangle\langle g_k g_l\rangle] + \check{\mathbb{A}}_{ij}\check{\mathbb{A}}_{kl}[\langle n_i n_l\rangle\langle g_k g_j\rangle] \\ & + \check{\mathbb{A}}_{ij}\check{\mathbb{A}}_{kl}[\langle n_i n_k\rangle\langle g_j g_l\rangle] + (\text{“}n\text{”} \leftrightarrow \text{“}g\text{”}) - \mu^2_{\mathcal{H}_1}.\end{aligned} \tag{A41}$$

Cancellations are again due to the uncorrelatedness and zero mean of the two series. Upon contraction the expression simplifies to

$$\sigma^2_{\mathcal{H}_1} = \mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n]^2 + 2\mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n\check{\mathbb{A}}\mathbb{C}_n] + \mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n]\mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_g] + 2\mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_g\check{\mathbb{A}}\mathbb{C}_n] + (\text{”}n\text{”} \leftrightarrow \text{”}g\text{”}) - \mu^2_{\mathcal{H}_1} \tag{A42}$$

$$= \sigma^2_{\mathcal{H}_0} + 4\mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n\check{\mathbb{A}}\mathbb{C}_g] + \mathcal{O}(\|\mathbb{C}_g\|^2) \simeq \sigma^2_{\mathcal{H}_0} + \mathrm{Tr}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_n\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_g]. \tag{A43}$$

Equations (A32), (A35), (A38), and (A43) prove results from the main text.

b. Gaussian diagonal-free case

Assuming uncorrelated noises across detectors (i.e., $\check{\mathbb{C}}^{\mathcal{AB}}_n \propto \delta^{\mathcal{AB}}$), we subtract by hand the diagonal terms from the statistics. We label the two “diagonal-free” hypotheses $\mathcal{H}_{0,\mathrm{G}}, \mathcal{H}_{1,\mathrm{G}}$, and we have

$$\hat{Y}(s) = \frac{1}{2}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}]^{\mathcal{AB}}_{ij}s^{\mathcal{A}}_i s^{\mathcal{B}}_j \to \hat{Y}_G(s) = \frac{1}{2}[\check{\mathbb{C}}_n^{-1}]^{\mathcal{AC}}_{ik}[\check{\mathbb{C}}_g]^{\mathcal{CD}}_{kl}[\check{\mathbb{C}}_n^{-1}]^{\mathcal{DB}}_{lj}s^{\mathcal{A}}_i s^{\mathcal{B}}_j(1-\delta^{\mathcal{AB}}). \tag{A44}$$

Therefore we obtain a robust cross-correlation statistic (although not necessarily optimal) with the following properties:

$$\mu_{\mathcal{H}_{0,\mathrm{G}}} \propto \frac{1}{2}[\check{\mathbb{C}}_n^{-1}]^{\mathcal{AC}}_{ik}[\check{\mathbb{C}}_g]^{\mathcal{CD}}_{kl}[\check{\mathbb{C}}_n^{-1}]^{\mathcal{DB}}_{lj}\delta^{\mathcal{AB}}(1-\delta^{\mathcal{AB}}) = 0, \tag{A45}$$

$$\mu_{\mathcal{H}_{1,\mathrm{G}}} = \frac{1}{2}[\check{\mathbb{C}}_n^{-1}]^{\mathcal{AC}}_{ik}[\check{\mathbb{C}}_g]^{\mathcal{CD}}_{kl}[\check{\mathbb{C}}_n^{-1}]^{\mathcal{DB}}_{lj}\mathbb{C}^{\mathcal{AB}}_g(1-\delta^{\mathcal{AB}}) \tag{A46}$$

$$= \frac{1}{2}\mathrm{Tr}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g], \tag{A47}$$

where $\check{\mathbb{C}}_g^{\mathcal{AB}} = \mathbb{C}_g^{\mathcal{AB}}(1-\delta^{\mathcal{AB}})$ defines a "diagonal-free" signal correlation. Since the noise is diagonal across detector indices, and detectors can have heterogeneous spectra, we write

$$[\check{\mathbb{C}}_n^{-1}]^{\mathcal{AC}} = \sum_\epsilon c^\epsilon \delta_\epsilon^{\mathcal{AC}} \tag{A48}$$

$$\delta_\epsilon^{\mathcal{AC}} \equiv \begin{cases} 1 & \epsilon = \mathcal{A} = \mathcal{C} \\ 0 & \text{otherwise} \end{cases}. \tag{A49}$$

Then the DS becomes

$$\hat{Y}_G(s) = \frac{1}{2}c_{ik}^\epsilon \delta_\epsilon^{\mathcal{AC}}[\check{\mathbb{C}}_g]_{kl}^{\mathcal{CD}} c_{lj}^\delta \delta_\delta^{\mathcal{DB}}(1-\delta^{\mathcal{AB}}) s_i^{\mathcal{A}} s_j^{\mathcal{B}}. \tag{A50}$$

Diagonal terms of $\check{\mathbb{C}}_g$ equal zero because $\delta_\epsilon^{\mathcal{AC}}\delta_\delta^{\mathcal{DB}}(1-\delta^{\mathcal{AB}}) = 0$ for $\mathcal{C} = \mathcal{D}$. Therefore we are free to subtract them,

$$\hat{Y}_G(s) = \frac{1}{2}c_{ik}^\epsilon \delta_\epsilon^{\mathcal{AC}}[\check{\mathbb{C}}_g]_{kl}^{\mathcal{CD}}(1-\delta^{\mathcal{CD}}) c_{lj}^\delta \delta_\delta^{\mathcal{DB}}(1-\delta^{\mathcal{AB}}) s_i^{\mathcal{A}} s_j^{\mathcal{B}}, \tag{A51}$$

For the same reason we can add them back in the rightmost factor, i.e., neglecting $(1-\delta^{\mathcal{AB}})$, because its effect is now taken care of by $\delta^{\mathcal{CD}}$. In conclusion,

$$\hat{Y}_G(s) = \frac{1}{2}c_{ik}^\epsilon \delta_\epsilon^{\mathcal{AC}}[\check{\mathbb{C}}_g]_{kl}^{\mathcal{CD}}(1-\delta^{\mathcal{CD}}) c_{lj}^\delta \delta_\delta^{\mathcal{DB}} s_i^{\mathcal{A}} s_j^{\mathcal{B}} \tag{A52}$$

$$= \frac{1}{2}[\check{\mathbb{C}}_n^{-1}]_{ik}^{\mathcal{AC}}[\check{\mathbb{C}}_g]_{kl}^{\mathcal{CD}}[\check{\mathbb{C}}_n^{-1}]_{lj}^{\mathcal{DB}} s_i^{\mathcal{A}} s_j^{\mathcal{B}}. \tag{A53}$$

