

**Detection methods for non-Gaussian gravitational wave stochastic backgrounds**

Steve Drasco\* and Éanna É. Flanagan†

*Newman Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14853*

(Received 10 October 2002; published 24 April 2003)

A gravitational wave stochastic background can be produced by a collection of independent gravitational wave events. There are two classes of such backgrounds, one for which the ratio of the average time between events to the average duration of an event is small (i.e., many events are on at once), and one for which the ratio is large. In the first case the signal is continuous, sounds something like a constant *hiss*, and has a Gaussian probability distribution. In the second case, the discontinuous or intermittent signal sounds something like popcorn popping, and is described by a non-Gaussian probability distribution. In this paper we address the issue of finding an optimal detection method for such a non-Gaussian background. As a first step, we examine the idealized situation in which the event durations are short compared to the detector sampling time, so that the time structure of the events cannot be resolved, and we assume white, Gaussian noise in two collocated, aligned detectors. For this situation we derive an appropriate version of the maximum likelihood detection statistic. We compare the performance of this statistic to that of the standard cross-correlation statistic both analytically and with Monte Carlo simulations. In general the maximum likelihood statistic performs better than the cross-correlation statistic when the stochastic background is sufficiently non-Gaussian, resulting in a gain factor in the minimum gravitational-wave energy density necessary for detection. This gain factor ranges roughly between 1 and 3, depending on the duty cycle of the background, for realistic observing times and signal strengths for both ground and space based detectors. The computational cost of the statistic, although significantly greater than that of the cross-correlation statistic, is not unreasonable. Before the statistic can be used in practice with real detector data, further work is required to generalize our analysis to accommodate separated, misaligned detectors with realistic, colored, non-Gaussian noise.

DOI: 10.1103/PhysRevD.67.082003

PACS number(s): 04.80.Nn, 04.30.Db, 07.05.Kf, 95.55.Ym

**I. INTRODUCTION AND SUMMARY**

Along with a new generation of gravitational wave detectors around the world [1–4], detection algorithms for a variety of sources are nearing completion. If the signals from these sources are detected, physicists stand to harvest unprecedented quantities of observational information concerning the nature of gravitation and the cosmos as a whole. The fruit of this harvest will be the outputs of detection algorithms. In this paper we introduce an algorithm designed for nearly optimal detection of a class of gravitational wave stochastic backgrounds. The non-Gaussian nature of this class of backgrounds causes the algorithm presented here to differ from the well studied cross-correlation based algorithms which are nearly optimal for Gaussian backgrounds.

**A. Gravitational wave stochastic backgrounds**

Consider a large collection of similar gravitational wave sources. If we cannot resolve the individual signals produced by these sources and know only their statistical properties, the signals form a stochastic background. A wide variety of candidate sources of gravitational wave stochastic backgrounds have been studied (for an excellent general review

see Ref. [5]). These include high redshift supernovae [6,7], the first stars or so-called population III objects [8], rapidly rotating young neutron stars [9,10], early universe phase transitions and cosmic strings [11,12], inflation [13], and high redshift compact binaries [14].

Detecting a gravitational wave stochastic background produced by any one of these candidate sources could provide information on a variety of topics ranging from the evolution of the star formation rate [15] to the numbers and sizes of posited extra dimensions [16]. Because of this, stochastic backgrounds have long been thought to be among the most interesting possible types of gravitational radiation.

**B. Gaussian stochastic backgrounds**

In order to develop detection methods, it is traditionally assumed that the individual events making up a background are uncorrelated and sufficiently frequent for the background to be Gaussian. That is, it is assumed that the conditions for applicability of the central limit theorem are satisfied.

Unlike electromagnetic waves, gravitational waves cannot be screened from a detector. Using a single gravitational wave detector, there is no practical way to distinguish between detector noise and a stochastic background of gravitational waves. As a consequence the sensitivity of a single detector to gravitational backgrounds is severely limited. By comparing the outputs of multiple detectors, sensitivity levels can be enhanced. Michelson [17] was the first to give a detailed description of such a detection method for a Gaussian stochastic background of gravitational waves in the pres-

\*Email address: sd68@cornell.edu

†Email address: eef3@cornell.edu. Also at Radcliffe Institute for Advanced Study, Putnam House, 10 Garden Street, Cambridge, MA 02138.

ence of Gaussian detector noise. His detection strategy and its later refinements [18–20] are often referred to as the cross-correlation method. Recently the cross-correlation method has been modified to treat more realistic detectors which themselves have sources of non-Gaussian noise [21–23].

We now briefly review the cross-correlation method. Consider two gravitational wave detectors. The output of each detector is a collection of dimensionless strain measurements. Suppose that  $N$  such measurements are made with each detector at regular time intervals. Denote these measurements by a  $N \times 2$  matrix  $h$  with components  $h_i^k$ , where  $i=1,2$  labels the detector, and  $k=1,2,\dots,N$  is a time index. To determine whether or not the data  $h$  contains some desired signal, one usually compares the value of some detection statistic  $\Gamma(h)$  to a threshold value  $\Gamma_*$ . That is, if  $\Gamma(h) > \Gamma_*$  one concludes that a signal is present and otherwise one concludes that no signal is present. A detection statistic is said to be optimal if it yields the smallest probability of mistakenly concluding a signal is present (false alarm probability) after choosing a threshold which fixes the probability for mistakenly concluding a signal is absent (false dismissal probability).

Assume that the two detectors are collocated and aligned, and that each detector has white Gaussian noise with vanishing mean with no correlations between the two detectors. Then the standard cross-correlation detection statistic  $\Lambda_{CC}$  for a Gaussian signal is

$$\Lambda_{CC}(h) = \frac{\hat{\alpha}^2}{\bar{\sigma}_1 \bar{\sigma}_2}, \quad (1.1)$$

where

$$\hat{\alpha}^2 = \bar{\alpha}^2 \theta(\bar{\alpha}^2), \quad (1.2)$$

$$\bar{\alpha}^2 = \frac{1}{N} \sum_{k=1}^N h_1^k h_2^k, \quad (1.3)$$

$$\bar{\sigma}_i^2 = \frac{1}{N} \sum_{k=1}^N (h_i^k)^2, \quad (1.4)$$

for  $i=1,2$ , and  $\theta(x)$  is the Heaviside step function defined by

$$\theta(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases} \quad (1.5)$$

This statistic is nearly optimal and can be derived from a maximum likelihood framework (see Sec. III B). The subscript CC in  $\Lambda_{CC}$  denotes ‘‘cross correlation.’’ The generalization of this statistic to allow for colored noise and non-collocated, non-aligned detectors is discussed in Refs. [17–20].

### C. Non-Gaussian stochastic backgrounds

A particular class of events will produce a Gaussian background if, on average, at any given moment, many individual events are arriving at the detector. However, if the ratio of average time between events to the average duration of events is large, then there are long stretches of ‘‘silence’’ or time during which no events arrive at the detector. The resulting stochastic background is non-Gaussian as the conditions for the applicability of the central limit theorem are not satisfied. Recent work has suggested that some candidate gravitational wave stochastic backgrounds, of both cosmological and astrophysical origin, may be non-Gaussian [7,8,11]. However, predictions concerning the properties of most gravitational wave background sources rely heavily on theoretical arguments which extrapolate well beyond observational support. Such extrapolations are always in some sense speculative. It is conceivable that backgrounds predicted to be Gaussian may in fact turn out to be non-Gaussian, or vice versa.

In Sec. III C below, we apply a maximum likelihood framework to derive a detection statistic for a particular model of non-Gaussian stochastic background, which we now describe. Let  $h_i^k$  be the outputs of two collocated aligned gravitational wave detectors with white, zero-mean, Gaussian noise with no correlations between the two detectors. The detector outputs  $h_i^k$  consist of noise  $n_i^k$  together with a common signal  $s^k$ :

$$h_1^k = n_1^k + s^k \quad (1.6)$$

$$h_2^k = n_2^k + s^k.$$

We wish to detect a non-Gaussian signal  $s^k$  composed of long stretches of silence which separate short bursts whose amplitudes are Gaussianly distributed, and whose durations are smaller than the detector resolution time (see Fig. 4). We therefore assume that each signal sample  $s^k$  is statistically independent with probability distribution [cf. Eq. (3.19) below]

$$p(s) = \xi \frac{1}{\sqrt{2\pi}\alpha} \exp\left[-\frac{s^2}{2\alpha^2}\right] + (1-\xi)\delta(s). \quad (1.7)$$

The parameter  $\xi$  is what we call the *Gaussianity parameter* of the stochastic background; it is the probability that, at any randomly chosen time, a burst is present in the detector. Thus  $\xi$  takes values in the interval  $0 \leq \xi \leq 1$ , and if  $\xi=1$  then the background is Gaussian. The parameter  $\xi$  can also be thought of as the duty cycle of the background. The parameter  $\alpha$  in Eq. (1.7) is the rms amplitude of the bursts.

Our nearly-optimal detection statistic  $\Lambda_{ML}^{NG}$  for the signal model (1.7) is given by [cf. Eq. (3.22) below]

$$\Lambda_{\text{ML}}^{\text{NG}}(h) = \max_{0 < \xi \leq 1} \max_{\alpha > 0} \max_{\sigma_1 \geq 0} \max_{\sigma_2 \geq 0} \prod_{k=1}^N \left\{ \frac{\bar{\sigma}_1 \bar{\sigma}_2 \xi}{\sqrt{\sigma_1^2 \sigma_2^2 + \sigma_1^2 \alpha^2 + \sigma_2^2 \alpha^2}} \exp \left[ \frac{\left( \frac{h_1^k}{\sigma_1^2} + \frac{h_2^k}{\sigma_2^2} \right)^2}{2 \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\alpha^2} \right)} - \frac{(h_1^k)^2}{2\sigma_1^2} - \frac{(h_2^k)^2}{2\sigma_2^2} + 1 \right] \right. \\ \left. + \frac{\bar{\sigma}_1 \bar{\sigma}_2}{\sigma_1 \sigma_2} (1 - \xi) \exp \left[ - \frac{(h_1^k)^2}{2\sigma_1^2} - \frac{(h_2^k)^2}{2\sigma_2^2} + 1 \right] \right\}. \quad (1.8)$$

Here the quantities  $\bar{\sigma}_1$  and  $\bar{\sigma}_2$  are defined by Eq. (1.4). The values of  $\xi$ ,  $\alpha^2$ ,  $\sigma_1^2$  and  $\sigma_2^2$  which achieve the maximum in Eq. (1.8) are, respectively, estimators of the signal's Gaussianity parameter, the variance of the signal events, and the variances of the noise in the two detectors. If we calculate the quantity (1.8) at  $\xi=1$ , instead of maximizing over  $\xi$ , the result is a statistic which is equivalent to the standard cross-correlation statistic  $\Lambda_{\text{CC}}$ .

