

Improved detection statistics for non-Gaussian gravitational wave stochastic backgrounds

Matteo Ballelli^{1,2}, Riccardo Buscicchio^{3,4,5,*}, Barbara Patricelli^{1,2}, Anirban Ain^{1,2} and Giancarlo Cella¹

¹*INFN Sezione di Pisa, Largo B. Pontecorvo 3, I-56127 Pisa, Italy*

²*Università di Pisa, Dipartimento di Fisica “E. Fermi”, Largo B. Pontecorvo 3, I-56127 Pisa, Italy*

³*Dipartimento di Fisica “G. Occhialini”, Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy*

⁴*INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy*

⁵*Institute for Gravitational Wave Astronomy and School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, United Kingdom*

 (Received 21 December 2022; revised 23 April 2023; accepted 17 May 2023; published 23 June 2023)

In a recent paper we described a novel approach to the detection and parameter estimation of a non-Gaussian stochastic background of gravitational waves. In this work we propose an improved version of the detection procedure, preserving robustness against imperfect noise knowledge at no cost of detection performance; in the previous approach, the solution proposed to ensure robustness reduced the performances of the detection statistics, which in some cases (namely, mild non-Gaussianity) could be outperformed by Gaussian ones established in literature. We show, through a simple toy model, that the new detection statistic performs better than the previous one (and than the Gaussian statistic) everywhere in the parameter space. It approaches the optimal Neyman-Pearson statistics monotonically with increasing non-Gaussianity and/or number of detectors. In this study we discuss in detail its efficiency. This is a second, important step towards the implementation of a nearly optimal detection procedure for a realistic non-Gaussian stochastic background. We discuss the relevance of results obtained in the context of the toy model used, and their importance for understanding a more realistic scenario.

DOI: [10.1103/PhysRevD.107.124044](https://doi.org/10.1103/PhysRevD.107.124044)

I. INTRODUCTION

The non-Gaussian stochastic gravitational wave (GW) background is an interesting and promising target for GW detectors. It originates from a superposition of many uncorrelated and unresolved events. An interesting example is the background originating from compact binary coalescences, which is within reach of the current generation detector network [1], but several others can be observed in the future, both of astrophysical [2] and cosmological origin [3].

The statistical properties of a non-Gaussian background are related to astrophysically relevant quantities of great interest, such as the event rate and the population distribution of source parameters. In the presence of strong overlap between individual signals generated by the different events, the stochastic background is well modeled by a

Gaussian stochastic process, fully characterized by its second order statistic. In absence of it, the statistical distribution of the strain signals measured by the detectors becomes non-Gaussian and contains larger information that can be in principle extracted. For example, accordingly with the rate estimates based on recent LVK observations [4] the background generated by binary black hole coalescences is expected to be strongly non-Gaussian, with a product between the length of the observable coalescence signal and the event rate $O(10^{-3})$. The same parameter is $O(10)$ for binary neutron star coalescences, and non-Gaussianity is expected to be less important.

Several detection strategies to efficiently detect a non-Gaussian background have been proposed in the literature. Drasco and Flanagan [5] devised a likelihood appropriate for the superposition of burstlike signals; therein optimal detection statistics was derived assuming well-separated, burstlike signals, coaligned and colocated detectors, and white noise. An approach applicable without these assumptions, but only with a negligible superposition probability for different events has been proposed by Thrane [6]; Smith and Thrane [7] further proposed an optimal Bayesian parameter estimation method suitable for a background of astrophysical events in the low overlap regime.

*riccardo.buscicchio@unimib.it

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

In all the above proposals the basic building block is the probability distribution p for a convenient statistics of the data ρ that can be evaluated for each segment of a measured signal. This is written as a weighted sum

$$p(\rho) = \xi p_1(\rho) + (1 - \xi) p_0(\rho), \quad (1)$$

where $\rho \sim p_1$ when the given segment contains an astrophysical event and $\rho \sim p_0$ otherwise. The weight ξ is just the probability of having an event in the segment.

Other approaches, based on semiparametric models [8] or higher-order statistics [9,10] have also been explored. A comprehensive review can be found in [11].

In a recent paper [12] we discussed a direct approach for detection and parameter estimation of a non-Gaussian stochastic background of GWs through a network of detectors. The core idea was to try to follow, as much as possible, an optimal approach. The target was to implement an optimal detection statistics (DS), and similarly a Bayesian approach to parameter estimation based on a realistic likelihood \mathcal{L} . To circumvent the lack of analytical closed-form expressions for the DS and for \mathcal{L} we opted to numerically evaluate them. This approach implies approximations of relevant integrals, that can be reduced at will with a large enough computational power. We proposed a Monte Carlo importance sampling procedure to achieve this objective. In this paper we focus on a further improvement of the DS, while not investigating parameter estimation anymore. In Sec. II we summarize the our previous relevant findings, leading to the improvement presented in this paper.

II. SUMMARY OF PREVIOUS RESULTS

As a first step in [12] we studied the optimal Neyman-Pearson DS for a simple toy model, where the stochastic signal is a sequence of independent pulses with amplitudes distributed as a Gaussian mixture. In this particular case the DS can be evaluated in a closed form as a nonlinear function of the data.

A conceptual issue arose in this context; the optimal Neyman-Pearson DS, which is obtained under the assumption of known noise, contains contributions proportional to single-detector data autocorrelations (of arbitrary order). These contributions are dominated by the noise,¹ and in a realistic regime the noise amplitude cannot be assumed known with a precision high enough to not spoil the DS effectiveness.

The same shortcoming has been encountered in literature also in the Gaussian case, and we will clarify it with an example. Let us suppose we want to discriminate between the models

$$\mathcal{H}_0: s_i^A = \sigma n_i^A, \quad (2)$$

¹We always assume a Gaussian noise.

$$\mathcal{H}_1: s_i^A = \sigma n_i^A + A h_i, \quad (3)$$

where s_i^A are measured values ($1 < A < N_D$ is an index labeling the detector and $i \leq 1 \leq N$), σ and A are the noise and signal amplitudes. The stochastic variables n_i, h are normally distributed and independent. Under the assumption of a known σ^2 the Neyman-Pearson test is easily found from

$$\frac{p(s_i^A | \sigma^2, A^2)}{p(s_i^A | \sigma^2, 0)} > \lambda \quad (4)$$

which is equivalent to

$$Y = \frac{1}{N} \sum_{i=1}^N \left[2 \sum_{A>B} s_i^A s_i^B + \sum_A (s_i^A)^2 \right] > \lambda. \quad (5)$$

The mean of Y is

$$\mu_Y = [N_D(N_D - 1)A^2 + N_D(A^2 + \sigma^2)]. \quad (6)$$

We see that the first term, originating from cross-correlations, is noise independent while the second, originating from autocorrelations, depends both upon signal and noise. As a matter of fact, for a realistic stochastic background σ^2 is uncertain with an uncertainty larger than A^2 . This means that we will not be able to detect confidently a stochastic background from the DS in (5); a given value of Y could be explained both by a background contribution or by a noise fluctuation. The problem disappears when the noise-dominated diagonal components are removed, obtaining

$$Y = \frac{1}{N} \sum_{i=1}^N \left[2 \sum_{A>B} s_i^A s_i^B \right] > \lambda. \quad (7)$$

This is a very simplified version of the statistic commonly employed to detect a Gaussian stochastic background, see for example Eq. (3.73) in [13] which defines the optimal cross-correlation in the Gaussian case

$$S \propto \int_{-\infty}^{\infty} \frac{\gamma(f) \Omega_{GW}(f)}{f^3 S_n^A(f) S_n^B(f)} s^A(f)^* s^B(f) df. \quad (8)$$

A simplification comes from the assumption of co-located and co-oriented detectors which makes the overlap reduction function $\gamma(f)$ equal to one. Furthermore, both the stochastic signal and the noise are assumed here to have a white spectrum, so $\Omega_{GW} \propto f^3$, $S_n^{A,B}$ are frequency independent and Eq. (8) becomes equivalent to Eq. (5). It is interesting to note that if we use a generalized likelihood ratio test like

$$\frac{p(s_i^A | \hat{\sigma}_1^2, \hat{A}_1^2)}{p(s_i^A | \hat{\sigma}_0^2, 0)} > \lambda, \quad (9)$$