Therefore we can equivalently neglect the diagonal in our modeled signal cross-correlation $\check{\mathbb{C}}_g$ or in the product of realizations $s_i^{\mathcal{A}} s_j^{\mathcal{B}}$. Consequently, with obvious definition

$$\check{\mathbb{A}}_{ij}^{\mathcal{AB}} = \frac{1}{2}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}]_{ij}^{\mathcal{AB}} \tag{A54}$$

we obtain

$$\mu_{\mathcal{H}_{0,\mathrm{G}}} = 0, \tag{A55}$$

$$\mu_{\mathcal{H}_{1,\mathrm{G}}} = \mathrm{Tr}[\check{\mathbb{A}}\check{\mathbb{C}}_g] = \mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_g]. \tag{A56}$$

For the variances

$$\sigma^2_{\mathcal{H}_{0,\mathrm{G}}} = \check{\mathbb{A}}_{ij}^{\mathcal{AB}}\check{\mathbb{A}}_{kl}^{\mathcal{CD}}\langle n_i^{\mathcal{A}} n_j^{\mathcal{B}} n_k^{\mathcal{C}} n_l^{\mathcal{D}}\rangle - (\mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n])^2 \tag{A57}$$

$$= \frac{1}{4}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}]_{ij}^{\mathcal{AB}}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}]_{kl}^{\mathcal{CD}}([\mathbb{C}_n]_{ik}^{\mathcal{AC}}[\mathbb{C}_n]_{jl}^{\mathcal{BD}} + [\mathbb{C}_n]_{il}^{\mathcal{AD}}[\mathbb{C}_n]_{jk}^{\mathcal{BC}}) \tag{A58}$$

$$= 2\mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n\check{\mathbb{A}}\mathbb{C}_n], \tag{A59}$$

$$\sigma^2_{\mathcal{H}_{1,\mathrm{G}}} = \sigma^2_{\mathcal{H}_{0,\mathrm{G}}} + \mathrm{Tr}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_n\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_g] + \frac{1}{2}\mathrm{Tr}[\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_n\check{\mathbb{C}}_n^{-1}\check{\mathbb{C}}_g\check{\mathbb{C}}_n^{-1}\mathbb{C}_g] \tag{A60}$$

$$= \sigma^2_{\mathcal{H}_{0,\mathrm{G}}} + 4\mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n\check{\mathbb{A}}\mathbb{C}_g] + \mathcal{O}(\|\mathbb{C}_g\|^2), \tag{A61}$$

and hence Eqs. (27), (28), (29), and (30). It is worth noting that $\sigma_{\mathcal{H}_0}$, $\sigma_{\mathcal{H}_1}$ and $\sigma_{\mathcal{H}_{0,\mathrm{G}}}$, $\sigma_{\mathcal{H}_{1,\mathrm{G}}}$ are, respectively, identical in functional form, with the substitution of $\mathbb{C}_g$ with its diagonal–free version $\mathscr{C}_g$.

c. Non-Gaussian case

We now turn our attention to the non-Gaussian case. By defining it as in the main text $\mathfrak{s}_i^{\mathcal{A}} = [\check{\mathbb{C}}_n^{-1}]_{ij}^{\mathcal{AB}} s_j^{\mathcal{B}}$, direct substitution in the general expression of Eq. (17) yields

$$\hat{Y}(s) = \frac{1}{2}\mathcal{G}(s,s) + \log\langle e^{-\frac{1}{2}\mathcal{W}_{n+g}(h,h)} e^{\mathcal{W}_{n+g}(s,h)}\rangle \tag{A62}$$

$$= \frac{1}{2}[\check{\mathbb{C}}_g]_{ij}^{\mathcal{AB}}\mathfrak{s}_i^{\mathcal{A}}\mathfrak{s}_j^{\mathcal{B}} + \log\chi_h \int_h p'_h[h] e^{\mathcal{W}_{n+g}(s,h)}, \tag{A63}$$

where we have defined a (normalized) probability distribution

$$p'_h[h] \equiv \chi_h^{-1} \exp\left[-\frac{1}{2}\mathcal{W}_{n+g}(h,h)\right] p_h[h] \tag{A64}$$

$$\chi_h \equiv \langle e^{-\frac{1}{2}\mathcal{W}_{n+g}(h,h)}\rangle. \tag{A65}$$

Now, focusing on

$$\log\chi_h \int_h p'_h[h] e^{\mathcal{W}_{n+g}(s,h)}, \tag{A66}$$

we expand $\mathcal{W}_{n+g}$

$$e^{\mathcal{W}_n(s,h)-\mathcal{G}(s,h)} = \exp[\mathfrak{s}_i^{\mathcal{A}} h_i^{\mathcal{A}} - \mathfrak{s}_i^{\mathcal{A}}[\check{\mathbb{C}}_g]_{ij}^{\mathcal{AB}}[\check{\mathbb{C}}_n^{-1}]_{jk}^{\mathcal{BC}} h_k^{\mathcal{C}}]. \tag{A67}$$

We stress here again an important point: the separation of the signal into a "Gaussian" and a "non-Gaussian" component is somewhat arbitrary. If we choose to set $g = 0$, the entirety of the gravitational wave signal is described by h. The double-whitened data points $\mathfrak{s}$ are not affected by this change, while Eq. (A63) becomes

$$\hat{Y}(s) = \log\left[\langle e^{-\frac{1}{2}\mathcal{W}_n(h,h)}\rangle \int_h p_h[h] e^{-\frac{1}{2}\mathcal{W}_n(h,h)} e^{\mathfrak{s}_i^{\mathcal{A}} h_i^{\mathcal{A}}}\right]. \tag{A68}$$

Therefore the rightmost term in Eq. (A63) is the generating function of *of the non-Gaussian component* "connected moments" (or "cumulants"), with $\mathfrak{s}$ acting as the auxiliary variable, and realizations h distributed according to $p_h[h] e^{-\frac{1}{2}\mathcal{W}_n(h,h)}$,

$$\hat{Y}(s) = \log\langle e^{\mathfrak{s}_i^{\mathcal{A}} h_i^{\mathcal{A}}}\rangle. \tag{A69}$$