The subscript ML on  $\Lambda_{\text{ML}}^{\text{NG}}$  stands for ‘‘maximum likelihood,’’ while the superscript NG stands for ‘‘non-Gaussian statistic.’’ The superscript NG does *not* necessarily mean that one is considering a non-Gaussian signal; both of the statistics  $\Lambda_{\text{CC}}$  and  $\Lambda_{\text{ML}}^{\text{NG}}$  can be applied to data containing either a Gaussian signal or a non-Gaussian signal.

If the burst-amplitude parameter  $\alpha$  is sufficiently large and the bursts are well separated in time, then the individual bursts can be seen in the detector output. In this case one could use, for example, the simple burst statistic<sup>1</sup>

$$\Lambda_{\text{B}} \equiv \max_{1 \leq k \leq N} |h_1^k| \quad (1.9)$$

on the data from detector 1 to detect the signal. The burst statistic (1.9) and the cross-correlation statistic  $\Lambda_{\text{CC}}$  are used as references for comparison for the maximum likelihood statistic below.

#### D. Main results

There are two main results in this paper. The first result is the detection statistic  $\Lambda_{\text{ML}}^{\text{NG}}$  given by Eq. (1.8), which is derived in Sec. III C. This statistic is nearly optimal for the

<sup>1</sup>In reality the statistic (1.9) would be especially susceptible to non-Gaussian noise bursts in the detector and so would not be used in practice; instead one would need search for events where  $|h_1^k|$  and  $|h_2^k|$  are simultaneously large. In this paper we restrict attention for simplicity to Gaussian detector noise; it will be important for future more general analyses to allow for (uncorrelated) non-Gaussian noise components in the two detectors.

detection of a class of non-Gaussian gravitational wave stochastic backgrounds incident on a pair of idealized detectors.

The second main result, summarized in Figs. 1 and 2, is a comparison of the performances of the maximum likelihood

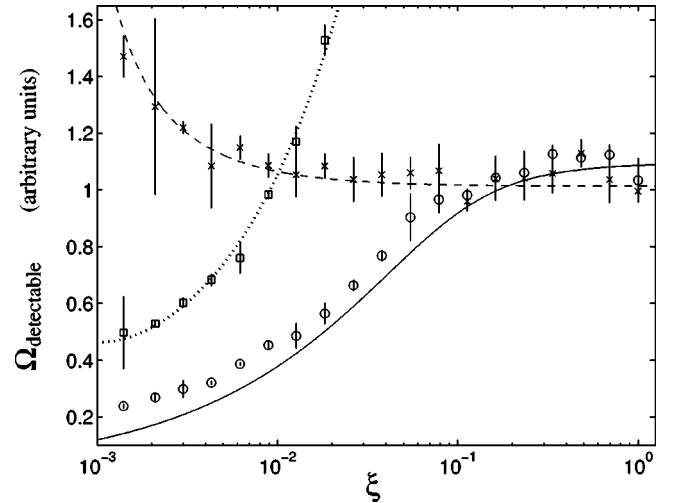


FIG. 1. This plot shows the minimum gravitational-wave energy density  $\Omega_{\text{detectable}}$  necessary for detection, for several different detection statistics, as a function of the background's Gaussianity parameter  $\xi$ . The Gaussianity parameter  $\xi$  is the probability that, at any randomly chosen time, the waves from an event are incident on the detectors, and thus takes values in the interval  $0 \leq \xi \leq 1$ . For a Gaussian background  $\xi=1$ . The circles are the results of our Monte Carlo simulations for the maximum likelihood statistic  $\Lambda_{\text{ML}}^{\text{NG}}$ , and the solid curve shows the approximate theoretical prediction (C46) and (C55) for this statistic (expected to be accurate only to within a few tens of percent). The crosses are the Monte Carlo results for the cross-correlation statistic  $\Lambda_{\text{CC}}$ , and the dashed curve shows the theoretical prediction (4.21) for this statistic. Finally the squares are the Monte Carlo results for the burst statistic (1.9), and the dotted curve shows the corresponding theoretical prediction given by Eqs. (4.9) and (4.10). For each statistic, the vertical error bars on the Monte Carlo simulation results give the fluctuations computed from 4 different runs, each with 2000 trials. The number of data points is  $N=10^4$ , and the false alarm and false dismissal probabilities are both 0.1. A detailed description of the simulations and the analytical predictions can be found in Sec. IV.

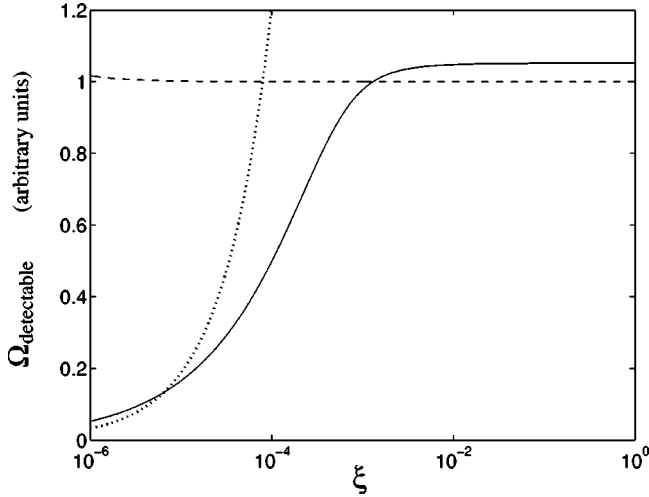


FIG. 2. The minimum gravitational-wave energy density  $\Omega_{\text{detectable}}$  necessary for detection as a function of the background's Gaussianity parameter  $\xi$  for  $N=10^9$  data points, which is a realistic number of data points for ground based detectors. The false alarm and false dismissal probabilities are both 0.01. The solid line is the theoretical prediction (C46) and (C55) for the maximum likelihood statistic, which is expected to be accurate to a few tens of percent. The dashed line is the theoretical prediction (4.6) for the cross correlation statistic, and the dotted line is the theoretical prediction (4.9), (4.10) for the burst statistic; see caption to Fig. 1. This plot indicates a maximum gain factor of  $\sim 2$  in energy density for duty cycles in a narrow band near  $\xi \sim 10^{-4}$ .

statistic  $\Lambda_{\text{ML}}^{\text{NG}}$ , the cross-correlation statistic  $\Lambda_{\text{CC}}$ , and the burst statistic  $\Lambda_{\text{B}}$ . That comparison is quantified in terms of the minimum gravitational-wave energy density  $\Omega_{\text{detectable}}$  necessary for detection. The values of this quantity for the three different statistics  $\Lambda_{\text{ML}}^{\text{NG}}$ ,  $\Lambda_{\text{CC}}$  and  $\Lambda_{\text{B}}$  we will denote by  $\Omega_{\text{detectable}}^{\text{ML}}$ ,  $\Omega_{\text{detectable}}^{\text{CC}}$ , and  $\Omega_{\text{detectable}}^{\text{B}}$ , respectively. Results for these three quantities obtained from Monte Carlo simulations are shown in Fig. 1, which gives  $\Omega_{\text{detectable}}$  as a function of  $\xi$  for  $N=10^4$  data points. The Monte Carlo simulations are described in Sec. IV B below. The figure shows that in the limit  $\xi \rightarrow 1$  of Gaussian signals, the statistics  $\Lambda_{\text{ML}}^{\text{NG}}$  and  $\Lambda_{\text{CC}}$  perform approximately equivalently (the cross-correlation statistic is slightly better). As the Gaussianity parameter  $\xi$  is decreased, the performance of  $\Lambda_{\text{ML}}^{\text{NG}}$  improves, until at  $\xi \sim 10^{-2.5}$  it is better than that of  $\Lambda_{\text{CC}}$  by about a factor of 3 in energy density. Finally, in the limit  $\xi \rightarrow 0$ , the individual bursts become visible and the burst statistic  $\Lambda_{\text{B}}$  becomes the best statistic.

Figure 1 also shows theoretical curves for the three quantities  $\Omega_{\text{detectable}}^{\text{ML}}$ ,  $\Omega_{\text{detectable}}^{\text{CC}}$ , and  $\Omega_{\text{detectable}}^{\text{B}}$ . These curves are derived and discussed in Sec. IV below. For the burst and cross-correlation statistics, the theoretical curves should have a fractional accuracy  $\sim 1/\sqrt{N}$ . For the maximum likelihood statistic, the theoretical prediction is expected to be accurate to a few tens of percent. These expected accuracies are confirmed by the Monte Carlo simulations, as seen in Fig. 1.

The value  $N=10^4$  of the number of data points is roughly appropriate for a space based detector like LISA, for which the duration of a measurement might be  $\sim 1$  year and the

effective bandwidth  $\sim 10^{-3}$  Hz. However, for year-long observations with ground based detectors, the effective bandwidth will be  $\sim 100$  Hz and consequently the appropriate value of  $N$  is  $\sim 10^9$ . We were unable to perform Monte Carlo simulations for this large value of  $N$  due to limitations in available computing power. However, we show in Fig. 2 the theoretical curves for the three different statistics as functions of  $\xi$  for  $N=10^9$ . In this case, the maximum likelihood statistic starts to outperform the cross-correlation statistic at  $\xi \sim 10^{-3}$ , and the maximum gain factor in energy density is of order  $\sim 2$ .

We next discuss the computational cost of the maximum likelihood statistic  $\Lambda_{\text{ML}}^{\text{NG}}$ . As is well known, the computational cost of trying to detect a stochastic background using the cross-correlation statistic  $\Lambda_{\text{CC}}$  is negligible when compared to, say, matched-filter-based inspiral waveform searches. However, because of the non-trivial maximization in Eq. (1.8), the maximum likelihood statistic  $\Lambda_{\text{ML}}^{\text{NG}}$  is computationally intensive. In fact, every evaluation of the function to be maximized over the four parameters  $\xi$ ,  $\alpha$ ,  $\sigma_1$ , and  $\sigma_2$  requires computing a length- $N$  sum or product, where  $N$  is the number of data points, and takes longer than the entire cross-correlation detection method. Depending on the method of calculation, the computational cost of computing  $\Lambda_{\text{ML}}^{\text{NG}}$  is larger than that of computing  $\Lambda_{\text{CC}}$  by a factor anywhere from  $10^2$  to  $10^4$ .