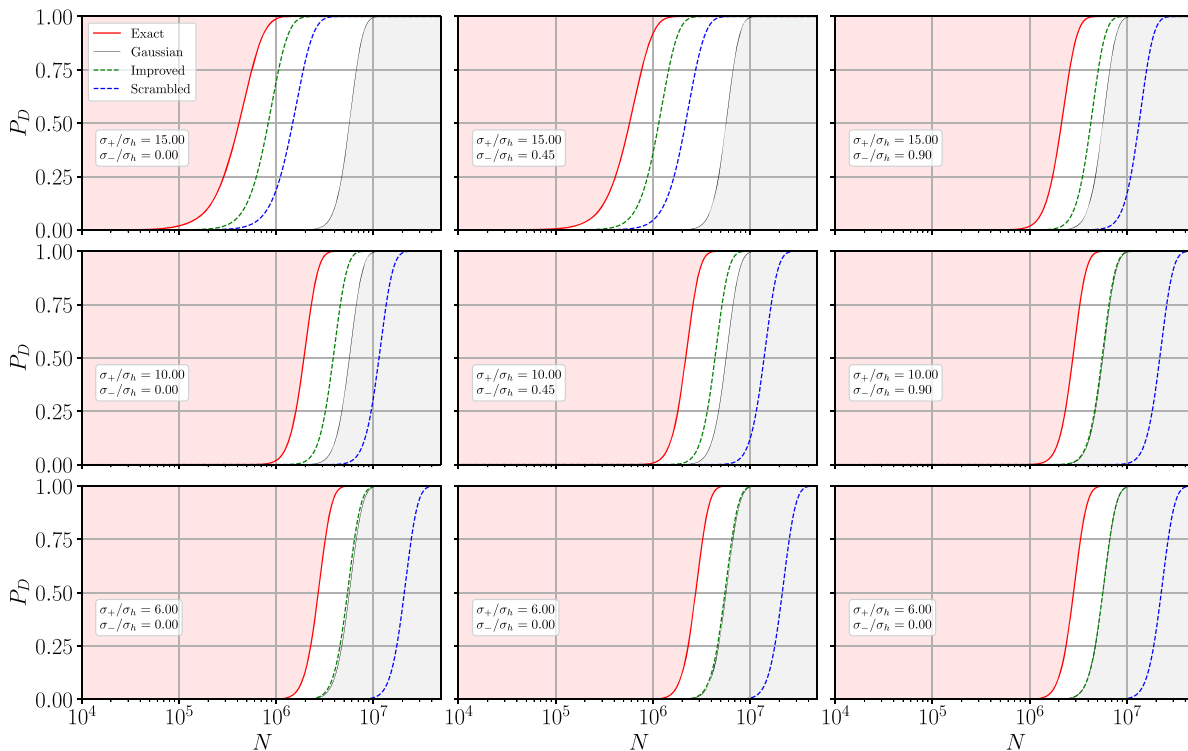


FIG. 1. Comparison between detections statistics; each panel contains the probability of detection P_D as a function of the number of data points N . We show the “Exact,” “Improved,” “Scrambled,” and “Gaussian” statistic as solid-red, dashed-green, dashed-blue, and solid-black lines, respectively. Several different signal models are considered, parametrized by σ_+/σ_h and σ_-/σ_h , of increasing non-Gaussianity, going from bottom-right to top-left panel. We fix $N_D = 2$ and $P_{FA} = 10^{-15}$. Shaded red areas delimit regions with beyond-optimal performances, right-bounded by the “Exact” DS. Shaded gray area delimit statistics with performances worse than the “Gaussian” one. We observe uniform improvement over the whole parameter space of the “Improved” statistics as compared to the Gaussian one. By contrast, the scrambling procedure introduced in [12] outperforms the Gaussian one for strong non-Gaussianities, only. In Figs. 2, 4, and 5 similar results for increasing number of detectors are shown, with the “Improved” statistics approaching the “Exact” one.

where $\hat{\sigma}_i$ and \hat{A}_i are the maximum-likelihood estimates of σ and A under the \mathcal{H}_1 hypothesis we obtain, when $\sigma^2 \gg A^2$,

$$Y = \sqrt{\frac{\hat{A}_1^2}{\hat{\sigma}_0^2}} = \frac{1}{N} \sum_i \sum_{\mathcal{A} > \mathcal{B}} s_i^{\mathcal{A}} s_i^{\mathcal{B}} > \lambda. \quad (10)$$

For large enough N , relative fluctuations in the denominator become negligible, and we recover the DS in Eq. (7).

Under the assumption that noises across different detectors are uncorrelated, the modified statistic has by construction a zero expectation value in the hypothesis \mathcal{H}_0 (i.e., absence of signal), and a detection procedure robust against noise mismodeling can be defined (see for example [13], or [14] and references therein).

Removing analogous contributions² from the nonlinear statistic, appropriate for a non-Gaussian background, is

²Because of the analogy with the Gaussian case we will henceforth refer to these terms as diagonal. The generalized definition of a diagonal term is the following; a correlation between the data, of arbitrary order, that has a nonzero average in the \mathcal{H}_0 hypothesis of no background.

tricky. In [12] we defined a procedure involving the subtraction of a set of auxiliary data constructed from the original ones. The basic idea was to construct a set of data streams (one for each detector) with the same autocorrelation but with zero cross-correlation among themselves and with the original set. This can easily be done, for example by introducing a time shift larger than the correlation length of the stochastic background. In this way autocorrelations can be independently estimated and subtracted.

However our approach was suboptimal; auxiliary data have by construction zero average estimates for cross-correlation, but these estimates—which are subtracted from the ones of the original data—carry additional fluctuations. These are summed to the ones of the original set; as a consequence, we obtain more noisy DS and the detection performance is reduced. In [12] this was evident in the regime where non-Gaussianity was not too high. It was quantified by the reduction of detection probability for a given false alarm, which in some cases was worse than the one of the Gaussian DS. We show these results again in this paper, comparing them with the improved ones (see Figs. 1, 2, 4, and 5).