Therefore by definition we can rewrite it as a power series in $\mathfrak{s}$,

$$\hat{Y}(s) = \frac{1}{2}[\mathbb{C}_g]_{ij}^{\mathcal{AB}}\mathfrak{s}_i^{\mathcal{A}}\mathfrak{s}_j^{\mathcal{B}} + \chi_h \sum_{n=1}^{\infty} \frac{1}{n!} \check{\Gamma}_{i_1\cdots i_n}^{\mathcal{A}_1\cdots\mathcal{A}_n} \mathfrak{s}_{i_1}^{\mathcal{A}_1} \cdots \mathfrak{s}_{i_n}^{\mathcal{A}_n}. \tag{A70}$$

This proves the general expression for the non-Gaussian DS in Eq. (32). Evaluating the expectation values under both hypotheses, we get Eqs. (35) for $\mathcal{H}_0$,

$$\mu_{\mathcal{H}_0} = \mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n] + \chi_h \sum_{n=1}^{\infty} \frac{1}{n!} \check{\Gamma}_{i_1\cdots i_n}^{\mathcal{A}_1\cdots\mathcal{A}_n} \overline{\mathfrak{s}_{i_1}^{\mathcal{A}_1} \cdots \mathfrak{s}_{i_n}^{\mathcal{A}_n}} \tag{A71}$$

$$= \mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n] + \chi_h \sum_{n=1}^{\infty} \frac{1}{n!} \check{\Gamma}^{\mathcal{A}_1\cdots\mathcal{A}_n}_{i_1\cdots i_n} \mathbb{N}^{\mathcal{A}_1\cdots\mathcal{A}_n}_{i_1\cdots i_n}. \tag{A72}$$

Using the Isserlis theorem we express the $\mathbb{N}$'s as products of over all pairs of indices (i.e., over $[\check{\mathbb{C}}_n^{-1}\mathbb{C}_n\check{\mathbb{C}}_n^{-1}]^{\mathcal{A}\mathcal{B}}_{ij}$). Terms with odd n cancel out to zero. Out of $n = 2m$ indices, we get $(2m-1)!!$ contractions in couples (the number of complete graphs with $2m$ vertices), which due to the symmetry of the Γ's contribute identically after full contraction:

$$\mu_{\mathcal{H}_0} = \mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n] + \chi_h \sum_{\substack{m=1 \\ n=2m}}^{\infty} \frac{(n-1)!!}{n!} \check{\Gamma}^{\mathcal{A}_1\cdots\mathcal{A}_n}_{i_1\cdots i_n} [\check{\mathbb{C}}_n^{-1}\mathbb{C}_n\check{\mathbb{C}}_n^{-1}]^{\mathcal{A}_1\mathcal{A}_2}_{i_1 j_2} \cdots [\check{\mathbb{C}}_n^{-1}\mathbb{C}_n\check{\mathbb{C}}_n^{-1}]^{\mathcal{A}_{n-1}\mathcal{A}_n}_{i_{n-1} j_n} \tag{A73}$$

$$= \mathrm{Tr}[\check{\mathbb{A}}\mathbb{C}_n] + \chi_h \sum_{\substack{m=1 \\ n=2m}}^{\infty} \frac{1}{n!!} \check{\Gamma}^{\mathcal{A}_1\cdots\mathcal{A}_n}_{i_1\cdots i_n} [\check{\mathbb{C}}_n^{-1}\mathbb{C}_n\check{\mathbb{C}}_n^{-1}]^{\mathcal{A}_1\mathcal{A}_2}_{i_1 j_2} \cdots [\check{\mathbb{C}}_n^{-1}\mathbb{C}_n\check{\mathbb{C}}_n^{-1}]^{\mathcal{A}_{n-1}\mathcal{A}_n}_{i_{n-1} j_n}, \tag{A74}$$

and hence Eq. (36). Results for the DS with "scrambled data" [Eqs. (40) and (41)] are proven directly in the main text.

4. Toy model derivations

We show here the detailed derivations of the quantities related to the DS $\hat{Y}(s)$ for the toy model. We start from Eq. (16), and we drop the Gaussian component. Moreover, we account for s_i, h_i being independent and diagonal across detectors, so we can factorize the $\langle\cdots\rangle$ into the product of expectation values for each h_i, obtaining

$$\hat{Y}(s) = \log\left\langle \exp\left[-\frac{1}{2}\mathcal{W}_n(h,h) + \mathcal{W}_n(s,h)\right]\right\rangle \tag{A75}$$

$$= \log\left\langle \exp\left[-\frac{1}{2}h_i^{\mathcal{A}}h_j^{\mathcal{B}}[\mathbb{C}_n^{-1}]^{ij}_{\mathcal{A}\mathcal{B}} + h_k^{\mathcal{C}}s_l^{\mathcal{D}}[\mathbb{C}_n^{-1}]^{kl}_{\mathcal{C}\mathcal{D}}\right]\right\rangle \tag{A76}$$

$$= \log\left\langle \prod_i \exp\left[-\frac{1}{2}\frac{h_i^{\mathcal{A}}h_j^{\mathcal{B}}}{\sigma_{\mathcal{A}}^2}\delta_{\mathcal{A}\mathcal{B}}\delta^{ij} + \frac{h_k^{\mathcal{C}}s_l^{\mathcal{D}}}{\sigma_{\mathcal{A}}^2}\delta_{\mathcal{C}\mathcal{D}}\delta^{kl}\right]\right\rangle \tag{A77}$$

$$= \sum_i \log\left\langle \exp\left[\sum_{\mathcal{A}} -\frac{h_i(h_i - 2s_i^{\mathcal{A}})}{2\sigma_{\mathcal{A}}^2}\right]\right\rangle, \tag{A78}$$

where $\langle\cdots\rangle$ in the last line is performed over a single data point h_i, but over multiple detectors. By using standard results on Gaussian integrals and defining

$$A_\alpha = \frac{1}{2\sigma_\alpha^2} + \sum_{\mathcal{A}} \frac{1}{2\sigma_{\mathcal{A}}^2}, \tag{A79}$$

$$B_i = \sum_{\mathcal{A}} \frac{s_i^{\mathcal{A}}}{\sigma_{\mathcal{A}}^2}, \tag{A80}$$