To summarize, under the idealized assumptions of this paper, if one searches for a stochastic background using the standard cross-correlation statistic, then one might not detect a signal that would have been detectable using our maximum likelihood statistic. This conclusion probably generalizes to realistic detector noise models and detector orientations.

### E. Outline of this paper

In Sec. II we introduce notation, review the general theory of signal detection and parameter measurement, and derive a general form of the maximum likelihood detection statistic. Then, in Sec. III, we derive the maximum likelihood statistics for both a Gaussian background (Sec. III B) and for the model (1.7) of a non-Gaussian background (Sec. III C), assuming two idealized detectors. In Sec. IV we discuss analytical calculations and Monte Carlo simulations comparing the performance of the maximum likelihood and cross-correlation detection statistics. Also in Sec. IV we show how the signal parameters  $\xi$  and  $\alpha$  can be estimated, with reasonable accuracy, for a strong non-Gaussian background. We conclude in Sec. V with a discussion of the results.

## II. GENERAL THEORY OF DETECTION STATISTICS AND PARAMETER ESTIMATION

In this section we review various formal aspects of the theory of signal detection and measurement. We derive a form of the maximum likelihood detection statistic that is more general than has been considered before in the context of gravitational wave data analysis [20,24–27]. The material in this section can be found in a variety of texts [28]; we

include this section for completeness and to introduce notation.

### A. Notational conventions

We use calligraphic letters  $\mathcal{A}, \mathcal{B}, \mathcal{C}, \dots$  to denote random variables. As described in Sec. I B, given  $D$  detectors we can assemble an  $N \times D$  detector output matrix  $\mathcal{H}$  with components  $\mathcal{H}_i^k$  where  $k=1,2,\dots,N$  is a time index, and  $i=1,2,\dots,D$  labels the detector. We assume that the detector outputs are made up of noise  $\mathcal{N}$  and signal  $\mathcal{S}$  with components  $\mathcal{N}_i^k$  and  $\mathcal{S}_i^k$  respectively, such that

$$\mathcal{H} = \mathcal{N} + \mathcal{S}. \quad (2.1)$$

Specific realizations of random variables will be denoted by lower case Roman symbols. For example,  $h = n + s$  is a specific realization of Eq. (2.1), where the components of  $h$  are  $h_i^k$ .

Probability densities for random variables will always be denoted by a lowercase  $p$  and will carry a subscript to indicate which random variable is being described. For example,  $p_{\mathcal{N}}(n)d^{ND}n$  is the probability that  $n < \mathcal{N} < n + dn$ , where  $d^{ND}n$  is the product

$$d^{ND}n = \prod_{k=1}^N \prod_{i=1}^D dn_i^k. \quad (2.2)$$

We write the normalization requirement for  $p_{\mathcal{N}}(n)$  as

$$1 = \int d^{ND}n p_{\mathcal{N}}(n). \quad (2.3)$$

Unless otherwise specified, integrals are over  $\mathbb{R}^{ND}$  where  $\mathbb{R}$  is the set of real numbers.

We assume a detector noise model with  $Q_n$  parameters. Let  $\mathcal{V}_n$  be a vector of length  $Q_n$  whose components are the parameters characterizing the noise in the detectors. We denote by  $\Theta_n$  the space of all possible values of  $\mathcal{V}_n$ . Here the subscript  $n$  is not an index; it is merely short for ‘‘noise.’’ We denote joint probabilities in the usual way. For example,  $p_{\mathcal{N}, \mathcal{V}_n}(n, \mathbf{v}_n)d^{ND}nd^{Q_n}v_n$  is the probability that  $n < \mathcal{N} < n + dn$  and  $\mathbf{v}_n < \mathcal{V}_n < \mathbf{v}_n + d\mathbf{v}_n$ , where  $d^{Q_n}v_n$  is defined by

$$d^{Q_n}v_n = \prod_{l=1}^{Q_n} dv_n^l, \quad (2.4)$$

and  $dv_n^l$  is the  $l$ th component of  $d\mathbf{v}_n$ . We also use vertical bars to denote conditional probabilities. For example

$$p_{\mathcal{N}|\mathcal{V}_n}(n|\mathbf{v}_n)d^{ND}n = \frac{p_{\mathcal{N}, \mathcal{V}_n}(n, \mathbf{v}_n)d^{Q_n}v_n}{p_{\mathcal{V}_n}(\mathbf{v}_n)d^{Q_n}v_n} d^{ND}n \quad (2.5)$$

is the probability that  $n < \mathcal{N} < n + dn$  given that  $\mathbf{v}_n < \mathcal{V}_n < \mathbf{v}_n + d\mathbf{v}_n$ .

We will often use the so-called total probability theorem [29] to write probability densities for a specific random variable as an integral over the functional dependencies of that random variable. An example is

$$p_{\mathcal{N}}(n) = \int_{\Theta_n} d^{Q_n}v_n p_{\mathcal{N}|\mathcal{V}_n}(n|\mathbf{v}_n) p_{\mathcal{V}_n}(\mathbf{v}_n). \quad (2.6)$$

Expanding probability densities in this way allows us to treat parameters, such as the noise parameters  $\mathcal{V}_n$  in Eq. (2.6), as unknowns. In fact, such a treatment of the noise parameters is the crucial difference between the derivations of this work and those in previous studies of gravitational wave data analysis techniques [20,24–27].

We assume that the signal model contains  $Q_s$  parameters, which we will treat as random variables like the noise parameters. We will denote by  $\mathcal{V}_s$  the random vector of length  $Q_s$  containing the signal parameters, and by  $\Theta_s$  the space of all possible values of  $\mathcal{V}_s$ .

We define the notions of ‘‘signal present’’ and ‘‘signal absent’’ in terms of a partition of the space  $\Theta_s$  of signal parameters into a disjoint union

$$\Theta_s = \Theta_{s0} \cup \Theta_{s1}, \quad (2.7)$$

where  $\Theta_{s0}$  corresponds to the signal being absent, and  $\Theta_{s1}$  the signal being present. We define the random variable  $\mathcal{T}$ , taking values  $\mathcal{T}=0$  or  $\mathcal{T}=1$ , according to

$$\mathcal{T} = \begin{cases} 1 & \text{if } \mathcal{V}_s \in \Theta_{s1}, \\ 0 & \text{if } \mathcal{V}_s \in \Theta_{s0}. \end{cases} \quad (2.8)$$

Thus  $\mathcal{T}=1$  corresponds to a signal being present, and  $\mathcal{T}=0$  to no signal being present. We define

$$p_{\mathcal{S}|\mathcal{V}_s, \mathcal{T}}(s|\mathbf{v}_s, 0) = \begin{cases} 0 & \text{if } \mathbf{v}_s \in \Theta_{s1} \\ \delta^{ND}(s) & \text{if } \mathbf{v}_s \in \Theta_{s0} \end{cases}, \quad (2.9)$$

where  $\delta^{ND}(s)$  is the  $N \times D$  dimensional Dirac delta function. We denote by  $p_{\mathcal{T}, \mathcal{H}}(t, h)d^{ND}h$  the probability that  $\mathcal{T}=t$  and that  $h < \mathcal{H} < h + dh$ , where  $t=0$  or 1. Similarly

$$p_{\mathcal{H}|\mathcal{T}}(h|t)d^{ND}h = \frac{p_{\mathcal{H}, \mathcal{T}}(h, t)}{P_{\mathcal{T}}(t)} d^{ND}h \quad (2.10)$$

is the probability that  $h < \mathcal{H} < h + dh$  given that  $\mathcal{T}=t$ .

We denote probabilities (as opposed to probability densities) with an uppercase  $P$ . For example  $P_{\mathcal{T}}(1)$  is the probability that a signal is present, and  $P_{\mathcal{T}}(0)$  is the probability that a signal is absent.

Before examining the detector outputs, we may have some idea, say from previous experiments, of the probability that a signal will be present. We denote this prior probability by  $P^{(0)}$ . We denote by  $P^{(1)}$  the posterior probability that the signal is present after examining  $\mathcal{H}$  in the context of all prior experiments etc. All posterior quantities have an implicit dependence on the detector outputs. To simplify the notation we will not explicitly show this dependence. For example, we write  $P^{(1)}$  rather than the more cumbersome  $P^{(1)}(\mathcal{H})$  for the posterior probability that a signal is present.

There are prior and posterior versions of all probability densities. When necessary we will append superscripts of (0) and (1) to distinguish priors and posteriors respectively. For example  $p_{\mathcal{V}_n}^{(1)}(\mathbf{v}_n) = p_{\mathcal{V}_n|\mathcal{H}}(\mathbf{v}_n|h)$  is the posterior probability

TABLE I. A summary of conventions introduced in Sec. II A.

Convention	Example
Random variables are denoted by upper case calligraphic letters	The detector output matrix is denoted by $\mathcal{H}$
Specific realizations of random variables are denoted by lower case Roman letters (see next convention)	A specific observation run may result in a specific detector output matrix $h$ or say $x$ . These results would be denoted $\mathcal{H}=h$ and $\mathcal{H}=x$ respectively
A lower case $p$ denotes a probability density function (PDF) Its subscript determines the quantities with which it is associated	The PDF for the detector output $\mathcal{H}$ as a function of $h$ , or say $x$ , is denoted by $p_{\mathcal{H}}(h)$ and $p_{\mathcal{H}}(x)$ respectively
A comma in a PDF subscript and argument indicates a joint PDF	The joint PDF for $\mathcal{N}$ and $\mathcal{V}_n$ as a function of $n$ and $\mathbf{v}_n$ respectively is denoted by $p_{\mathcal{N},\mathcal{V}_n}(n,\mathbf{v}_n)$
A vertical bar in a PDF subscript and argument indicates a conditional PDF	The conditional PDF for $\mathcal{N}$ and $\mathcal{V}_n$ as a function of $n$ and $\mathbf{v}_n$ respectively is denoted by $p_{\mathcal{N} \mathcal{V}_n}(n \mathbf{v}_n)$
An upper case $P$ denotes a probability	The probability that $\mathcal{T}=1$ is denoted by $P_{\mathcal{T}}(1)$
Prior and posterior quantities are denoted by superscripts of (0) and (1) respectively	The prior probability that a signal is present is denoted by $P^{(0)}$ , while the posterior probability that a signal is present, after an observation $\mathcal{H}=h$ , is denoted by $P^{(1)} = P_{\mathcal{T} \mathcal{H}}(1 h)$

density for  $\mathcal{V}_n$ . The posterior distribution for the noise can be expanded in terms of  $p_{\mathcal{V}_n}^{(1)}(\mathbf{v}_n)$  as

$$p_{\mathcal{N}}^{(1)}(n) = \int_{\Theta_n} d^{Q_n} v_n p_{\mathcal{N}|\mathcal{V}_n}(n|\mathbf{v}_n) p_{\mathcal{V}_n}^{(1)}(\mathbf{v}_n). \quad (2.11)$$

The conventions and symbols which have been introduced above are summarized in Tables I and II respectively.