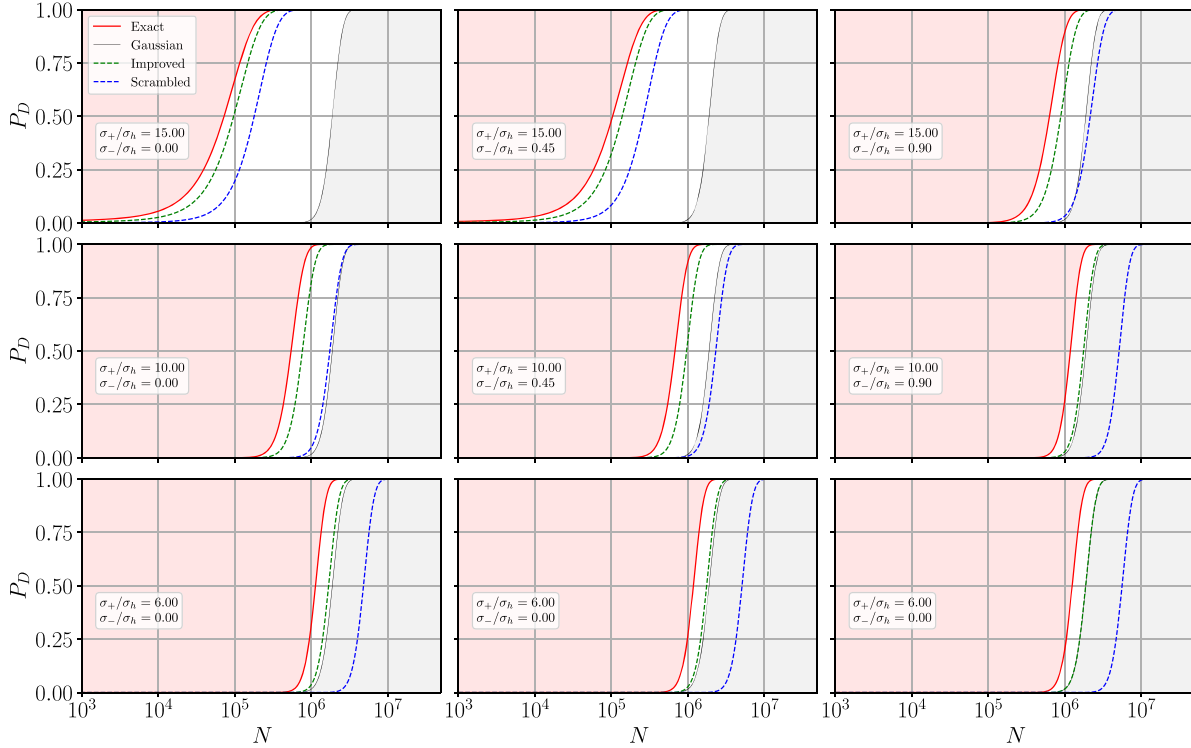


FIG. 2. Comparison between detections statistics: each panel contains the probability of detection P_D as a function of the number of datapoints N . We show the “Exact,” “Improved,” “Scrambled,” and “Gaussian” statistic as solid red, dashed green, dashed blue, and solid black lines, respectively. Several different signal models are considered, parameterized by σ_+/σ_h and σ_-/σ_h , of increasing non-Gaussianity, going from bottom right to top left panel. We fix $N_D = 3$ and $P_{FA} = 10^{-15}$. Shaded red areas delimit regions with beyond-optimal performances, right-bounded by the “Exact” detections statistic. Shaded gray areas delimit statistics with performances worse than the “Gaussian” one. We observe uniform improvement over the whole parameter space of the “Improved” statistics as compared to the “Gaussian” one. By contrast, the scrambling procedure introduced in [12] outperforms the Gaussian one for strong non-Gaussianities, only.

In this work we define a different, straightforward procedure to remove diagonal contributions from the nonlinear statistics. This procedure does not introduce additional fluctuations, and has improved performances compared to [12], when measured in term of figures of merit related to detection probability, as will be discussed extensively in Sec. IV. We apply the new procedure to the same toy model used previously, and we quantify the resulting improvement. Finally, we show that the new procedure can be applied to a realistic background.

Henceforth, we will discuss and compare a number of detections statistics, with the following naming convention:

- (i) The “Exact” DS is constructed via the Neyman-Pearson lemma [15]. It is formally the optimal one, but the presence of diagonal terms makes it unusable in a realistic setup, i.e., when uncertainty on the noise amplitude is larger than the target GW signal under the hypothesis \mathcal{H}_1 .
- (ii) The “Scrambled” DS is obtained from the “Exact” statistic with the subtraction procedure defined and characterized in [12], and shortly described in this section.

- (iii) The “Improved” DS is obtained from the “Exact” statistic with the new subtraction procedure described in Sec. III.
- (iv) The “Gaussian” DS is the optimal Neyman-Pearson detector for a Gaussian stochastic background with a spectrum matching that of the toy model described in Sec. III, the same considered in [12]. It is suboptimal when applied to a non-Gaussian stochastic background, and is used as a fiducial reference for performance comparison.

In this work we provide only minimal details on the “Exact,” “Scrambled,” and “Gaussian” DSs construction. The interested reader will find relevant ones in [12] and references therein. We will instead focus on the “Improved” statistics. In Sec. III we construct it explicitly, providing a proof of its optimality with respect to a set of robustness requirements. In Sec. IV A we present performance comparisons between the above statistics, exploring their behavior for varying signal duration, their amount of non-Gaussianity, and the number of detectors. In Sec. IV B, we briefly discuss a realistic implementation of the procedure, highlighting its advantages. Finally, in Sec. V we

draw conclusions and prospects for further developments of our approach.

III. IMPROVED DETECTION

Our discussion will be mainly in the context of the toy model studied in [12]. It allows for a quantitative comparison of detector performances. In this model, the data measured by each detector are time series of the form

$$s_i^A = n_i^A + h_i, \quad (11)$$

where i is a discrete time index, $1 \leq i \leq N$, and $t_i = i\delta t$ being the measurement time and δt the sampling time. We suppose to have a network of N_D detectors labeled by the index \mathcal{A} , with $1 \leq \mathcal{A} \leq N_D$. The noise component of the data is n_i^A , and we will assume that n_i^A s are Gaussian, zero-average random numbers with covariances defined by

$$\langle n_i^A n_j^B \rangle = \sigma_{\mathcal{A}}^2 \delta_{ij} \delta^{AB}. \quad (12)$$

The signal h_i is the same on each detector, so they can be thought as identical, aligned, and colocated. It follows a Gaussian mixture model distribution

$$h_i \sim \sum_{\alpha} p_{\alpha} \mathcal{N}(\cdot | 0, \sigma_{\alpha}), \quad (13)$$

$$\sigma_h^2 = \langle h_i^2 \rangle = \sum_{\alpha} p_{\alpha} \sigma_{\alpha}^2. \quad (14)$$

By hypothesis, different signal data points are therefore independent; the waveform associated to our events are very short, deltalike burst, without a resolved structure. Let us write the single event in the form

$$h_i(h, j) = h \delta_{ij}. \quad (15)$$

We have two parameters; the amplitude h and the index j associated to the event time. As the model is time independent, all the values of j have the same probability. If $p(h)$ is the probability distribution for the amplitude parameter, h_i will be distributed as

$$h_i \sim e^{-\Gamma \delta t} \left(\delta(\cdot) + \sum_{n=1}^{\infty} \frac{(\Gamma \delta t)^n}{n!} \underbrace{(p \star p \star \dots \star p)}_{n \text{ terms}}(\cdot) \right), \quad (16)$$

where Γ is the event rate. This can be approximated by the toy model with two mixture components $\alpha \in \{+, -\}$ introduced in [12] if $p \sim \mathcal{N}(\cdot | 0, \sigma_{\pm})$ and $\Gamma \delta t \ll 1$, in such a way that terms with $n > 1$ can be neglected in Eq. (16)

$$h_i \sim (1 - \Gamma \delta t) \delta(\cdot) + \Gamma \delta t \mathcal{N}(\cdot | 0, \sigma_{\pm}). \quad (17)$$

TABLE I. Relation between physical quantities and model parameters.

Physical quantity	Definition	Model parameters
Γ	Astrophysical background events rate	$\frac{1}{\delta t} \frac{\sigma_h^2 - \sigma_-^2}{\sigma_+^2 - \sigma_-^2} = \frac{p_+}{\delta t}$
$S_{h, \text{Astro}}$	Astrophysical background power spectrum	$(\sigma_+^2 - \sigma_-^2) \delta t$
$S_{h, \text{Gauss}}$	Gaussian background power spectrum	$\sigma_-^2 \delta t$
T	Observing time	$N \delta t$

This corresponds to $\sigma_- = 0$, $p_+ = \Gamma \delta t$ and $\sigma_+ / \sigma_h = 1 / \sqrt{\Gamma \delta t}$. The parameter $\Gamma \delta t$ is the average number of events that contribute to a given h_i , a measure of overlap and Gaussianity. For our deltalike signals the approximation $\Gamma \delta t \ll 1$ is not a problem, because the sampling time δt is a parameter unrelated to the physics and could be assumed to be small. The two main limits of the model are the lack of any time structure for the event waveform, which does not allow for relevant overlap structure, and the approximation of identical (coaligned and colocated) detectors. To accommodate for the general case our model should be interpreted as a superposition of an astrophysical contribution like the previous one, with $p \sim \mathcal{N}(\cdot | 0, \sigma_+)$, and of a Gaussian contribution with variance σ_-^2 . In this case (again in the approximation $\Gamma \delta t \ll 1$) we obtain

$$h_i \sim (1 - \Gamma \delta t) \mathcal{N}(\cdot | 0, \sigma_-) + \Gamma \delta t \mathcal{N}(\cdot | 0, \sqrt{\sigma^2 + \sigma_-^2}) \quad (18)$$

with $\sigma_h = \sqrt{\sigma_-^2 + \Gamma \delta t \sigma^2}$ and $\sigma_+ = \sqrt{\sigma^2 + \sigma_-^2}$. We summarize the relation between the model parameters and physical quantities in Table I. The Neyman-Pearson optimal DS for the two-component mixture model, \hat{Y} , reads

$$\hat{Y} = \sum_i \hat{y}(w_i), \quad (19)$$

where

$$\hat{y}(w) = \log \left[\sum_{\alpha} \frac{p_{\alpha} \sigma}{\sqrt{\sigma^2 + \sigma_{\alpha}^2}} \exp \left(\frac{1}{2} \frac{\sigma_{\alpha}^2 w}{\sigma^2 + \sigma_{\alpha}^2} \right) \right], \quad (20)$$

$$w_i = \sum_{AB} u_i^A u_i^B, \quad (21)$$

$$u_i^A = \frac{\sigma s_i^A}{\sigma_{\mathcal{A}}^2}, \quad (22)$$

$$\frac{1}{\sigma^2} = \sum_{\mathcal{A}} \frac{1}{\sigma_{\mathcal{A}}^2}. \quad (23)$$