we obtain

$$\hat{Y}(s) = \sum_i \log\left[\sum_{\alpha=+,-} \frac{\gamma_\alpha}{\sqrt{2\pi\sigma_\alpha^2}} \int \mathrm{d}h_i \exp\left[-A_\alpha h_i^2 + B_i h_i\right]\right] = \sum_i \hat{y}\left(\sigma \sum_{\mathcal{A}} \frac{s_i^{\mathcal{A}}}{\sigma_{\mathcal{A}}^2}\right) \tag{A81}$$

with

$$\hat{y}(u) = \log\left[\sum_{\alpha=+,-} \frac{\gamma_\alpha \sigma}{\sqrt{\sigma_\alpha^2 + \sigma^2}} \exp\left[\frac{\sigma_\alpha^2}{2(\sigma_\alpha^2 + \sigma^2)}u^2\right]\right] \tag{A82}$$

$$\frac{1}{\sigma^2} = \sum_{\mathcal{A}} \frac{1}{\sigma_{\mathcal{A}}^2}, \tag{A83}$$

and hence Eqs. (60) to (63). The basic building blocks for estimating the performances of the toy model Neyman-Pearson DS are the means and variances of the $\hat{y}$ statistic. It is a nontrivial function of a single scalar, combination of all the detector signals, under both hypotheses. Evaluating them can be achieved numerically as follows.

a. Nonscrambled data

Under the $\mathcal{H}_0$ hypothesis we need

$$\mu_0 = \langle \hat{y}(u(s)) \rangle, \tag{A84}$$

$$\sigma_0^2 = \langle \hat{y}(u(s))^2 \rangle - \mu_0^2, \tag{A85}$$

where

$$u(s) = \sigma \sum_{\mathcal{A}} \frac{n_i^{\mathcal{A}}}{\sigma_{\mathcal{A}}^2} \tag{A86}$$

is a Gaussian variable with zero mean and unit variance. Under $\mathcal{H}_1$ the $u(s)$ becomes

$$u(s) = \sigma \sum_{\mathcal{A}} \frac{n_i^{\mathcal{A}}}{\sigma_{\mathcal{A}}^2} + \frac{h_i}{\sigma}, \tag{A87}$$

the sum of a normal variable (the first term) and a (scaled) variable distributed according to the mixture model. So the overall distribution is given by

$$p_1(u) = \gamma_+ \mathcal{N}\left(u, \sqrt{1 + \frac{\sigma_+^2}{\sigma^2}}\right) + \gamma_- \mathcal{N}\left(u, \sqrt{1 + \frac{\sigma_-^2}{\sigma^2}}\right). \tag{A88}$$

Therefore we get

$$\mu_0 = \int du \hat{y}(u) \mathcal{N}(u, 1), \tag{A89}$$

$$\sigma_0^2 = \int du \hat{y}(u)^2 \mathcal{N}(u, 1) - \mu_0^2, \tag{A90}$$

$$\mu_1 = \int du \hat{y}(u) p_1(u), \tag{A91}$$

$$\sigma_1^2 = \int du \hat{y}(u)^2 p_1(u) - \mu_1^2, \tag{A92}$$

which can easily be evaluated numerically.

b. Scrambled data

The DS is

$$\mathring{Y}(s) = \hat{Y}(s) - \hat{Y}(\mathring{s}) \tag{A93}$$

$$= \sum_i \hat{y}(u(s_i)) - \hat{y}(u(\mathring{s}_i)) \tag{A94}$$

$$= \sum_i \hat{y}(u(s_i)) - \sum_i \hat{y}(u(\mathring{s}_i)). \tag{A95}$$

We need to evaluate the mean and the variance of the DS. Under the hypothesis $\mathcal{H}_0$ we get

$$u(\mathring{s}_i)|\mathcal{H}_0 = \sigma \sum_{\mathcal{A}} \frac{\mathring{n}_i^{\mathcal{A}}}{\sigma_{\mathcal{A}}^2}, \tag{A96}$$

and scrambling the noise realizations makes them uncorrelated across detectors; however, the nonscrambled ones were already so. So $u(\mathring{s}_i)$ is a zero-mean unit variance variable, and therefore when evaluating the differences in (A94), we have

$$\mathring{\mu}_0 = \langle \hat{y}(u(s)) \rangle - \langle \hat{y}(u(\mathring{s})) \rangle. \tag{A97}$$

The two averages are identical, as $\mathring{s}$ and s are identically distributed. Therefore $\mathring{\mu}_0 = 0$ as expected. Computing explicitly the variance

$$\mathring{\sigma}_0^2 = \langle (\hat{y}(u(s)) - \hat{y}(u(\mathring{s})))^2 \rangle - \mathring{\mu}_0^2 \tag{A98}$$

$$= \langle \hat{y}(u(s))^2 \rangle + \langle \hat{y}(u(\mathring{s}))^2 \rangle - 2\langle \hat{y}(u(s)) \rangle \langle \hat{y}(u(\mathring{s})) \rangle \tag{A99}$$

$$= \sigma_0^2 + \mu_0^2 + \sigma_0^2 + \mu_0^2 - 2\mu_0^2 \tag{A100}$$

$$= 2\sigma_0^2, \tag{A101}$$

where in Eq. (A101) we used the statistical independence by construction of the scrambled data $\mathring{s}$ upon the initial ones. Under the hypothesis $\mathcal{H}_1$ we get

$$\mathring{\mu}_1 = \langle \hat{y}(u(s)) \rangle - \langle \hat{y}(u(\mathring{s})) \rangle, \tag{A102}$$

where now

$$u(\mathring{s}) = \sigma \sum_{\mathcal{A}} \frac{\mathring{n}_i^{\mathcal{A}}}{\sigma_{\mathcal{A}}^2} + \sigma \sum_{\mathcal{A}} \frac{\mathring{h}_i^{\mathcal{A}}}{\sigma_{\mathcal{A}}^2}. \tag{A103}$$

The first term is a normal Gaussian variable, as before. In the second term, each $x = \sigma \mathring{h}_i^{\mathcal{A}}/\sigma_{\mathcal{A}}^2$ is a different realization, distributed according to

$$p_{\mathcal{A}}(x) = \gamma_+ \mathcal{N}\left(x, \frac{\sigma\sigma_+}{\sigma_{\mathcal{A}}^2}\right) + \gamma_- \mathcal{N}\left(x, \frac{\sigma\sigma_-}{\sigma_{\mathcal{A}}^2}\right), \tag{A104}$$

so the final distribution is given by the overall convolution

$$\mathring{p}(u) = \mathcal{N}(u, 1) \star p_1(u) \star \cdots \star p_{N_D}(u), \tag{A105}$$

where N_D is the number of detectors available. This can be expressed as a sum of Gaussian distributions, remembering that

$$\mathcal{N}(u, \sigma_1) \star \mathcal{N}(u, \sigma_2) = \mathcal{N}\left(u, \sqrt{\sigma_1^2 + \sigma_2^2}\right). \tag{A106}$$