### B. Detection statistics

To detect a signal one uses a detection statistic, say  $\Gamma = \Gamma(\mathcal{H})$ , that is some function of the detector outputs  $\mathcal{H}$ . A signal is said to have been detected when  $\Gamma$  exceeds some threshold value  $\Gamma_*$ .

TABLE II. A summary of symbols introduced in Sec. II A.

Symbol	Meaning
$\mathcal{H}, h$	detector output matrix
$\mathcal{N}, n$	noise contribution to detector output matrix
$\mathcal{S}, s$	signal contribution to detector output matrix
$N$	number of strain samples taken from one detector
$D$	number of detectors
$Q_n$	number of parameters in the model noise PDF
$Q_s$	number of parameters in the model signal PDF
$\mathcal{V}_n, \mathbf{v}_n$	the parameters of the model noise PDF
$\mathcal{V}_s, \mathbf{v}_s$	the parameters of the model signal PDF
$\Theta_n$	the space of all possible values of $\mathcal{V}_n$
$\Theta_s$	the space of all possible values of $\mathcal{V}_s$
$\Theta_{s0}$	the subspace of $\Theta_s$ for which a signal is absent
$\Theta_{s1}$	the subspace of $\Theta_s$ for which a signal is present
$\mathcal{T}, t$	1 if a signal is present ( $\mathcal{V}_s \in \Theta_{s1}$ ), otherwise 0
$P^{(0)}$	prior probability that a signal is present
$P^{(1)}$	posterior probability that a signal is present

Denote by  $P_{\text{FD}}(\Gamma_*)$  the probability of false dismissal, that is, the probability that we fail to detect a signal which is actually present. Similarly, let  $P_{\text{FA}}(\Gamma_*)$  be the probability that we claim to have detected a signal which in fact is absent—the probability of false alarm. For given signal and noise models and for a given statistic  $\Gamma$ , the false alarm and false dismissal probabilities generate a curve in the  $P_{\text{FA}}-P_{\text{FD}}$  plane parametrized by the threshold  $\Gamma_*$ . Such curves depend on the number of detectors  $D$ , the number of data points  $N$ , the signal parameters  $\mathcal{V}_s$ , and the noise parameters  $\mathcal{V}_n$ .

Suppose that the statistic  $\Gamma$  is bounded in the sense that there exist numbers  $\Gamma_{\min}$  and  $\Gamma_{\max}$  such that  $\Gamma_{\min} < \Gamma < \Gamma_{\max}$  for all  $\mathcal{H}$ . Then it is clear that  $P_{\text{FD}}(\Gamma_{\min})=0$  and that  $P_{\text{FA}}(\Gamma_{\min})=1$ . As the threshold  $\Gamma_*$  increases toward  $\Gamma_{\max}$ ,  $P_{\text{FD}}(\Gamma_*)$  will increase while  $P_{\text{FA}}(\Gamma_*)$  decreases, until finally at  $\Gamma_* = \Gamma_{\max}$ ,  $P_{\text{FD}}=1$ , and  $P_{\text{FA}}=0$ . Thus, false dismissal-false alarm curves generally look something like those sketched in Fig. 3.

Note that if one uses a different statistic  $f(\Gamma)$ , where  $f$  is any function, then the shape of the  $P_{\text{FA}}-P_{\text{FD}}$  curve does not change as long as  $f$  is monotonic in the sense that

$$\Gamma > \Gamma_* \Rightarrow f(\Gamma) > f(\Gamma_*). \quad (2.12)$$

Only the parametrization of the curve changes under such a transformation. Statistics related by transformations  $f$  satisfying the monotonicity property (2.12) have identical false dismissal versus false alarm curves.

In 1933 Neyman and Pearson considered a simple signal detection scenario where the sets  $\Theta_n$ ,  $\Theta_{s1}$ , and  $\Theta_{s0}$  each contain a single element [30]. They showed that for this scenario the detection statistic which minimizes  $P_{\text{FD}}$  for any  $P_{\text{FA}}$  is the so-called *likelihood ratio*  $\Lambda$ , defined by

$$\Lambda = \frac{p_{\mathcal{T}|\mathcal{H}}(h|1)}{p_{\mathcal{T}|\mathcal{H}}(h|0)}. \quad (2.13)$$

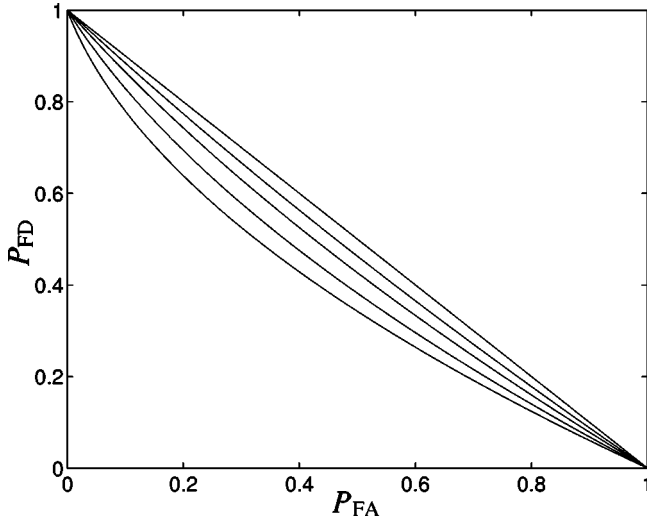


FIG. 3. False dismissal versus false alarm curves for typical detection statistics.

One notion of optimality for detection statistics is that the statistic should minimize the false dismissal probability at a fixed value of the false alarm probability. For the simple scenario above, this criteria, known as the Neyman-Pearson criteria, uniquely determines the likelihood ratio as the optimal statistic [31]. However in general, when any of  $\Theta_n$ ,  $\Theta_{s1}$ , or  $\Theta_{s0}$  contains more than one element, the statistic

selected by this criteria is a function of the unknown parameters  $\mathcal{V}_s$  and  $\mathcal{V}_n$ . Thus, as is well known, the Neyman-Pearson criteria does not single out a unique statistic in such cases.

In this paper we will obtain our detection statistics from Bayesian considerations, but we will quantify their effectiveness using the Neyman and Pearson criteria of comparing false dismissal probabilities at fixed false alarm probabilities.

### C. Likelihood ratio and likelihood function

From a Bayesian point of view, a natural criterion for deciding that a signal is present is for the posterior probability  $P^{(1)}$  to exceed some threshold [32]. The posterior probability  $P^{(1)}$  is related to the prior probability  $P^{(0)}$  and to the likelihood ratio  $\Lambda$  defined by Eq. (2.13) by

$$\frac{P^{(1)}}{1 - P^{(1)}} = \Lambda \frac{P^{(0)}}{1 - P^{(0)}}. \quad (2.14)$$

See Appendix A for a derivation of Eq. (2.14) in the most general context where the sets  $\Theta_n$ ,  $\Theta_{s1}$ , and  $\Theta_{s0}$  are all non-trivial. It follows from Eq. (2.14) that  $P^{(1)}$  is a monotonic function of  $\Lambda$ , so thresholding on  $P^{(1)}$  is equivalent to thresholding on  $\Lambda$ . This makes  $\Lambda$ , or approximate versions of it, the natural choice for a detection statistic.

We derive in Appendix A the following general formula for the likelihood ratio as a function of the data  $\mathcal{H} = h$ :

$$\Lambda = \frac{\int_{\Theta_{s1}} d^{\mathcal{Q}_s} v_s \int d^{ND} s \int_{\Theta_n} d^{\mathcal{Q}_n} v_n p_{\mathcal{M}|\mathcal{V}_n}(h - s | \mathbf{v}_n) p_{\mathcal{V}_n}(\mathbf{v}_n) p_{\mathcal{S}|\mathcal{V}_s, \mathcal{T}}(s | \mathbf{v}_s, 1) p_{\mathcal{V}_s|\mathcal{T}}(\mathbf{v}_s | 1)}{\int_{\Theta_n} d^{\mathcal{Q}_n} v'_n p_{\mathcal{M}|\mathcal{V}'_n}(h | \mathbf{v}'_n) p_{\mathcal{V}'_n}(\mathbf{v}'_n)}. \quad (2.15)$$

The various probability distributions that appear in Eq. (2.15) are (i) the prior distribution  $p_{\mathcal{V}_s|\mathcal{T}}(\mathbf{v}_s | 1)$  for the signal parameters  $\mathbf{v}_s$ ; (ii) the distribution  $p_{\mathcal{S}|\mathcal{V}_s, \mathcal{T}}(s | \mathbf{v}_s, 1)$  for the signal  $s$  given the signal parameters  $\mathbf{v}_s$ ; (iii) the prior distribution  $p_{\mathcal{V}_n}(\mathbf{v}_n)$  for the noise parameters  $\mathbf{v}_n$ ; and (iv) the distribution  $p_{\mathcal{M}|\mathcal{V}_n}(h | \mathbf{v}_n)$  for the noise  $n$  given the noise parameters  $\mathbf{v}_n$ .