We can expand the DS as

$$\hat{y}(w) = \sum_{n=0}^{\infty} \frac{1}{n!} \hat{y}^{(n)}(0) \left(\sum_{\mathcal{A}} u^{\mathcal{A}} \right)^{2n}, \quad (24)$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \hat{y}^{(n)}(0) Z(n, N_D), \quad (25)$$

with

$$Z(n, N_D) = \sum_{k_1 + \dots + k_{N_D} = 2n} \binom{2n}{k_1, \dots, k_{N_D}} \prod_{\mathcal{A}=1}^{N_D} (u^{\mathcal{A}})^{k_{\mathcal{A}}}, \quad (26)$$

where $k_{\mathcal{A}}$ are non-negative integers. Following the argument provided in Sec. II we want to remove from Eq. (25) all terms with nonzero expectation value under the hypothesis \mathcal{H}_0 (i.e., in absence of GW signal). Under such a hypothesis, the $u^{\mathcal{A}}$ s are independent normal variables and by virtue of Isserlis' theorem the terms to be canceled are those in the large sum of (26) with all even $k_{\mathcal{A}}$ s.

This can be written explicitly, up to an irrelevant multiplicative constant, as

$$\hat{y}_s = \hat{y} \left[\left(\sum_{\mathcal{A}} u^{\mathcal{A}} \right)^2 \right] - \frac{1}{2^{N_D}} \sum_{\varepsilon_1 = -1, 1} \dots \sum_{\varepsilon_{N_D} = -1, 1} \hat{y} \left[\left(\sum_{\mathcal{A}} \varepsilon_{\mathcal{A}} u^{\mathcal{A}} \right)^2 \right]. \quad (27)$$

A formal proof that this is the desired DS follows.

Theorem III. 1. For a given statistics $\mathcal{S}(u^1, \dots, u^{N_D})$, where $u^{\mathcal{A}} = h^{\mathcal{A}} + n^{\mathcal{A}}$, if $n^{\mathcal{A}}$ are statistically-independent stochastic variables (i.e., the noise in our context) with nonzero even momenta and zero odd momenta, we consider the formal Taylor expansion in powers of $u^{\mathcal{A}}$. A modified statistics where under \mathcal{H}_0 terms with nonzero expectation values are canceled while preserving others is

$$\mathcal{S}_s(u^1, \dots, u^{N_D}) = \mathcal{S}(u^1, \dots, u^{N_D}) + \frac{1}{N_D} \sum_{\varepsilon_1 = -1, 1} \dots \times \sum_{\varepsilon_{N_D} = -1, 1} \mathcal{S}(\varepsilon_1 u^1, \dots, \varepsilon_{N_D} u^{N_D}). \quad (28)$$

Proof.—Let us consider a generic term of the Taylor expansion, which reads

$$(u^1)^{k_1} \dots (u^{N_D})^{k_{N_D}}, \quad (29)$$

where k_i are positive indices. If $h^{\mathcal{A}} = 0$, its expectation value is different from zero if and only if all the k_i s are even. In fact,

$$\langle (n^1)^{k_1} \dots (n^{N_D})^{k_{N_D}} \rangle = \langle (n^1)^{k_1} \rangle \dots \langle (n^{N_D})^{k_{N_D}} \rangle, \quad (30)$$

as the $n^{\mathcal{A}}$ are statistically independent. We observe that

$$\begin{aligned} & \frac{1}{N_D} \sum_{\varepsilon} (\varepsilon_1 n^1)^{k_1} \dots (\varepsilon^{N_D} n^{N_D})^{k_{N_D}} \\ &= \left\{ \frac{1}{N_D} \sum_{\varepsilon} \varepsilon_1^{k_1} \dots \varepsilon_{N_D}^{k_{N_D}} \right\} (n^1)^{k_1} \dots (n^{N_D})^{k_{N_D}}, \end{aligned} \quad (31)$$

and the expression between curly braces is equal to one if and only if all the k_i are even, and equal to zero otherwise. Therefore, the statistics in Eq. (28) cancels the fully even terms, while preserving all the others. ■

It is worth noting that the Gaussian noise assumption is not mandatory; independence and zero odd momenta are required, only. Gaussian, zero average noise is a particular case. We focus on two particular examples, $N_D = 2, 3$. For $N_D = 2$ we get

$$\hat{y}_s = \frac{1}{2} \hat{y}[(u^1 + u^2)^2] - \frac{1}{2} \hat{y}[(u^1 - u^2)^2]. \quad (32)$$

The first two nonzero orders in a power expansion are given by

$$\hat{y}_s = 2\hat{y}'_s(0)u^1u^2 + 2\hat{y}''_s(0)[(u^1)^3u^2 + u^1(u^2)^3] + O((u)^6), \quad (33)$$

and we see that for a linear function \hat{y} the usual Gaussian optimal DS is recovered, while higher-order corrections are proportional to higher order correlation of the measured data. As desired, diagonal terms such as $(u^1)^2$, $(u^1)^2(u^2)^2$ etc., are canceled under the null hypothesis.

We also note here that under the competing hypothesis ($h \neq 0$), $O(u^4)$ terms proportional to h^2n^2 and to h^4 are preserved. The first provide additional information on the GW spectra, boosted by the noises' spectra frequently assumed to be much larger. The other contains information about non-Gaussianity. Similar contributions will appear at higher orders.

In the same way for $N_D = 3$ we have

$$\begin{aligned} \hat{y}_s &= \frac{3}{4} \hat{y}[(u^1 + u^2 + u^3)^2] - \frac{1}{4} \hat{y}[(-u^1 + u^2 + u^3)^2] + \\ & - \frac{1}{4} \hat{y}[(u^1 - u^2 + u^3)^2] - \frac{1}{4} \hat{y}[(u^1 + u^2 - u^3)^2], \end{aligned} \quad (34)$$

and the first two nonzero orders are

$$\begin{aligned} \hat{y}_s &= 2\hat{y}'_s(0)[u^1u^2 + u^2u^3 + u^3u^1] \\ &+ 2\hat{y}''_s(0)[3(u^1)^2u^2u^3 + 3u^1(u^2)^2u^3 + 3u^1u^2(u^3)^2 \\ &+ (u^1)^3(u^2 + u^3) + (u^2)^3(u^1 + u^3) + (u^3)^3(u^1 + u^2)] \\ &+ O((u)^6). \end{aligned} \quad (35)$$

We get again the Gaussian optimal DS at the lowest order, which is just a sum over all pairings of the weighted cross-correlations

$$u^1u^2 + u^2u^3 + u^3u^1 = \sigma^2 \left(\frac{s^1s^2}{\sigma_1^2\sigma_2^2} + \frac{s^2s^3}{\sigma_2^2\sigma_3^2} + \frac{s^3s^1}{\sigma_3^2\sigma_1^2} \right), \quad (36)$$

and no diagonal higher-order terms.

IV. RESULTS

In order to contextualize the performances of the improved DS, we first discuss the connection between the toy model used for the tests and a realistic stochastic background. Let us remind that the former can be described as a superposition of statistically independent events. Each event will produce a contribution to the strain

$$h_{\mu\nu}(\mathbf{x}, t) = u_{\mu\nu}(\mathbf{x}, t - \tau_I, \lambda_I), \quad (37)$$

where τ_I is the event time and $\lambda_I = (\lambda_I^1, \dots, \lambda_I^{N_p})$ an appropriate set of parameters describing the event.