In closed form, it reads

$$\mathring{p}(u) = \sum_{k=0}^{N_D} \gamma_+^k \gamma_-^{N_D - k} \sum_{\mathsf{s} \in \mathcal{S}_k} \mathcal{N}\left(u, \sqrt{1 + \sigma^2 \sum_{\mathcal{A}=1}^{N_D} \frac{\mathsf{S}_{\mathcal{A}}}{\sigma_{\mathcal{A}}^4}}\right), \tag{A107}$$

where $\mathcal{S}_k$ is the set of ordered sequences of σ_+^2 and σ_-^2, of length N_D, containing exactly k σ_+^2s. For example, for $N_D = 3$,

$$\mathcal{S}_0 = \{\sigma_-^2, \sigma_-^2, \sigma_-^2\}, \tag{A108}$$

$$\mathcal{S}_1 = \{\sigma_+^2, \sigma_-^2, \sigma_-^2\}, \{\sigma_-^2, \sigma_+^2, \sigma_-^2\}, \{\sigma_-^2, \sigma_-^2, \sigma_+^2\}, \tag{A109}$$

$$\mathcal{S}_2 = \{\sigma_+^2, \sigma_+^2, \sigma_-^2\}, \{\sigma_+^2, \sigma_-^2, \sigma_+^2\}, \{\sigma_-^2, \sigma_+^2, \sigma_+^2\}, \tag{A110}$$

$$\mathcal{S}_3 = \{\sigma_+^2, \sigma_+^2, \sigma_+^2\}. \tag{A111}$$

So the new mean is corrected by a term μ_D with respect to the statistics on the nonscrambled data

$$\mathring{\mu}_1 = \mu_1 - \mu_D, \tag{A112}$$

$$\mu_D = \int d\mathring{s}\, \hat{y}(u)\mathring{p}(u). \tag{A113}$$

And similarly for the variance, which gets a correction $\sigma_D^2 = \int du \hat{y}(u)^2 \mathring{p}(u) - \mu_D^2$,

$$\mathring{\sigma}_1^2 = \langle (\hat{y}(u(s)) - \hat{y}(u(\mathring{s})))^2 \rangle - \mathring{\mu}_1^2 \tag{A114}$$

$$= \langle \hat{y}(u(s))^2 \rangle + \langle \hat{y}(u(\mathring{s}))^2 \rangle - 2\langle \hat{y}(u(s))\hat{y}(u(\mathring{s})) \rangle - \mathring{\mu}_1^2 \tag{A115}$$

$$= \sigma_1^2 + \sigma_D^2. \tag{A116}$$

In conclusion, with respect to Eqs. (A91) and (A92), Eqs. (A112) and (A116) constitute a correction to the DSs.

c. Gaussian search of a non-Gaussian background

If we ignore the non-Gaussianity of the toy model, and we model only its Gaussian part, we have

$$\langle s_i^{\mathcal{A}} s_j^{\mathcal{B}} \rangle = \delta^{\mathcal{AB}} \delta_{ij} \sigma_{\mathcal{A}}^2 \qquad \text{under } \mathcal{H}_0, \tag{A117}$$

$$\langle s_i^{\mathcal{A}} s_j^{\mathcal{B}} \rangle = \delta_{ij}(\delta^{\mathcal{AB}} \sigma_{\mathcal{A}}^2 + \gamma_+ \sigma_+^2 + \gamma_- \sigma_-^2) = \delta_{ij}(\delta^{\mathcal{AB}} \sigma_{\mathcal{A}}^2 + \sigma_h^2) \text{ under } \mathcal{H}_1. \tag{A118}$$

The standard Gaussian detector is

$$\hat{Y}_G(s) = \sum_i \sum_{\mathcal{A}\neq\mathcal{B}} \frac{s_i^{\mathcal{A}} s_i^{\mathcal{B}}}{\sigma_{\mathcal{A}}^2 \sigma_{\mathcal{B}}^2}. \tag{A119}$$

Without loss of generality, we focus on a single data point and omit the i index. Under $\mathcal{H}_0$ we get mean and second order moment

$$\mu_{0,G} = \sum_{\mathcal{A}\neq\mathcal{B}} \frac{\langle s^{\mathcal{A}} s^{\mathcal{B}} \rangle_0}{\sigma_{\mathcal{A}}^2 \sigma_{\mathcal{B}}^2} = \sum_{\mathcal{A}\neq\mathcal{B}} \frac{\sigma_{\mathcal{A}}^2 \delta_{\mathcal{AB}}}{\sigma_{\mathcal{A}}^2 \sigma_{\mathcal{B}}^2} = 0 \tag{A120}$$

and

$$\begin{aligned}\sigma_{0,G}^2 + \mu_{0,G}^2 &= \sum_{\mathcal{A}\neq\mathcal{B}} \sum_{\mathcal{C}\neq\mathcal{D}} \frac{\langle s^{\mathcal{A}} s^{\mathcal{B}} s^{\mathcal{C}} s^{\mathcal{D}} \rangle_0}{\sigma_{\mathcal{A}}^2 \sigma_{\mathcal{B}}^2 \sigma_{\mathcal{C}}^2 \sigma_{\mathcal{D}}^2} \\ &= \sum_{\mathcal{A}\neq\mathcal{B}} \sum_{\mathcal{C}\neq\mathcal{D}} \frac{\delta_{\mathcal{AC}} \delta_{\mathcal{BD}} \sigma_{\mathcal{A}}^2 \sigma_{\mathcal{B}}^2}{\sigma_{\mathcal{A}}^2 \sigma_{\mathcal{B}}^2 \sigma_{\mathcal{C}}^2 \sigma_{\mathcal{D}}^2} + \sum_{\mathcal{A}\neq\mathcal{B}} \sum_{\mathcal{C}\neq\mathcal{D}} \frac{\delta_{\mathcal{AD}} \delta_{\mathcal{BC}} \sigma_{\mathcal{A}}^2 \sigma_{\mathcal{B}}^2}{\sigma_{\mathcal{A}}^2 \sigma_{\mathcal{B}}^2 \sigma_{\mathcal{C}}^2 \sigma_{\mathcal{D}}^2} \\ &= 2 \sum_{\mathcal{A}\neq\mathcal{B}} \frac{1}{\sigma_{\mathcal{A}}^2 \sigma_{\mathcal{B}}^2}.\end{aligned} \tag{A121}$$