We can interpret Eq. (2.15) as follows. In the simple signal detection scenario, we choose between a pair of simple claims: (i)  $\mathcal{V}_s = \mathbf{v}_{s0}$  or (ii)  $\mathcal{V}_s = \mathbf{v}_{s1}$ . In general we choose between a pair of complicated, or composite, claims: (i)  $\mathcal{V}_s \in \Theta_{s0}$  or (ii)  $\mathcal{V}_s \in \Theta_{s1}$ , where both  $\Theta_{s0}$  and  $\Theta_{s1}$  contain many elements. Equation (2.15) says that the best way to choose between a pair of complicated claims is to first break the complicated pair of claims into pairs of simple claims, then compute the likelihood ratio for each pair of simple claims, and sum the results of each choice. That is, the likelihood ratio can be written as an integral over the parameters of the composite claims

$$\Lambda = \int_{\Theta_{s1}} d^{\mathcal{Q}_s} v_s \int_{\Theta_n} d^{\mathcal{Q}_n} v_n \Lambda(\mathbf{v}_s, \mathbf{v}_n), \quad (2.16)$$

where the integrand  $\Lambda(\mathbf{v}_s, \mathbf{v}_n)$ , which we refer to as the *likelihood function*, can be read off from Eq. (2.15):

$$\Lambda(\mathbf{v}_s, \mathbf{v}_n) = \frac{\int d^{ND} s p_{\mathcal{M}|\mathcal{V}_n}(h - s | \mathbf{v}_n) p_{\mathcal{S}|\mathcal{V}_s, \mathcal{T}}(s | \mathbf{v}_s, 1) p_{\mathcal{V}_n}(\mathbf{v}_n) p_{\mathcal{V}_s|\mathcal{T}}(\mathbf{v}_s | 1)}{\int_{\Theta_n} d^{\mathcal{Q}_n} v'_n p_{\mathcal{M}|\mathcal{V}'_n}(h | \mathbf{v}'_n) p_{\mathcal{V}'_n}(\mathbf{v}'_n)}. \quad (2.17)$$

The likelihood function<sup>2</sup>  $\Lambda(\mathbf{v}_s, \mathbf{v}_n)$  can be used to compute the posterior probability density  $p_{\mathcal{V}_s, \mathcal{V}_n | \mathcal{T}}^{(1)}(\mathbf{v}_s, \mathbf{v}_n | 1)$  for the signal and noise parameters given that a signal is present, via the formula

$$\frac{P^{(1)}}{1 - P^{(1)}} p_{\mathcal{V}_s, \mathcal{V}_n | \mathcal{T}}^{(1)}(\mathbf{v}_s, \mathbf{v}_n | 1) = \Lambda(\mathbf{v}_s, \mathbf{v}_n) \frac{P^{(0)}}{1 - P^{(0)}}. \quad (2.18)$$

A derivation of Eq. (2.18) can be found in Appendix A.

#### D. Maximum likelihood detection statistics and parameter estimators

In many applications, it is impractical to compute the detection statistic (2.15) because of the multi-dimensional integrals involved [33]. However, approximate versions of the statistic are often easier to compute and useful. If a signal is present with sufficiently large amplitude, then the integrand in the numerator of Eq. (2.15) will be sharply peaked. The integrand in the denominator of Eq. (2.15) will also be sharply peaked when there is sufficient data that the noise is well characterized. Under these circumstances, the integrals can be written as the values of the corresponding integrands at the peaks multiplied by “width factors,” where the width factors depend only weakly on the data  $h$  and can be neglected without affecting much the performance of the statistic. (The width factors from the integrals over the noise parameters will tend to cancel between the numerator and denominator.) Also, frequently the prior distributions for  $\mathcal{V}_s$  and  $\mathcal{V}_n$  are slowly varying, and neglecting those distributions has a negligible effect on the performance of the statistic. Under these conditions the maximum likelihood detection statistic  $\Lambda_{\text{ML}}$  defined by

$$\Lambda_{\text{ML}} = \frac{\max_{\mathbf{v}_s \in \Theta_{s1}} \max_{\mathbf{v}_n \in \Theta_n} \int d^{ND} s p_{\mathcal{M}|\mathcal{V}_n}(h-s|\mathbf{v}_n) p_{\mathcal{S}|\mathcal{V}_s, \mathcal{T}}(s|\mathbf{v}_s, 1)}{\max_{\mathbf{v}' \in \Theta_n} p_{\mathcal{M}|\mathcal{V}_n}(h|\mathbf{v}')} \quad (2.19)$$

is a natural approximate version of  $\Lambda$ .<sup>3</sup> The subscript ML denotes that Eq. (2.19) is the maximum likelihood approxi-

<sup>2</sup>There are two different conventions for the definition of the likelihood function. Some authors include the probability distributions for  $\mathcal{V}_s$  and  $\mathcal{V}_n$  in the definition of  $\Lambda(\mathbf{v}_s, \mathbf{v}_n)$  as we have in Eq. (2.17), while others leave these out of  $\Lambda(\mathbf{v}_s, \mathbf{v}_n)$  and would show these distributions explicitly in Eq. (2.16).

<sup>3</sup>In the event that the priors for  $\mathcal{V}_s$  and  $\mathcal{V}_n$  restrict these parameters to regions  $\Theta'_{s1} \subset \Theta_{s1}$  and  $\Theta'_n \subset \Theta_n$ , the bounds of the maximizations in Eq. (2.19) should be changed to  $\Theta_{s1} \rightarrow \Theta'_{s1}$  and  $\Theta_n \rightarrow \Theta'_n$ .

mate version of  $\Lambda$ . See Ref. [28] for further discussion of  $\Lambda_{\text{ML}}$  as an approximate version of  $\Lambda$ .<sup>4</sup>

A particular special case of the detection statistic (2.19), which is widely used, is the following. Assume that the noise parameters have some known values  $\mathcal{V}_n = \mathbf{v}_n$ . Then the noise priors and the  $\Theta_n$  integrals in Eq. (2.15) are trivial, and one obtains the detection statistic

$$\tilde{\Lambda}_{\text{ML}} = \frac{\max_{\mathbf{v}_s \in \Theta_{s1}} \int d^{ND} s p_{\mathcal{M}|\mathcal{V}_n}(h-s|\mathbf{v}_n) p_{\mathcal{S}|\mathcal{V}_s, \mathcal{T}}(s|\mathbf{v}_s, 1)}{p_{\mathcal{M}|\mathcal{V}_n}(h|\mathbf{v}_n)}. \quad (2.20)$$

See Ref. [26] for an exploration of the statistic (2.20) in the context of stochastic backgrounds. We will show below that for a Gaussian stochastic background,  $\Lambda_{\text{ML}}$  reduces to the standard cross-correlation statistic while the more specialized statistic  $\tilde{\Lambda}_{\text{ML}}$  does not. Thus for stochastic backgrounds, treating the noise parameters as unknowns is crucial [22].

When the noise and signal parameters  $\mathcal{V}_n$  and  $\mathcal{V}_s$  can take on many values, one naturally would like to know which values are realized. Equation (2.18) suggests using the values  $\hat{\mathbf{v}}_n$  and  $\hat{\mathbf{v}}_s$  defined by

$$\Lambda(\hat{\mathbf{v}}_s, \hat{\mathbf{v}}_n) = \max_{\mathbf{v}_s \in \Theta_{s1}} \max_{\mathbf{v}_n \in \Theta_n} \Lambda(\mathbf{v}_s, \mathbf{v}_n). \quad (2.21)$$

The estimators  $\hat{\mathbf{v}}_n$  and  $\hat{\mathbf{v}}_s$  are known as maximum likelihood estimators. Note that  $\mathbf{v}_s = \hat{\mathbf{v}}_s$  and  $\mathbf{v}_n = \hat{\mathbf{v}}_n$  also maximize the numerator in Eq. (2.19). For the remainder of this paper we will use  $\Lambda_{\text{ML}}$ , defined by Eq. (2.19), as our detection statistic, and  $\hat{\mathbf{v}}_s$  and  $\hat{\mathbf{v}}_n$ , defined by Eq. (2.21), as parameter estimators.

### III. APPLICATION TO STOCHASTIC BACKGROUND SEARCHES

In this section we derive the maximum likelihood detection statistic (2.19) for a simplified model of the detection problem for stochastic gravitational waves, and for a specific simple model of a non-Gaussian stochastic background.

#### A. Assumptions

We assume two detectors with outputs  $\mathcal{H}_i^k$ , where  $i = 1, 2$  labels the detector and  $k = 1, 2, \dots, N$  is a time index. We assume that the noise in detector one is uncorrelated with the noise in detector two. We will require the noise in both detectors to have vanishing mean and to be both Gaussian and white, so that

<sup>4</sup>Note that  $\Lambda_{\text{ML}}$  is an approximate version of  $\Lambda$  only in the sense that the false dismissal versus false alarm curves of the two statistics will be close to one another. The numerical values of  $\Lambda_{\text{ML}}$  and  $\Lambda$  will in general differ significantly, due to the width factors and priors. Therefore the statistic  $\Lambda_{\text{ML}}$  cannot be used in Eq. (2.14) to compute Bayesian thresholds for detection given a desired value of  $P^{(1)}$ .



$$p_{\mathcal{N}|\mathcal{V}_n}[n|(\sigma_1, \sigma_2)] = \prod_{k=1}^N \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{(n_1^k)^2}{2\sigma_1^2} - \frac{(n_2^k)^2}{2\sigma_2^2}\right]. \quad (3.1)$$

The parameters  $\sigma_1$  and  $\sigma_2$  in Eq. (3.1) are the square roots of the variances of the noise in the two detectors. For this model  $\mathbf{v}_n = (\sigma_1, \sigma_2)$  and  $\Theta_n = \{(\sigma_1, \sigma_2) \mid \sigma_1 \geq 0 \text{ and } \sigma_2 \geq 0\}$ .

We assume that the detectors are collocated and aligned, so that the same signal is present in both detectors

$$S_1^k = S_2^k = S^k. \quad (3.2)$$

Lastly we assume that the individual signal samples are uncorrelated and identically distributed, i.e., the signal is white, so that

$$p_S(s) = \prod_{k=1}^N p_{S^k}(s^k). \quad (3.3)$$

Our assumptions (3.1)–(3.3) are unrealistic for both ground-based and space-based detectors: we expect the noise to be colored with significant non-Gaussian components, and in general detectors will not be collocated and aligned. Our

analysis is therefore just a first step, and will need to be generalized. However, we expect that our central conclusion—the existence of statistics which outperform the standard cross-correlation statistic for non-Gaussian signals—is robust, and will not be altered when these complications are taken into account.