These can be intrinsic (e.g., the chirp mass of a coalescing binary) or extrinsic (e.g., the source luminosity distance or its position in the sky). It is convenient for our modeling to isolate the arrival time and not describe it on the same footing as other parameters.

An useful formalism for the description of a sequence of independent events with a fixed rate is based on the introduction of a function $Q(\lambda)$ [16].

We will not give the explicit derivations (see [16] for details); the interesting final result is that the statistical cumulants of strain field can be written as

$$\langle\langle h_{\mu_1\nu_1}(\mathbf{x}_1, t_1) \cdots h_{\mu_n\nu_n}(\mathbf{x}_n, t_n) \rangle\rangle \quad (38)$$

$$= \int d\tau \int d\lambda Q(\lambda) \quad (39)$$

$$\times u_{\mu_1\nu_1}(\mathbf{x}_1, t_1 - \tau, \lambda) \cdots u_{\mu_n\nu_n}(\mathbf{x}_n, t_n - \tau, \lambda), \quad (40)$$

where we assumed that the average value $\langle h_{\mu\nu}(\mathbf{x}, t) \rangle$ is zero. The quantity $Q(\lambda)d\lambda$ can be interpreted as the rate of events in a given infinitesimal volume of the parameter space. This is more evident if we write

$$Q(\lambda) = \Gamma P(\lambda), \quad (41)$$

where Γ is the event rate and $P(\lambda)$ the probability density of λ over its parameter space. We use this notation to also rewrite (38) as

$$\begin{aligned} &\langle\langle h_{\mu_1\nu_1}(\mathbf{x}_1, t_1) \cdots h_{\mu_n\nu_n}(\mathbf{x}_n, t_n) \rangle\rangle \\ &= \Gamma \int d\tau \overline{u_{\mu_1\nu_1}(\mathbf{x}_1, t_1 - \tau, \lambda) \cdots u_{\mu_n\nu_n}(\mathbf{x}_n, t_n - \tau, \lambda)} \end{aligned} \quad (42)$$

or, in the frequency domain,

$$\begin{aligned} &\langle\langle \tilde{h}_{\mu_1\nu_1}(\mathbf{x}_1, \omega_1) \cdots \tilde{h}_{\mu_n\nu_n}(\mathbf{x}_n, \omega_n) \rangle\rangle \\ &= 2\pi\Gamma\delta(\omega_1 + \cdots + \omega_n) \overline{\tilde{u}_{\mu_1\nu_1}(\mathbf{x}_1, \omega_1, \lambda) \cdots \tilde{u}_{\mu_n\nu_n}(\mathbf{x}_n, \omega_n, \lambda)}. \end{aligned} \quad (43)$$

Here \bar{X} is the average over the parameters λ of X .

Note that, from an observational point of view, we are interested in cumulants of the strain measured by the detectors, namely

$$h^A(t) = \int D_{ij}^A(t, t') h_{ij}(t', \mathbf{x}^A(t')) dt'. \quad (44)$$

Here D_{ij}^A is a generalized detector tensor of the A th interferometer located in \mathbf{x}^A , where we included the time dependency connected to the detector's movement and a whitening transformation that can be chosen in such a way to normalize the signal to a white noise spectrum. Moments of measured strains defined in Eq. (44) can be easily recovered from (38).

All the information of astrophysical interest are encoded in $Q(\lambda)$. Q functions are additive; if a given stochastic background is the sum of several contributions, each described by $Q^{(a)}(\lambda)$, we have

$$Q(\lambda) = \sum_a Q^{(a)}(\lambda). \quad (45)$$

This mathematical description let us highlight an important fact; let us suppose that the rate is increased $Q(\lambda) \rightarrow \eta Q(\lambda)$, and at the same time individual events amplitudes are scaled accordingly with $u_{\mu\nu} \rightarrow \eta^{-1/2} u_{\mu\nu}$. The second order cumulant $\langle\langle h_{\mu_1\nu_1}(\mathbf{x}_1, t_1) h_{\mu_2\nu_2}(\mathbf{x}_2, t_2) \rangle\rangle$ does not change, while higher-order ones are rescaled as

$$\begin{aligned} &\langle\langle h_{\mu_1\nu_1}(\mathbf{x}_1, t_1) \cdots h_{\mu_n\nu_n}(\mathbf{x}_n, t_n) \rangle\rangle \\ &\rightarrow \eta^{(2-n)/2} \langle\langle h_{\mu_1\nu_1}(\mathbf{x}_1, t_1) \cdots h_{\mu_n\nu_n}(\mathbf{x}_n, t_n) \rangle\rangle \end{aligned} \quad (46)$$

and become less and less important in the $\eta \rightarrow \infty$ limit. This offers a straightforward interpretation; when the background is generated by a very large number of weak events, it becomes a Gaussian stochastic field—only its second-order cumulant, which is related to spectral characteristics, is relevant.

Let us now connect the formalism above with the proposed toy model. Considering $\sigma_- = 0$ [Eq. (17)] we have

$$\tilde{h}(\omega) = h\delta t \quad (47)$$

and we can use Eq. (43) directly to get the even cumulants

$$\begin{aligned} \langle\langle \tilde{h}^{A_1}(\omega_1) \cdots \tilde{h}^{A_n}(\omega_n) \rangle\rangle &= 2\pi\Gamma\delta(\omega_1 + \cdots + \omega_n)\bar{h}^n \\ &= 2\pi\Gamma(n-1)!! p_+ \sigma_+^n \delta t^n \delta(\omega_1 + \cdots + \omega_n) \end{aligned} \quad (48)$$

while the odd ones are zero. The general case in Eq. (18) can be obtained by substituting in this result $\sigma_+^2 \rightarrow \sigma_+^2 - \sigma_-^2$ and a by adding the contribution of Gaussian background with variance σ_-^2 . This contributes only to the $n = 2$ cumulant and we get

$$\begin{aligned} \langle\langle \tilde{h}^{A_1}(\omega_1) \cdots \tilde{h}^{A_n}(\omega_n) \rangle\rangle &= 2\pi\Gamma(n-1)!! (p_+(\sigma_+^2 - \sigma_-^2)^{n/2} + \sigma_-^2 \delta_{n,2}) \\ &\quad \times \delta t^n \delta(\omega_1 + \cdots + \omega_n). \end{aligned} \quad (49)$$

In particular the power spectrum is flat and given by

$$S_h(f) = \sigma_h^2 \delta t \quad (50)$$

which corresponds to an adimensional energy density for logarithmic interval of frequency

$$\Omega_{GW}(f) = \frac{4\pi^2}{3H_0^2} f^3 \sigma_h^2 \delta t. \quad (51)$$

It is important to note that the distribution of signal amplitudes for each event depends on the specific details of $Q(\lambda)$: e.g., it is sensitive to the redshift z distribution of the sources. The signal power spectrum depends only on σ_h , while higher-order momenta carry information, and therefore can distinguish additional signal properties.

The noise power spectrum can be written similarly as $S_n^A(f) = \sigma_A^2 \delta t$. Finally, we write the signal-to-noise ratio for a pair of detectors as a function of the toy model parameters

$$\text{SNR}^2 = \frac{8}{25} T \Delta f \frac{\sigma_h^4}{\sigma_1^2 \sigma_2^2}, \quad (52)$$

where T is the observing time and Δf is the sensitivity bandwidth of the detectors. We stress that this is an appropriate figure of merit for the performances of the Gaussian analysis.

In order to compare DSs' performances, we follow the procedure in [12] closely. In addition, we observe that momenta of the DS can be evaluated as an explicit integral, therefore we avoid a direct simulation of the data under \mathcal{H}_1 . For the ‘‘Exact’’ DS we write

$$\begin{aligned} \langle \hat{y}^K \rangle &= \int \cdots \int \prod_A ds^A \log \left\{ \sum_{\alpha=+,-} \frac{p_\alpha \sigma}{\sqrt{\sigma_\alpha^2 + \sigma^2}} \right. \\ &\quad \times \exp \left[\frac{\sigma_\alpha^2}{2(\sigma_\alpha^2 + \sigma^2)} \left(\sigma \sum_A s^A \right) \right] \left. \right\}^K \\ &\quad \times \left[\sum_{\alpha=+,-} \frac{p_\alpha e^{-\frac{1}{2}[C_\alpha^{-1}]^{AB} s^A s^B}}{\sqrt{(2\pi)^{N_D} \det C_\alpha}} \right], \end{aligned} \quad (53)$$

where C_\pm are matrices in detector space with entries defined by

$$C_\pm^{AB} = \sigma_A^2 \delta_{AB} \quad \text{under } \mathcal{H}_0, \quad (54)$$

$$C_\pm^{AB} = \sigma_A^2 \delta_{AB} + \sigma_\pm^2 \quad \text{under } \mathcal{H}_1. \quad (55)$$

and they describe the two multivariate Gaussian distributions entering the competing hypothesis. Similar expressions can be obtained for the other DSs listed in Sec. I (detailed derivations are provided in [12]).