Under $\mathcal{H}_1$ we get

$$\mu_{1,G} = \sum_{\mathcal{A}\neq\mathcal{B}} \frac{\langle h^{\mathcal{A}} h^{\mathcal{B}}\rangle_1}{\sigma^2_{\mathcal{A}}\sigma^2_{\mathcal{B}}} = \sigma_h^2 \sum_{\mathcal{A}\neq\mathcal{B}} \frac{1}{\sigma^2_{\mathcal{A}}\sigma^2_{\mathcal{B}}} \tag{A122}$$

and

$$\begin{aligned}\sigma^2_{1,G} + \mu^2_{1,G} &= \sum_{\mathcal{A}\neq\mathcal{B}}\sum_{\mathcal{C}\neq\mathcal{D}} \frac{\langle n^{\mathcal{A}} n^{\mathcal{B}} n^{\mathcal{C}} n^{\mathcal{D}}\rangle_1}{\sigma^2_{\mathcal{A}}\sigma^2_{\mathcal{B}}\sigma^2_{\mathcal{C}}\sigma^2_{\mathcal{D}}} + \sum_{\mathcal{A}\neq\mathcal{B}}\sum_{\mathcal{C}\neq\mathcal{D}} \frac{\langle h^{\mathcal{A}} h^{\mathcal{B}} h^{\mathcal{C}} h^{\mathcal{D}}\rangle_1}{\sigma^2_{\mathcal{A}}\sigma^2_{\mathcal{B}}\sigma^2_{\mathcal{C}}\sigma^2_{\mathcal{D}}} \\ &\quad + \sum_{\mathcal{A}\neq\mathcal{B}}\sum_{\mathcal{C}\neq\mathcal{D}} \frac{\langle n^{\mathcal{A}} h^{\mathcal{B}} n^{\mathcal{C}} h^{\mathcal{D}}\rangle_1}{\sigma^2_{\mathcal{A}}\sigma^2_{\mathcal{B}}\sigma^2_{\mathcal{C}}\sigma^2_{\mathcal{D}}} + \frac{\langle n^{\mathcal{A}} h^{\mathcal{B}} h^{\mathcal{C}} n^{\mathcal{D}}\rangle_1}{\sigma^2_{\mathcal{A}}\sigma^2_{\mathcal{B}}\sigma^2_{\mathcal{C}}\sigma^2_{\mathcal{D}}} + \frac{\langle h^{\mathcal{A}} n^{\mathcal{B}} n^{\mathcal{C}} h^{\mathcal{D}}\rangle_1}{\sigma^2_{\mathcal{A}}\sigma^2_{\mathcal{B}}\sigma^2_{\mathcal{C}}\sigma^2_{\mathcal{D}}} + \frac{\langle h^{\mathcal{A}} n^{\mathcal{B}} h^{\mathcal{C}} n^{\mathcal{D}}\rangle_1}{\sigma^2_{\mathcal{A}}\sigma^2_{\mathcal{B}}\sigma^2_{\mathcal{C}}\sigma^2_{\mathcal{D}}} \qquad \text{(A123)} \\ &= 2\sum_{\mathcal{A}\neq\mathcal{B}} \frac{1}{\sigma^2_{\mathcal{A}}\sigma^2_{\mathcal{B}}} + 3(\gamma_+\sigma_+^4 + \gamma_-\sigma_-^4)\left(\sum_{\mathcal{A}\neq\mathcal{B}} \frac{1}{\sigma^2_{\mathcal{A}}\sigma^2_{\mathcal{B}}}\right)^4 + 4\sigma_h^2 \sum_{\mathcal{A}\neq\mathcal{B}}\sum_{\mathcal{A}\neq\mathcal{C}} \frac{1}{\sigma^2_{\mathcal{A}}\sigma^2_{\mathcal{B}}\sigma^2_{\mathcal{C}}}. \qquad \text{(A124)}\end{aligned}$$

5. Bayesian analysis for the toy model

The Bayesian inference can be constructed by parametrizing the signal hypothesis $\mathcal{H}_1$ with the model parameters $\mathcal{M}$. Therefore the posterior reads

$$p(\mathcal{M}|s) = \mathcal{L}(s|\mathcal{M})\pi(\mathcal{M}) \propto \mathcal{N}_{n+g} \int_h e^{-\frac{1}{2}\mathcal{W}_{n+g}(s-h,s-h)}\, p_h[h|\mathcal{M}]\pi(\mathcal{M}). \tag{A125}$$

For the toy model in Sec. III, $\mathcal{M}$ is specified by $(\sigma_h, \sigma_+, \sigma_-)$ and, assuming stationary, uncorrelated noises across detectors and $g = 0$, the following simplifications occur:

$$\mathcal{N}_{n+g} = \mathcal{N}_n = \left(\prod_{\mathcal{A}=1}^{N_D} \frac{1}{\sqrt{2\pi\sigma^2_{\mathcal{A}}}}\right)^{N_s}, \tag{A126}$$

$$\mathcal{W}_{n+g} = \mathcal{W}_n(s-h, s-h) = \sum_i \sum_{\mathcal{A}} \frac{(s_i^{\mathcal{A}} - h_i^{\mathcal{A}})^2}{\sigma^2_{\mathcal{A}}}. \tag{A127}$$