We now derive a general formula for the maximum likelihood statistic (2.19), which we apply in both the Gaussian and non-Gaussian cases in the following two subsections. The denominator in Eq. (2.19) can be written, from Eq. (3.1), as

$$\max_{\sigma_1 \geq 0} \max_{\sigma_2 \geq 0} \left\{ (2\pi\sigma_1\sigma_2)^{-N} \exp\left[-\frac{N}{2} \left(\frac{\bar{\sigma}_1^2}{\sigma_1^2} + \frac{\bar{\sigma}_2^2}{\sigma_2^2}\right)\right] \right\}, \quad (3.4)$$

where  $\bar{\sigma}_1^2$  and  $\bar{\sigma}_2^2$  are defined by

$$\bar{\sigma}_i^2 = \frac{1}{N} \sum_{k=1}^N (h_i^k)^2 \quad (3.5)$$

for  $i=1,2$ . It is easily shown that the maximum in Eq. (3.4) is achieved at  $\sigma_i = \bar{\sigma}_i$ . From Eq. (2.19) this yields

$$\Lambda_{\text{ML}} = \frac{\max_{\mathbf{v}_s \in \Theta_{s1}} \max_{\mathbf{v}_n \in \Theta_n} \int d^{ND} s p_{\mathcal{N}|\mathcal{V}_n}(h-s|\mathbf{v}_n) p_{S|\mathcal{V}_s, \mathcal{T}}(s|\mathbf{v}_s, 1)}{(2\pi\bar{\sigma}_1\bar{\sigma}_2)^{-N} \exp(-N)}. \quad (3.6)$$

Combining this with Eq. (3.3) yields the following final general expression for the maximum likelihood statistic:

$$\Lambda_{\text{ML}} = \max_{\mathbf{v}_s \in \Theta_{s1}} \max_{\sigma_1 \geq 0} \max_{\sigma_2 \geq 0} \prod_{k=1}^N \frac{\bar{\sigma}_1 \bar{\sigma}_2}{\sigma_1 \sigma_2} \int_{-\infty}^{\infty} ds^k p_{S^k|\mathcal{V}_s, \mathcal{T}}(s^k|\mathbf{v}_s, 1) \exp\left[-\frac{(h_1^k - s^k)^2}{2\sigma_1^2} - \frac{(h_2^k - s^k)^2}{2\sigma_2^2} + 1\right]. \quad (3.7)$$

### B. Gaussian signal

We now consider the case where the signal is Gaussian and has a vanishing mean. We denote by  $\alpha^2$  the variance of the signal, so the prior for  $S$  is given by

$$p_{S^k|\mathcal{V}_s, \mathcal{T}}(s^k|\alpha, 1) = \frac{1}{\sqrt{2\pi\alpha}} \exp\left[-\frac{(s^k)^2}{2\alpha^2}\right]. \quad (3.8)$$

For this model  $\mathbf{v}_s = (\alpha)$  has only one component, and  $\Theta_{s1} = \{\alpha \mid \alpha > 0\}$ .

Substituting the signal probability distribution (3.8) into the general expression (3.7) for  $\Lambda_{\text{ML}}$  yields a Gaussian integral which is straightforward to evaluate. The result is

$$\Lambda_{\text{ML}}^{\text{G}} = \max_{\alpha > 0} \max_{\sigma_1 \geq 0} \max_{\sigma_2 \geq 0} \left\{ \frac{\bar{\sigma}_1 \bar{\sigma}_2}{\sqrt{\sigma_1^2 \sigma_2^2 + \sigma_1^2 \alpha^2 + \sigma_2^2 \alpha^2}} \times \exp\left[ \frac{\frac{\bar{\sigma}_1^2}{\sigma_1^4} + \frac{\bar{\sigma}_2^2}{\sigma_2^4} + \frac{2\bar{\alpha}^2}{\sigma_1^2 \sigma_2^2}}{2\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\alpha^2}\right)} - \frac{\bar{\sigma}_1^2}{2\sigma_1^2} - \frac{\bar{\sigma}_2^2}{2\sigma_2^2} + 1 \right] \right\}^N, \quad (3.9)$$

where

$$\bar{\alpha}^2 = \frac{1}{N} \sum_{k=1}^N h_1^k h_2^k, \quad (3.10)$$

and we have appended a superscript G on  $\Lambda_{\text{ML}}^{\text{G}}$  to indicate the maximum likelihood detection statistic for a Gaussian signal.

One can show that the maximum in Eq. (3.9) is achieved at  $\alpha = \hat{\alpha}$ ,  $\sigma_1 = \hat{\sigma}_1$ , and  $\sigma_2 = \hat{\sigma}_2$ , where

$$\hat{\alpha}^2 = \bar{\alpha}^2 \theta(\bar{\alpha}^2), \quad (3.11)$$

$$\hat{\sigma}_i^2 = (\bar{\sigma}_i^2 - \hat{\alpha}^2) \theta(\bar{\sigma}_i^2 - \hat{\alpha}^2), \quad (3.12)$$

for  $i=1,2$ , and  $\bar{\sigma}_1$  and  $\bar{\sigma}_2$  are given by Eq. (3.5). Here  $\theta(x)$  is the step function (1.5). The quantities (3.11) and (3.12) are the maximum likelihood estimators for the variance  $\alpha^2$  of the signal and the variances  $\sigma_1^2$  and  $\sigma_2^2$  of the noise in the two detectors. The step functions in Eqs. (3.11) and (3.12) arise as a result of the bounds of the maximization in Eq. (3.7).

The corresponding detection statistic is, from Eq. (3.9),<sup>5</sup>

$$\Lambda_{\text{ML}}^{\text{G}} = \left[ 1 - \frac{\hat{\alpha}^4}{\bar{\sigma}_1^2 \bar{\sigma}_2^2} \right]^{-N/2}. \quad (3.13)$$

The cross-correlation statistic  $\Lambda_{\text{CC}}$  can be obtained from  $\Lambda_{\text{ML}}^{\text{G}}$  via a monotonic transformation which preserves false dismissal versus false alarm curves [cf. Eq. (2.12) above]:

$$\Lambda_{\text{CC}} = \sqrt{1 - (\Lambda_{\text{ML}}^{\text{G}})^{-2/N}} = \frac{\hat{\alpha}^2}{\bar{\sigma}_1 \bar{\sigma}_2}. \quad (3.14)$$

Note that if we had assumed the noise parameters  $\mathbf{v}_n = (\sigma_1, \sigma_2)$  were known, and derived a statistic from Eq. (2.20) rather than Eq. (2.19), we would have found instead the detection statistic  $\bar{\Lambda}_{\text{ML}}^{\text{G}} = \bar{\Lambda}_{\text{ML}}^{\text{G}} \theta(\bar{\Lambda}_{\text{ML}}^{\text{G}})$ , where

$$\bar{\Lambda}_{\text{ML}}^{\text{G}} = \bar{\alpha}^2 + \frac{1}{2} \left[ \frac{\sigma_2^2}{\sigma_1^2} (\bar{\sigma}_1^2 - \sigma_1^2) + \frac{\sigma_1^2}{\sigma_2^2} (\bar{\sigma}_2^2 - \sigma_2^2) \right], \quad (3.15)$$

which is different from the standard cross-correlation statistic. This non-standard result is obtained because of the unrealistic assumption that the noise parameters  $\mathbf{v}_n = (\sigma_1, \sigma_2)$  are known. Different derivations of the result (3.15) can be found in Refs. [22,26].

It is often useful to characterize the “strength” of a stochastic background in terms of the signal-to-noise ratio of the cross-correlation statistic (3.14), which we now define. First note that for large  $N$ , the fractional fluctuations in  $\hat{\alpha}^2$

will be much larger than those in  $\bar{\sigma}_1 \bar{\sigma}_2$ .<sup>6</sup> For the purpose of defining the signal-to-noise ratio, we assume that  $N$  is large enough that  $\bar{\sigma}_1$  and  $\bar{\sigma}_2$  in Eq. (3.14) can be taken to be independent of  $h$ , so that  $\Lambda_{\text{CC}}$  and  $\hat{\alpha}^2$  are equivalent detection statistics. We also use  $\bar{\alpha}^2$  instead of  $\hat{\alpha}^2$  in the computations that follow, as is conventional when defining signal-to-noise ratios. If a signal is present, then the expected value of  $\bar{\alpha}^2$  is, from Eqs. (2.1), (3.1)–(3.3), (3.8) and (3.10),

$$\langle \bar{\alpha}^2 \rangle = \alpha^2. \quad (3.16)$$

If no signal is present, so that  $\alpha^2 = 0$ , then the fluctuations in  $\bar{\alpha}^2$  are given by

$$\Delta(\bar{\alpha}^2) = \frac{\sigma_1 \sigma_2}{\sqrt{N}}. \quad (3.17)$$

The signal-to-noise ratio  $\rho$  is defined to be the ratio of these two quantities:

$$\rho = \frac{\alpha^2 \sqrt{N}}{\sigma_1 \sigma_2}. \quad (3.18)$$

### C. Non-Gaussian signal

As mentioned in the Introduction, the traditional assumption that a gravitational wave stochastic background will be Gaussian requires the individual events to be sufficiently frequent and uncorrelated. Our model for a non-Gaussian signal assumes instead that the events are infrequent.

Consider a collection of similar events generating a stochastic background  $\mathcal{S}$ . Let  $\xi$  be the probability that, at any randomly chosen time, the waves from an event are arriving at the detectors. We assume that the time structure of individual events cannot be resolved by the detectors. That is, we assume that the events occur over time scales smaller than the detectors’ resolution time, as illustrated in Fig. 4. We assume that the distribution of the amplitudes of the events is Gaussian with variance  $\alpha^2$ . The probability distribution for the signal given the signal parameters  $(\xi, \alpha)$  is therefore given by

$$p_{\mathcal{S}^k | \gamma_s} \mathcal{P}^{\mathcal{S}^k}[(\xi, \alpha), 1] = \frac{\xi}{\sqrt{2\pi\alpha}} \exp\left[-\frac{(s^k)^2}{2\alpha^2}\right] + (1 - \xi) \delta(s^k), \quad (3.19)$$

together with Eq. (3.3). Thus the signal model parameters are  $\mathbf{v}_s = (\xi, \alpha)$ , which give respectively the “event probability” and “event variance” characterizing the stochastic background. The parameter space  $\Theta_s$  for this model is

$$\Theta_s = \{(\xi, \alpha) \mid 0 \leq \xi \leq 1 \text{ and } \alpha \geq 0\}, \quad (3.20)$$

and the subset corresponding to a signal being present is

<sup>5</sup>To simplify the formula for  $\Lambda_{\text{ML}}^{\text{G}}$  we assume that  $\bar{\sigma}_i^2 - \alpha^2 > 0$ . This will be true for any realistic value of  $N$  since  $\bar{\sigma}_i^2 - \alpha^2 = \sigma_{i,\text{true}}^2 + \mathcal{O}(1/\sqrt{N})$ , where  $\sigma_{i,\text{true}}$  is the true value of  $\sigma_i$  and the second term describes the statistical fluctuations.