A. Performance comparison

When a series of N_d data is collected from each detector, the DS is the sum of Eq. (19) and for large enough N_d , \hat{Y} is approximately Gaussian distributed. Using this property and the connected momenta being proportional to N_d , i.e.,

$$\mu|_{N_d=N} = N\mu|_{N_d=1}, \quad (56)$$

$$\sigma^2|_{N_d=N} = N\sigma^2|_{N_d=1}, \quad (57)$$

we write a relation between false alarm probability, detection probability, and N in the form

$$\begin{aligned} P_D &= \frac{1}{2} \operatorname{erfc} \left(\frac{\sigma_{\mathcal{H}_0}}{\sigma_{\mathcal{H}_1}} \Big|_{N_d=1} \operatorname{erfc}^{-1}(2P_{FA}) + \right. \\ &\quad \left. - \sqrt{\frac{N}{2}} \frac{\mu_{\mathcal{H}_1} - \mu_{\mathcal{H}_0}}{\sigma_{\mathcal{H}_1}} \Big|_{N_d=1} \right). \end{aligned} \quad (58)$$

We note that this relation is valid for all the DS considered, for a large enough number of data N_d (or, equivalently, for a large enough data taking time $T = \delta t N_d$). It follows that a naive definition of a signal-to-noise ratio

$$\text{SNR} = \frac{\mu_{\mathcal{H}_1} - \mu_{\mathcal{H}_0}}{\sigma_{\mathcal{H}_1}} \quad (59)$$

can always be used. However, it is also clear that this signal-to-noise ratio does not describes completely the performances of a given DS, as the ratio $\frac{\sigma_{\mathcal{H}_0}}{\sigma_{\mathcal{H}_1}}$ is also relevant. For this reason, we argue that the best and unambiguous figure of merit for a comparison between the Gaussian and

the non-Gaussian case is the value of the detection probability.

Performances are presented in Figs. 1, 2, 4, and 5 using Eq. (58) for $N_D = 2, 3, 4$ and 5, respectively. We plot the detection probability against the number of data points N , for $P_{FA} = 10^{-15}$. The advantage of the procedure is that it gives precise estimates of the relevant quantities also for very small values of P_{FA} , otherwise inaccessible through simulated data. On the other hand, the functional form of Eq. (58) is based on the assumption that the statistic is Gaussian, so it should be considered reliable only for high enough number of data.

Every figure shows several plots with different choices of σ_+/σ_h and σ_-/σ_h , to explore different levels of non-Gaussianity, from negligible (bottom right panel) to strong (top left panel). The ‘‘Improved’’ DS systematically outperforms the ‘‘Scrambled’’ one studied in [12], with greater probabilities of detections for every N and level of signal Gaussianity. In particular, the ‘‘Improved’’ performances are always equal to or better than the ‘‘Gaussian’’ ones. The ‘‘Improved’’ DS does not suffer of the loss of performances affecting the ‘‘Scrambled’’ one, even for signals very close to Gaussian.

Remarkably, increasing the number of DSs (i.e., in Fig. 2) we see that the ‘‘Improved’’ DS has performances similar to the ‘‘Exact’’ one, showing an important gain in having greater N_D . The improved DS performances converge to those of the ‘‘Exact’’ one for large N_D .

A semiquantitative understanding of this behavior is obtained as follows: we notice that for a given order $O(u^{2n})$, there are

$$N_{n,N_D} = \binom{2n + (N_D - 1)}{2n} \quad (60)$$

possible terms in the Taylor expansion of a DS. The number of terms with even powers of all variables is

$$N_{n,N_D}^{(s)} = \binom{n + (N_D - 1)}{n} \quad (61)$$

and the ratio $N_{n,N_D}^{(s)}/N_{n,N_D}$ scales as N_D^{-n} in the large N_D limit,

$$\frac{N_{n,N_D}^{(s)}}{N_{n,N_D}} = \frac{2^n (2n - 1)!!}{\prod_{i=1}^n (N_D + 2n - i)}. \quad (62)$$

This suggests that subtracted terms impact less and less the performances when N_D grows larger, the ‘‘Improved’’ DS getting closer and closer to the ‘‘Exact’’ one. We note that a network with $N_D = 5$ is not unrealistic [17], and is already large enough to make the ‘‘Exact’’ and ‘‘Improved’’ DSs nearly equivalent.

A further argument in support of such behavior follows: we rewrite Eq. (58) as

$$N_d = \left[\frac{\sigma_{\mathcal{H}_1}}{\mu_{\mathcal{H}_1} - \mu_{\mathcal{H}_0}} \Big|_{N_d=1} \times \left(\frac{\sigma_{\mathcal{H}_0}}{\sigma_{\mathcal{H}_1}} \Big|_{N_d=1} \operatorname{erfc}^{-1}(2P_{FA}) - \operatorname{erfc}^{-1}(2P_D) \right) \right]^2, \quad (63)$$

and we define N_G and N_I its value computed for the ‘‘Gaussian’’ and ‘‘Improved’’ DSs, respectively. Therefore, the ratio N_G/N_I gives the multiplicative factor for the number of data needed for the ‘‘Gaussian’’ DS to reach the same detection probability of the ‘‘Improved’’ one, for fixed competing hypotheses and probability of false alarm.

We show a plot of such ratio at fixed $P_D = 0.5$ and $P_{FA} = 10^{-15}$ in Fig. 3 (right panel) for a set of N_D values. The ratios σ_+/σ_h and σ_-/σ_h are shown on the vertical and horizontal axis, respectively. The improved statistic has always better performances than the Gaussian statistic, as N_G/N_I is globally greater than one. Such improvement grows larger in the upper-left corner, where the signal kurtosis is larger, as expected. In the right panel of Fig. 3 we represent the relation between the reduced kurtosis and N_G/N_I , for a uniform sampling in the parameters σ_+/σ_h , σ_-/σ_h of the region shown in the right panel. Results for $N_D = 2, 3, 4, 5$ can be compared, and we see again that the performance of the ‘‘Improved’’ DS relative to the ‘‘Gaussian’’ one increase with N_D . This is equivalent to a comparison of the time required to achieve a detection at the same level of significance using the two-detection statistics. From Fig. 3 we see that a confident detection would require a data taking time three times larger with the standard ‘‘Gaussian’’ DS compared with the ‘‘Improved’’ DS we introduced with two detector and a mild non-Gaussianity. Larger improvements are obtained with larger non-Gaussianities and more than two detectors.

We note that having $N_G/N_I \geq 1$ is *a priori* expected. In fact, the ‘‘Improved’’ version of the optimal statistic is obtained subtracting nonzero mean terms from the ‘‘Exact’’ one, which is optimal in the Neyman-Pearson sense. In the Gaussian limit the ‘‘Improved’’ statistic becomes exactly the ‘‘Gaussian’’ one, as discussed in the previous section [see Eq. (36) for a specific example], and can be considered its natural generalization. However, we do not provide a rigorous proof here.