Therefore the posterior reads

$$p(\mathcal{M}|s) \propto \left(\prod_{\mathcal{A}=1}^{N_D} \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{A}}}\right)^{N_s} \int \prod_i \mathrm{d}h_i \times \exp\left[-\frac{1}{2}\sum_i \sum_{\mathcal{A}} \frac{(s_i^{\mathcal{A}} - h_i)^2}{\sigma^2_{\mathcal{A}}}\right] p(h_i|\mathcal{M})\pi(\mathcal{M}), \tag{A128}$$

which, as expected, decomposes into the product of likelihoods for each data point

$$p(\mathcal{M}|s) = \pi(\mathcal{M}) \prod_i \mathcal{L}(s_i|\mathcal{M}), \tag{A129}$$

$$\mathcal{L}(s_i|\mathcal{M}) = \left(\prod_{\mathcal{A}=1}^{N_D} \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{A}}}\right) \times \int \mathrm{d}h \exp\left[-\frac{1}{2}\sum_{\mathcal{A}} \frac{(s_i^{\mathcal{A}} - h)^2}{\sigma^2_{\mathcal{A}}}\right] p(h|\mathcal{M}) \tag{A130}$$

$$= \sum_{\alpha=+,-} \frac{\gamma_\alpha}{\sqrt{1 + \frac{\sigma_\alpha^2}{\sigma^2}}} \times \exp\left\{-\frac{1}{2}\mathcal{Q}^\alpha_{\mathcal{A}\mathcal{B}} s_i^{\mathcal{A}} s_i^{\mathcal{B}} - \sum_{\mathcal{A}} \log\sqrt{2\pi}\sigma_{\mathcal{A}}\right\}, \tag{A131}$$

where the single data points collected across detectors are weighted by the quadratic form

$$\mathcal{Q}^{\alpha}_{\mathcal{AB}} = \frac{1}{\sigma^2_{\mathcal{A}}\sigma^2_{\mathcal{B}}}\left(\delta_{\mathcal{AB}} - \frac{\sigma^2_\alpha}{\sigma^2+\sigma^2_\alpha}\frac{\sigma^{-1}_{\mathcal{A}}\sigma^{-1}_{\mathcal{B}}}{\sigma^{-2}}\right). \tag{A132}$$

6. Cumulants for the toy model

We provide here an explicit calculation for the cumulant generating function of the model presented in Sec. III. This serves the reader with a mapping of previous approaches in literature into our formalism [41]. It is straightforward to compute the cumulant generating function of a single h distributed according to Eq. (51):

$$K(z) = \log\left[\gamma_- \exp\left(\frac{z^2\sigma_-^2}{2}\right) + \gamma_+ \exp\left(\frac{z^2\sigma_+^2}{2}\right)\right]. \tag{A133}$$

With each h_i independent (upon scrambling) and equally distributed across detectors, the cumulant generating function of a set of N data points is simply a sum of K's with independent auxiliary variables.

$$\mathcal{K}(z^{\mathcal{A}_1}_{i_1}, \ldots, z^{\mathcal{A}_n}_{i_n}) = \sum_{j=1}^{n} K(z^{\mathcal{A}_j}_{i_j}). \tag{A134}$$

Previous studies approximate $p(h)$ with suitable asymptotic expansions (e.g., *Gram-Charlier A* or *Edgeworth* expansions), and then make use of a subset of cumulants. Those approaches may be reproduced by using

$$\Gamma^{\mathcal{A}_1\cdots\mathcal{A}_n}_{i_1\cdots i_n} = \partial^{\mathcal{A}_1}_{i_1}\cdots\partial^{\mathcal{A}_n}_{i_n}\mathcal{K}(z^{\mathcal{A}_1}_{i_1}, \ldots, z^{\mathcal{A}_n}_{i_n})\big|_{z^{\mathcal{A}_j}_{i_j}=0} \tag{A135}$$

$$= \mathbf{1}^{\mathcal{A}_1\cdots\mathcal{A}_n}\delta_{i_1\cdots i_n} n\left[\frac{\partial^n}{\partial s^n}K(s)\right]_{s=0} \tag{A136}$$

$$\equiv \mathbf{1}^{\mathcal{A}_1\cdots\mathcal{A}_n}\delta_{i_1\cdots i_n} n\Gamma_n. \tag{A137}$$

Differentiating $k(s)$ yields the nth cumulant Γ_n of the single h. The value is, as a function of the parameter models $(\sigma_+, \sigma_-, \sigma_h)$ (assuming without loss of generality $\gamma_+ > \gamma_-$) [68],

$$\Gamma_{2r} = \delta_{1r}\sigma_+^2 - \sum_{q=1}^{2r}\sum_{k=1}^{\infty} k^{q-1}\left(-\frac{\gamma_-}{\gamma_+}\right)^k B_{2r,q}(\sigma_+^2, \sigma_-^2), \tag{A138}$$

$$\Gamma_{2r+1} = 0, \tag{A139}$$

with $B_{k,q}$ the partial exponential Bell polynomials [69].

[1] J. Aasi, B. P. Abbott, R. Abbott *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **32**, 074001 (2015).

[2] F. Acernese, M. Agathos, K. Agatsuma, D. Aisa *et al.*, Classical Quantum Gravity **32**, 024001 (2015).

[3] T. Akutsu, M. Ando, K. Arai *et al.* (Kagra Collaboration), Nat. Astron. **3**, 35 (2019).

[4] R. Abbott *et al.* (The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration), arXiv:2111.03606.

[5] R. Abbott *et al.* (The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration), arXiv:2111.03634 [Phys. Rev. X (to be published)].

[6] B. P. Abbott *et al.* (KAGRA Collaboration, LIGO Scientific Collaboration, and Virgo Collaboration), Living Rev. Relativity **21**, 3 (2018).

[7] M. Punturo, M. Abernathy, F. Acernese *et al.*, Classical Quantum Gravity **27**, 194002 (2010).

[8] P. Amaro-Seoane, H. Audley, S. Babak, J. Baker *et al.*, arXiv:1702.00786.