<sup>6</sup>This is true at fixed signal-to-noise ratio  $\rho$ .

$$\Theta_{s_1} = \{(\xi, \alpha) \mid 0 < \xi \leq 1 \text{ and } \alpha > 0\}. \quad (3.21)$$

Note that our assumption that the time structure of events is not resolved by the detector is unrealistic. Detector resolution times can be as small as 0.1 ms in the case of ground-based detectors like LIGO,<sup>7</sup> and even supernova bursts are

expected to have time scales  $\geq 10$  ms [34,35]. It will be important for future studies to relax this assumption.

We now compute the maximum likelihood detection statistic  $\Lambda_{\text{ML}}^{\text{NG}}$  for our simple non-Gaussian signal model by substituting Eq. (3.19) into Eq. (3.7). This yields

$$\Lambda_{\text{ML}}^{\text{NG}} = \max_{0 < \xi \leq 1} \max_{\alpha > 0} \max_{\sigma_1 \geq 0} \max_{\sigma_2 \geq 0} \prod_{k=1}^N \left\{ \frac{\bar{\sigma}_1 \bar{\sigma}_2 \xi}{\sqrt{\sigma_1^2 \sigma_2^2 + \sigma_1^2 \alpha^2 + \sigma_2^2 \alpha^2}} \exp \left[ \frac{\left( \frac{h_1^k}{\sigma_1^2} + \frac{h_2^k}{\sigma_2^2} \right)^2}{2 \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\alpha^2} \right)} - \frac{(h_1^k)^2}{2\sigma_1^2} - \frac{(h_2^k)^2}{2\sigma_2^2} + 1 \right] \right. \\ \left. + \frac{\bar{\sigma}_1 \bar{\sigma}_2}{\sigma_1 \sigma_2} (1 - \xi) \exp \left[ -\frac{(h_1^k)^2}{2\sigma_1^2} - \frac{(h_2^k)^2}{2\sigma_2^2} + 1 \right] \right\}. \quad (3.22)$$

The values of  $\xi$ ,  $\alpha^2$ ,  $\sigma_1^2$ , and  $\sigma_2^2$  which achieve the maximum in Eq. (3.22) are, respectively, estimators of the signal's Gaussianity parameter, the variance of the signal events, and the noise variances in the two detectors.<sup>8</sup> Note that if we evaluate Eq. (3.22) at  $\xi = 1$ , rather than maximizing over  $\xi$ , we recover Eq. (3.9) and the statistic  $\Lambda_{\text{ML}}^{\text{G}}$ .

We mention in passing an approximate version of the statistic (3.22) which is significantly easier to compute. Expanding the logarithm of the quantity to be maximized in Eq. (3.22) as a power series in  $\alpha^2$  to fourth order about  $\alpha^2 = 0$  yields the approximate statistic  $\hat{\Lambda}_{\text{ML}}^{\text{NG}}$  given by

$$\ln \hat{\Lambda}_{\text{ML}}^{\text{NG}} = \max_{0 < \xi \leq 1} \max_{\alpha > 0} \max_{\sigma_1 \geq 0} \max_{\sigma_2 \geq 0} \sum_{n=0}^4 \sum_{l=0}^8 \sum_{m=0}^8 \\ \times \left( \frac{\alpha^2}{\sigma_1 \sigma_2} \right)^n C_{nlm}(\xi, \sigma_1, \sigma_2) \sum_{k=1}^N (h_1^k)^l (h_2^k)^m, \quad (3.23)$$

where the coefficients  $C_{nlm}(\xi, \sigma_1^2, \sigma_2^2)$  vanish unless  $l + m$  is even and  $l + m \leq 8$ . In evaluating the statistic (3.23), one can first evaluate the 24 sums

$$\sum_{k=1}^N (h_1^k)^l (h_2^k)^m \quad (3.24)$$

for the required values of  $l$  and  $m$ , and subsequently numerically maximize over the parameters  $\xi$ ,  $\alpha$ ,  $\sigma_1$ , and  $\sigma_2$ . Thus the length- $N$  sums need only be performed once, rather than each time one tries a new set of values for  $\xi$ ,  $\alpha$ ,  $\sigma_1$ , and  $\sigma_2$ . Therefore the computational cost of  $\hat{\Lambda}_{\text{ML}}^{\text{NG}}$  is only about an order of magnitude greater than that of the cross correlation statistic  $\Lambda_{\text{CC}}$ , and this statistic may be useful to explore.

We now derive the signal-to-noise ratio  $\rho$  for the cross-correlation statistic and for the non-Gaussian signal (3.19). If the signal is present, then from Eqs. (2.1), (3.3), (3.10), (3.11) and (3.19) the expected value of  $\bar{\alpha}^2$  is

$$\langle \bar{\alpha}^2 \rangle = \xi \alpha^2. \quad (3.25)$$

If no signal is present then the fluctuations in  $\bar{\alpha}^2$  are given by

$$\Delta(\bar{\alpha}^2) = \frac{\sigma_1 \sigma_2}{\sqrt{N}}. \quad (3.26)$$

Therefore, taking the ratio of Eqs. (3.25) and (3.26), the signal-to-noise ratio  $\rho$  is

$$\rho = \frac{\xi \alpha^2 \sqrt{N}}{\sigma_1 \sigma_2}. \quad (3.27)$$

<sup>7</sup>For ground-based detectors, the effective resolution time in a cross-correlation between two detectors can be considerably longer than 0.1 ms [20], which may help with this issue.

<sup>8</sup>See Ref. [36] for a derivation of a statistic similar to  $\Lambda_{\text{ML}}^{\text{NG}}$  and designed for the same non-Gaussian signals which is based on Eq. (2.20) rather than Eq. (2.19).

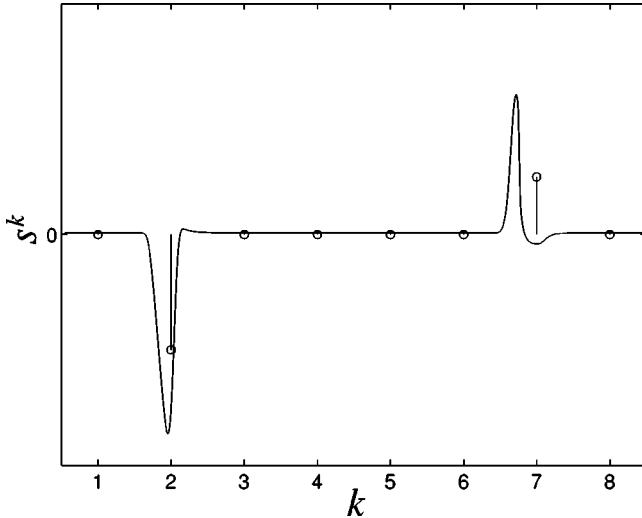


FIG. 4. Sketched segment of the signal produced by a model non-Gaussian stochastic background of events unresolved by the detectors. Here we show two events. The solid curve is the exact signal. This exact signal's contributions to the detector outputs, shown as stemmed O's, are averages of the exact signal over the detector resolution time scale.

#### IV. PERFORMANCE COMPARISON

In this section we compare the performances of the cross-correlation statistic (3.14), the burst statistic (1.9), and the maximum likelihood statistic (3.22) for our model non-Gaussian signal described in Sec. III C. The comparison is quantified in terms of the false alarm versus false dismissal curves, as discussed in Sec. II above. In Sec. IV A we discuss analytic predictions for these curves for the three different statistics. Section IV B describes our Monte Carlo simulation algorithm, and Secs. IV C and IV D describe the results.

##### A. Analytic computation of asymptotic behavior of statistics

We start by discussing the set of parameters on which the false dismissal versus false alarm curves can depend. As before, we assume two detectors with noise characterized by Eq. (3.1) with  $\mathcal{V}_n = (\sigma_1, \sigma_2)$ , and a non-Gaussian signal characterized by Eqs. (3.3) and (3.19) with  $\mathcal{V}_s = (\xi, \alpha)$ . The curves for each statistic are given by some function

$$P_{\text{FD}} = P_{\text{FD}}(P_{\text{FA}}, \xi, \alpha, \sigma_1, \sigma_2, N) \quad (4.1)$$

of the false alarm probability  $P_{\text{FA}}$ , the Gaussianity parameter  $\xi$ , the rms amplitude  $\alpha$  of events, the noise variances  $\sigma_1^2$  and  $\sigma_2^2$ , and the number of data points  $N$ . We can simplify Eq. (4.1) by replacing  $\alpha$  with the signal-to-noise ratio  $\rho$  using the definition (3.27), and noting from dimensional analysis that  $P_{\text{FA}}$  depends on  $\sigma_1$  and  $\sigma_2$  at fixed  $\rho$  only through the ratio  $\sigma_1/\sigma_2$ . This gives

$$P_{\text{FD}} = P_{\text{FD}}(P_{\text{FA}}, \xi, \rho, \sigma_1/\sigma_2, N). \quad (4.2)$$

For simplicity, we specialize to  $\sigma_1 = \sigma_2$  for the remainder of this paper. This implies that

$$P_{\text{FD}} = P_{\text{FD}}(P_{\text{FA}}, \xi, \rho, N). \quad (4.3)$$

##### 1. Cross correlation statistic

The false dismissal versus false alarm curves for the cross-correlation statistic can be computed analytically in the large  $N$  limit, as we now describe. Our derivation generalizes the analysis of Ref. [20] from Gaussian to non-Gaussian signals. For any detection statistic  $\Gamma$ , we can express  $P_{\text{FA}}$  and  $P_{\text{FD}}$  in terms of the detection threshold  $\Gamma_*$  as

$$P_{\text{FA}}(\Gamma_*, \sigma_1, \sigma_2, N) = \int_{\Gamma_*}^{\infty} dx p_{\Gamma|\mathcal{T}}(x|0), \quad (4.4)$$

$$P_{\text{FD}}(\Gamma_*, \xi, \rho, \sigma_1, \sigma_2, N) = 1 - \int_{\Gamma_*}^{\infty} dx p_{\Gamma|\mathcal{T}}(x|1). \quad (4.5)$$

Here the definition of the random variable  $\mathcal{T}$  is such that if  $\mathcal{T}=0$  then no signal is present ( $\xi = \rho = 0$ ), and if  $\mathcal{T}=1$  then a signal is present ( $\xi \neq 0$  and  $\rho \neq 0$ ); cf. Sec. II A above. Note that by eliminating  $\Gamma_*$  between Eqs. (4.4) and (4.5), we recover Eq. (4.1).