B. Subtraction for realistic backgrounds

Given the effectiveness of the new procedure, we explore its application to a realistic non-Gaussian stochastic background. We consider the model

$$s_i^A = n_i^A + h_i^A, \quad (64)$$

where the GW signals measured by each detector are now different, and the joint probability distribution reads

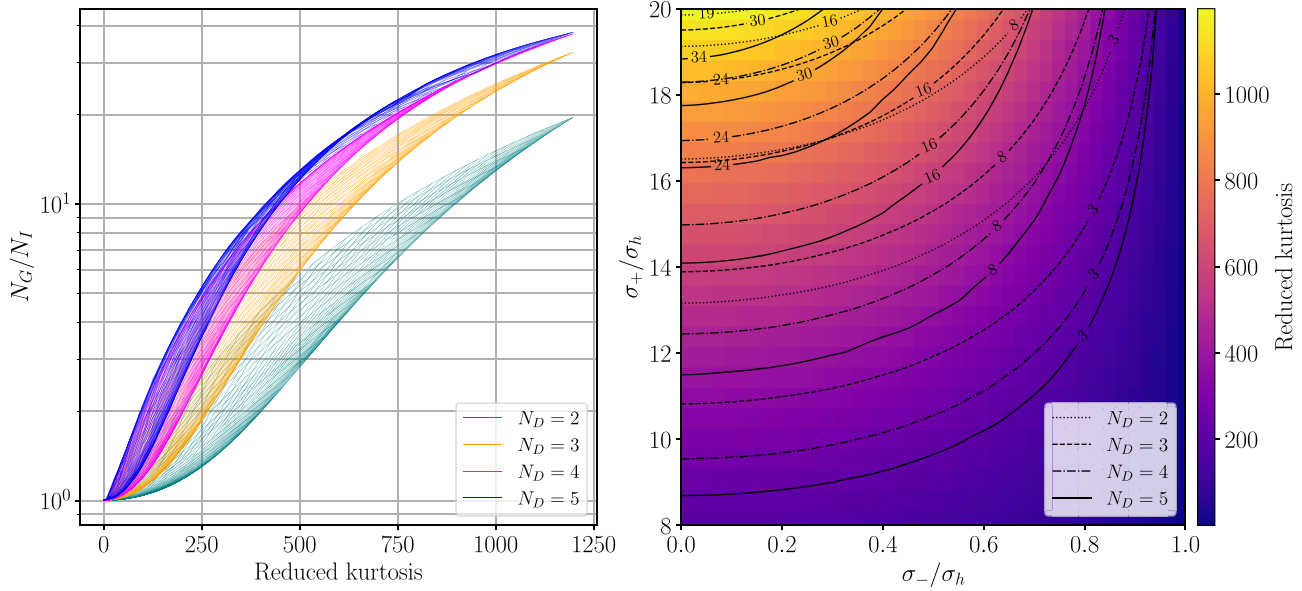


FIG. 3. (Right panel) Contour levels of the improvement factor N_G/N_I in a range of the model parameters $\sigma_+/\sigma_h, \sigma_-/\sigma_h$. Solid, dot-dashed, dashed, and dotted lines denote the improvement factor for $N_D = 2, 3, 4, 5$, respectively. Pixels shading denotes the model-reduced kurtosis. We observe N_G/N_I greater than one, up to a few tens, over the explored parameter space, systematically increasing for increasing number of detectors. (Left panel) The above performances are shown as a function of the reduced kurtosis and N_G/N_I . Uniform grids over the signal parameter space are plotted as teal, orange, purple, and blue meshes for $N_D = 2, 3, 4$ and 5 , respectively. We fix as fiducial values $P_D = 0.5$ and $P_{FA} = 10^{-15}$.

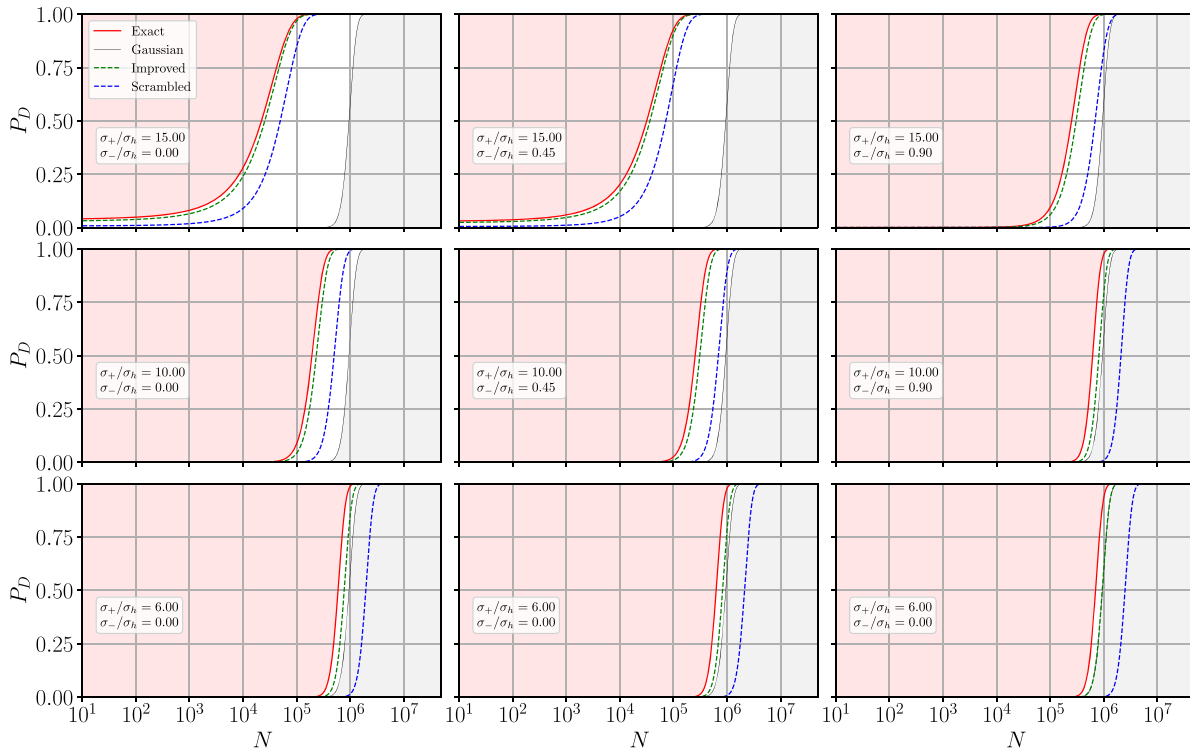


FIG. 4. Comparison between the detection probability of the considered DSs as a function of the number of used data N . Several different signal models are considered, parametrized by σ_+/σ_h and σ_-/σ_h . Here $N_D = 4$ and $P_{FA} = 10^{-15}$.