[9] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese *et al.*, Phys. Rev. D **104**, 022004 (2021).
[10] A. Buonanno, G. Sigl, G. G. Raffelt, H.-T. Janka, and E. Müller, Phys. Rev. D **72**, 084001 (2005).
[11] E. Howell, D. Coward, R. Burman, D. Blair, and J. Gilmore, Mon. Not. R. Astron. Soc. **351**, 1237 (2004).
[12] P. Sandick, K. A. Olive, F. Daigne, and E. Vangioni, Phys. Rev. D **73**, 104024 (2006).
[13] V. Ferrari, S. Matarrese, and R. Schneider, Mon. Not. R. Astron. Soc. **303**, 258 (1999).
[14] T. Regimbau and J. A. de Freitas Pacheco, Astron. Astrophys. **376**, 381 (2001).
[15] P. D. Lasky, M. F. Bennett, and A. Melatos, Phys. Rev. D **87**, 063004 (2013).
[16] X.-J. Zhu, E. J. Howell, D. G. Blair, and Z.-H. Zhu, Mon. Not. R. Astron. Soc. **431**, 882 (2013).
[17] S. Marassi, R. Schneider, G. Corvino, V. Ferrari, and S. Portegies Zwart, Phys. Rev. D **84**, 124037 (2011).
[18] C. Wu, V. Mandic, and T. Regimbau, Phys. Rev. D **85**, 104024 (2012).
[19] P. A. Rosado, Phys. Rev. D **84**, 084004 (2011).
[20] A. J. Farmer and E. S. Phinney, Mon. Not. R. Astron. Soc. **346**, 1197 (2003).
[21] R. R. Caldwell and B. Allen, Phys. Rev. D **45**, 3447 (1992).
[22] T. Damour and A. Vilenkin, Phys. Rev. D **71**, 063510 (2005).
[23] S. Ölmez, V. Mandic, and X. Siemens, Phys. Rev. D **81**, 104028 (2010).
[24] L. P. Grishchuk, Phys. Rev. D **48**, 3513 (1993).
[25] R. Easther, J. Giblin, John T., and E. A. Lim, Phys. Rev. Lett. **99**, 221301 (2007).
[26] J. L. Cook and L. Sorbo, Phys. Rev. D **85**, 023534 (2012).
[27] S. Clesse, J. García-Bellido, and S. Orani, arXiv:1812.11011.
[28] T. Callister, M. Fishbach, D. E. Holz, and W. M. Farr, Astrophys. J. Lett. **896**, L32 (2020).
[29] S. S. Bavera, G. Franciolini, G. Cusin, A. Riotto, M. Zevin, and T. Fragos, Astron. Astrophys. **660**, A26 (2022).
[30] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese *et al.*, Astrophys. J. **923**, 14 (2021).
[31] R. Buscicchio, C. J. Moore, G. Pratten, P. Schmidt, M. Bianconi, and A. Vecchio, Phys. Rev. Lett. **125**, 141102 (2020).
[32] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. Lett. **118**, 121101 (2017); **119**, 029901(E) (2017).
[33] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. D **100**, 061101 (2019).
[34] X.-J. Zhu, E. Howell, T. Regimbau, D. Blair, and Z.-H. Zhu, Astrophys. J. **739**, 86 (2011).
[35] S. Mukherjee and J. Silk, Mon. Not. R. Astron. Soc. **491**, 4690 (2020).
[36] S. Drasco and E. E. Flanagan, Phys. Rev. D **67**, 082003 (2003).
[37] E. Thrane, Phys. Rev. D **87**, 043009 (2013).
[38] R. Smith and E. Thrane, Phys. Rev. X **8**, 021019 (2018).
[39] L. Martellini and T. Regimbau, Phys. Rev. D **92**, 104025 (2015); **97**, 049903(E) (2018).
[40] T. Regimbau, S. Giampanis, X. Siemens, and V. Mandic, Phys. Rev. D **85**, 066001 (2012).
[41] L. Martellini and T. Regimbau, Phys. Rev. D **89**, 124009 (2014).
[42] R. Flauger, N. Karnesis, G. Nardini, M. Pieroni A. Ricciardone, and J. Torrado, J. Cosmol. Astropart. Phys. 01 (2021) 059.
[43] M. Georgousi, N. Karnesis, V. Korol, M. Pieroni, and N. Stergioulas, arXiv:2204.07349.
[44] N. Karnesis, S. Babak, M. Pieroni, N. Cornish, and T. Littenberg, Phys. Rev. D **104**, 043019 (2021).
[45] E. Racine and C. Cutler, Phys. Rev. D **76**, 124033 (2007).
[46] A. I. Renzini, B. Goncharov, A. C. Jenkins, and P. M. Meyers, Galaxies **10**, 34 (2022).
[47] P. Jaranowski, *Analysis of Gravitational-Wave Data* (Cambridge University Press, Cambridge, New York, 2009).
[48] J. Neyman and E. S. Pearson, Phil. Trans. R. Soc. A **231**, 289 (1933).
[49] S. Kay, *Fundamentals of Statistical Signal Processing: Detection theory*, Fundamentals of Statistical Signal Processing (PTR Prentice-Hall, Englewood Cliffs, NJ, 1993).
[50] L. Isserlis, Biometrika **12**, 134 (1918).
[51] M. Was, M.-A. Bizouard, V. Brisson, F. Cavalier, M. Davier, P. Hello, N. Leroy, F. Robinet, and M. Vavoulidis, Classical Quantum Gravity **27**, 015005 (2010).
[52] P. J. Green, Biometrika **82**, 711 (1995).
[53] L. Martellini and T. Regimbau, Phys. Rev. D **89**, 124009 (2014).
[54] T. Regimbau, Res. Astron. Astrophys. **11**, 369 (2011).
[55] D. Cox and V. Isham, Point Processes (Routledge, London, UK, 1980).
[56] J. Skilling, Bayesian Anal. **1**, 833 (2006).
[57] J. Veitch, W. Del Pozzo, A. Lyttle, M. Williams *et al.*, johnveitch/cpnest: v0.11.3, Zenodo, 10.5281/zenodo.4470001 (2021).
[58] R. Buscicchio, E. Roebber, J. M. Goldstein, and C. J. Moore, Phys. Rev. D **100**, 084041 (2019).
[59] D. Meacher, E. Thrane, and T. Regimbau, Phys. Rev. D **89**, 084063 (2014).
[60] R. Buscicchio, An improved detector for non Gaussian stochastic background of gravitational waves., Master's thesis, Università di Pisa, 2016.
[61] N. V. Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, New York, 1992).
[62] G. Ashton, S. Thiele, Y. Lecoeuche, J. McIver, and L. K. Nuttall, Classical Quantum Gravity **39**, 175004 (2022).
[63] W. R. Inc., *Mathematica*, Version 13.1, Champaign, IL, 2022, https://www.wolfram.com/mathematica.
[64] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).
[65] J. D. Hunter, Comput. Sci. Eng. **9**, 90 (2007).
[66] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers *et al.*, Nature (London) **585**, 357 (2020).
[67] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland *et al.*, Nat. Methods **17**, 261 (2020).
[68] C. S. Withers, S. Nadarajah, and S. H. Shih, Methodol. Comput. Appl. Probab. **17**, 541 (2015).
[69] L. Comtet, *Advanced Combinatorics: The Art of Finite and Infinite Expansions; Rev. Version* (Reidel, Dordrecht, 1974); *Translation of: Analyse Combinatoire* (Presses University de France, Paris, 1970).