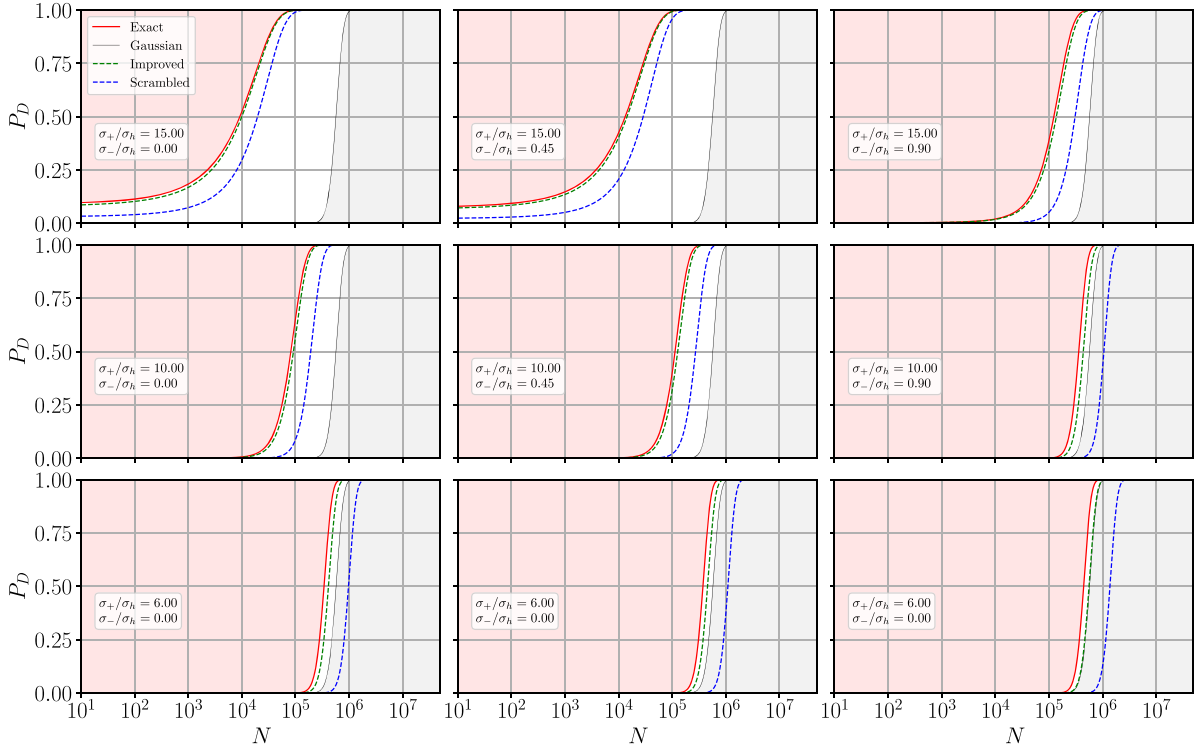


FIG. 5. Comparison between the detection probability of the considered DSs as a function of the number of used data N . Several different signal models are considered, parametrized by σ_+/σ_h and σ_-/σ_h . Here $N_D = 5$ and $P_{FA} = 10^{-15}$.

$$dP = \mathcal{D}(h) \prod_{A,i} dh_i^A, \quad (65)$$

where $\mathcal{D}(h)$ is some generic function of the data h_i^A (a probability density) and $\prod_{A,i} dh_i^A$ a measure in the space (of dimension $N_D \times N$) of possible signal time series.

We first observe that the statistical properties of h are irrelevant, because we only want to obtain the cancellation of diagonal terms under the \mathcal{H}_0 hypothesis. However, realistic detector noises can have nontrivial time-domain correlations, and non-Gaussian contaminations are expected. A generic term of a Taylor expansion of the DSs with degree $\kappa_1 + \dots + \kappa_{N_D}$ reads

$$\underbrace{(n_{i,1}^1 \dots n_{i,\kappa_1}^1)}_{\text{Detector 1}} \dots \underbrace{(n_{i_{N_D},1}^{N_D} \dots n_{i_{N_D},\kappa_{N_D}}^{N_D})}_{\text{Detector } N_D}, \quad (66)$$

and its expectation is

$$\langle n_{i,1}^1 \dots n_{i,\kappa_1}^1 \rangle \dots \langle n_{i_{N_D},1}^{N_D} \dots n_{i_{N_D},\kappa_{N_D}}^{N_D} \rangle \quad (67)$$

as we assume again statistically independent noises across different detectors. After applying the subtraction procedure defined in Eq. (27) we get

$$\begin{aligned} & (n_{i_{1,1}}^1 \dots n_{i_{1,\kappa_1}}^1) \dots (n_{i_{N_D},1}^{N_D} \dots n_{i_{N_D},\kappa_{N_D}}^{N_D}) \\ & \rightarrow \left(1 - \frac{1}{2^{N_D}} \sum_{\mathbf{\epsilon}} \epsilon_1^{\kappa_1} \dots \epsilon_{N_D}^{\kappa_{N_D}} \right) \\ & \quad \times (n_{i_{1,1}}^1 \dots n_{i_{1,\kappa_1}}^1) \dots (n_{i_{N_D},1}^{N_D} \dots n_{i_{N_D},\kappa_{N_D}}^{N_D}). \end{aligned}$$

As in the simplified case, a given term is cancelled if all the $\kappa_{A,S}$ are even and is preserved otherwise. The expectation values of all the preserved contributions are zero, unless some odd-order noise momenta are non-null. Therefore, the subtraction procedure is effective in a very general sense, and is also applicable when non-Gaussian contaminations (without nonzero odd momenta) are present. In particular, this procedure can be incorporated easily in the general DS defined in Eq. (17) of [12]. However, it should be noted that non-Gaussian noise contributions can change the expectation value of the statistic under \mathcal{H}_1 .

V. CONCLUSIONS AND PERSPECTIVES

The ‘‘Improved’’ DS introduced in this paper can be seen as the natural generalization to the non-Gaussian case of the optimal statistic used for a Gaussian stochastic background. As a study of a realistic model is computationally expensive, we performed an exploration based on a simplified toy model, and focused in this manuscript on our findings. The toy model we used to quantify its performances is

simplified. In particular, it completely neglects the non-trivial time structure of the signals associated to a realistic event. However, we showed that the core procedure to remove diagonal contributions is applicable to the general case, since it is insensitive to time or frequency structure, and only performs a transformation in the detector space.

Performance results are promising, and we expect that the improvement with respect to the “Gaussian” DS will be preserved in a realistic situation. We observed that the “Improved” DS becomes nearly equivalent to the “Exact” one for a large enough (although reasonable) number of detectors in the network. For completeness, it should be noted that the improvement that can be obtained in a real scenario depends on the details of the background non-Gaussianity, so one should extrapolate with care our findings, and ultimately support them through direct investigation.

From a practical point of view, the main difference in a realistic scenario where no analytical expression for the “Improved” DS is available, arises from the computational cost in evaluating it as discussed in [12] [e.g., the discussion following Eq. (17) therein], which requires an efficient numerical procedure. The computational cost is related to the number of configurations needed for an accurate statistic estimate, which we investigated in our

previous work using the importance-sampling procedure. On this ground, we expect a large number of detectors to help in reducing the effective computational cost.

Computational cost estimate of the method in a realistic case is a key step that needs to be fully addressed to assess the applicability of our method. We argue that an expensive, although manageable, computational cost would certainly be justified by the prospect of an earlier and more confident detection of an astrophysical stochastic background, e.g., generated by a superposition of binary coalescence events, within reach in the near future [1]. We expect that no major additional obstructions will forbid the application of the method to other detectors, such as LISA or pulsar timing arrays.

ACKNOWLEDGMENTS

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* [18] and of the following Python [19] packages for the analysis, post-processing and production of results throughout: Matplotlib [20], NumPy [21], SciPy [22].

-
- [1] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese *et al.*, *Phys. Rev. D* **104**, 022004 (2021).
 - [2] T. Regimbau, *Res. Astron. Astrophys.* **11**, 369 (2011).
 - [3] T. Damour and A. Vilenkin, *Phys. Rev. Lett.* **85**, 3761 (2000).
 - [4] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration, *Phys. Rev. X* **13**, 011048 (2023).
 - [5] S. Drasco and E. E. Flanagan, *Phys. Rev. D* **67**, 082003 (2003).
 - [6] E. Thrane, *Phys. Rev. D* **87**, 043009 (2013).
 - [7] R. Smith and E. Thrane, *Phys. Rev. X* **8**, 021019 (2018).
 - [8] L. Martellini and T. Regimbau, *Phys. Rev. D* **89**, 124009 (2014).
 - [9] N. Seto, *Astrophys. J. Lett.* **683**, L95 (2008).
 - [10] N. Seto, *Phys. Rev. D* **80**, 043003 (2009).
 - [11] J. D. Romano and N. J. Cornish, *Living Rev. Relativity* **20**, 2 (2017).
 - [12] R. Busicchio, A. Ain, M. Ballelli, G. Cella, and B. Patricelli, *Phys. Rev. D* **107**, 063027 (2023).
 - [13] B. Allen and J. D. Romano, *Phys. Rev. D* **59**, 102001 (1999).
 - [14] A. I. Renzini, B. Goncharov, A. C. Jenkins, and P. M. Meyers, *Galaxies* **10**, 34 (2022).
 - [15] J. Neyman and E. S. Pearson, *Phil. Trans. R. Soc. A* **231**, 289 (1933).
 - [16] N. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier Science Publishers, Amsterdam, 1992).
 - [17] B. P. Abbott *et al.*, *Living Rev. Relativity* **23**, 3 (2020).
 - [18] W. R. Inc., *Mathematica*, Version 13.1 (2022), Champaign, IL.
 - [19] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).
 - [20] J. D. Hunter, *Comput. Sci. Eng.* **9**, 90 (2007).
 - [21] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers *et al.*, *Nature (London)* **585**, 357 (2020).
 - [22] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland *et al.*, *Nat. Methods* **17**, 261 (2020).