In the large  $N$  limit, the distribution  $p_{\Lambda_{\text{CC}}|\mathcal{T}}(x|t)$  is a Gaussian by the central limit theorem, and the integrals (4.4) and (4.5) can be evaluated analytically (see Appendix B) to give

$$P_{\text{FD}}(P_{\text{FA}}, \xi, \rho, N) = 1 - \frac{1}{2} \operatorname{erfc} \left[ \frac{\operatorname{erfc}^{-1}(2P_{\text{FA}}) - \frac{\rho}{\sqrt{2}}}{\sqrt{\frac{\rho^2}{N} \left( \frac{3}{\xi} - 1 \right) + \frac{2\rho}{\sqrt{N}} + 1}} \right] + O\left(\frac{1}{\sqrt{N}}\right). \quad (4.6)$$

Here the function  $\operatorname{erfc}(x)$  (known as the complement of the error function) is defined by

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} dy e^{-y^2}, \quad (4.7)$$

and  $\operatorname{erfc}^{-1}(x)$  is the inverse of  $\operatorname{erfc}(x)$ . The formula (4.6) is valid only for  $P_{\text{FA}} < 1/2$ ;  $P_{\text{FD}}$  is undefined for  $1/2 \leq P_{\text{FA}} < 1$ . In deriving Eq. (4.6), we assumed that the statistics  $\Lambda_{\text{CC}}$  and  $\hat{\alpha}^2$  are equivalent, and that the distribution for  $\hat{\alpha}^2$  is

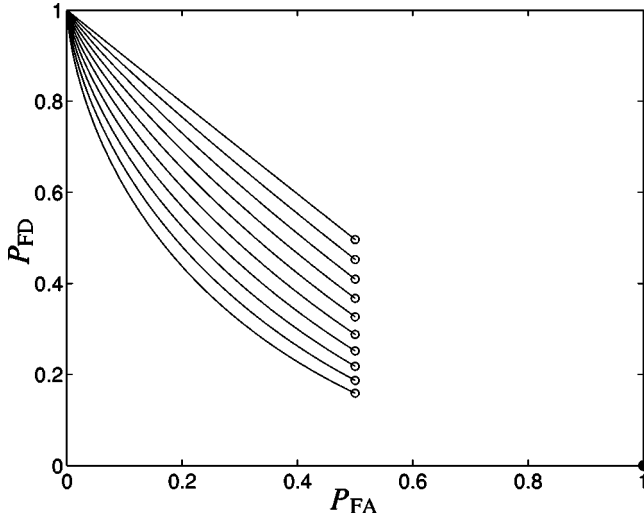


FIG. 5. Sample false dismissal versus false alarm curves for the cross correlation statistic  $\Lambda_{CC}$  in the large  $N$  limit, as prescribed by Eq. (4.8). For these curves the signal-to-noise ratio  $\rho$  has equally spaced values from 0.01 to 1. Note that here  $P_{FD}$  is undefined for  $1/2 \leq P_{FA} < 1$ .

Gaussian. Those assumptions are only valid up to fractional correction terms of order  $1/\sqrt{N}$ ; hence the indicated correction term in Eq. (4.6).

In the regime where  $\rho^2 \ll N\xi$  in addition to  $N \gg 1$ , the result (4.6) simplifies to

$$P_{FD}(P_{FA}, \xi, \rho, N) = 1 - \frac{1}{2} \operatorname{erfc} \left[ \operatorname{erfc}^{-1}(2P_{FA}) - \frac{\rho}{\sqrt{2}} \right] + O\left(\frac{1}{\sqrt{N}}\right) + O\left(\frac{\rho}{\sqrt{N}}\right) + O\left(\frac{\rho^2}{N\xi}\right). \quad (4.8)$$

Note that the false dismissal versus false alarm relation (4.8) is independent of both  $N$  and  $\xi$ . Sample curves from Eq. (4.8) are shown in Fig. 5. The discontinuities at  $P_{FA} = 1/2$  are a result of the step functions in the definition (3.11) of  $\hat{\alpha}^2$ .

## 2. Burst statistic

By combining the definition (1.9) of the burst statistic together with the decomposition (1.6), the noise and signal distributions (3.1) and (3.19), and the change of variables (3.27) it is straightforward to derive the exact false alarm versus false dismissal relation. The result is given by

$$(1 - P_{FA})^{1/N} = \operatorname{erf} \left( \frac{\Lambda_*}{\sqrt{2}} \right) \quad (4.9)$$

and

$$P_{FD}^{1/N} = \xi \operatorname{erf} \left[ \frac{\Lambda_*}{\sqrt{2 + \frac{2\rho}{\xi\sqrt{N}}}} \right] + (1 - \xi) \operatorname{erf} \left( \frac{\Lambda_*}{\sqrt{2}} \right), \quad (4.10)$$

where  $\Lambda_*$  is the value of the threshold.

## 3. Maximum likelihood statistic

We start by discussing the different regimes present in the space of signal parameters  $\xi$ ,  $\rho$  and  $N$ , treating the false alarm probability  $P_{FA}$  as fixed. There are several different constraints on the three parameters  $\xi$ ,  $\rho$ , and  $N$  that define the regime in parameter space where we expect our maximum likelihood statistic to work well. First, it is clear that the total number of events  $\sim \xi N$  in the data set must be large compared to one:

$$\xi \gg \frac{1}{N}. \quad (4.11)$$

Second, if the signal-to-noise ratio  $\alpha^2/(\sigma_1\sigma_2)$  of individual burst events is large compared to one, then one can detect the individual events using the burst statistic (1.9) and the method of this paper is not needed. From Eq. (3.18) we can write the constraint  $\alpha^2/(\sigma_1\sigma_2) \leq 1$  as

$$\xi \geq \frac{\rho}{\sqrt{N}}. \quad (4.12)$$

A more precise version of this requirement can be obtained by noting that the detection threshold for the signal-to-noise ratio  $\alpha^2/(\sigma_1\sigma_2)$  is  $\sim \sqrt{2 \ln N}$ , since there are  $N$  independent trials. This yields the constraint

$$\xi \geq \frac{\rho}{\sqrt{2N \ln N}}. \quad (4.13)$$

The regime  $\xi \sim \rho/\sqrt{2N \ln N}$  is where the burst statistic  $\Lambda_B$  starts becoming as sensitive as the cross correlation statistic, as can be seen by combining Eqs. (4.8), (4.9) and (4.10) above. This behavior can also be seen in Figs. 1 and 2 above.

A third constraint on the space of signal parameters is derived as follows. Consider the statistic

$$\eta = \frac{1}{N} \sum_{k=1}^N (h_1^k)^2 (h_2^k)^2. \quad (4.14)$$

We can use this statistic to estimate the Gaussianity parameter  $\xi$  in the following way. The mean value of  $\eta$  when a signal is present is given by

$$\langle \eta \rangle = 3\xi\alpha^4 + \xi\alpha^2(\sigma_1^2 + \sigma_2^2) + \sigma_1^2\sigma_2^2, \quad (4.15)$$

and the variance when a signal is absent is

$$(\Delta\eta)^2 = \sigma_1^4\sigma_2^4 \frac{8}{N}. \quad (4.16)$$

It follows from Eqs. (4.15), (4.16), and the relation  $\langle \hat{\alpha}^2 \rangle = \xi \alpha^2$  that the estimator  $\hat{\xi}$  of  $\xi$  defined by

$$\hat{\xi} = \frac{3\hat{\alpha}^4}{\eta - \hat{\alpha}^2(\hat{\sigma}_1^2 + \hat{\sigma}_2^2) - \hat{\sigma}_1^2\hat{\sigma}_2^2} \quad (4.17)$$

has a fractional accuracy of order

$$\frac{\Delta \hat{\xi}}{\hat{\xi}} \sim \frac{\xi \sqrt{N}}{\rho^2}. \quad (4.18)$$

Now in the regime  $\Delta \hat{\xi}/\hat{\xi} \ll 1$ , we expect our maximum likelihood detection statistic to work well, since one's first guess for a nonlinear statistic (4.14) can be used to detect the non-Gaussianity of the signal to high accuracy. In the regime  $\Delta \hat{\xi}/\hat{\xi} \gg 1$ , it is not obvious how the maximum likelihood detection statistic will perform, since it could have a performance much better than that of the statistic  $\eta$ . However, our Monte Carlo simulations (Sec. IV B below) and analytic computations (Appendix C) indicate that the maximum likelihood statistic does indeed perform poorly in the regime  $\Delta \hat{\xi}/\hat{\xi} \gg 1$ . Thus, our third constraint is  $\Delta \hat{\xi}/\hat{\xi} \lesssim 1$ , which from Eq. (4.18) can be written as

$$\hat{\xi} \lesssim \frac{\rho^2}{\sqrt{N}}. \quad (4.19)$$

Our Monte Carlo simulations show that for  $\rho^2/\sqrt{N} \lesssim \hat{\xi} \lesssim 1$ , the maximum likelihood and cross-correlation statistics perform roughly equivalently, and that once  $\hat{\xi}$  becomes smaller than  $\rho^2/\sqrt{N}$ , the maximum likelihood statistic starts to perform significantly better than the cross-correlation statistic; see Figs. 1 and 2 above.

In Appendix C we derive analytically the approximate expression (C46) for the false dismissal probability for the maximum likelihood statistic, which we expect to be accurate up to corrections of order  $1/\rho^4$  or a few tens of percent. We also derive the expression (C55) for the false alarm probability using a combination of analytical and numerical techniques. Combining these results gives the curves which are associated with the maximum likelihood statistic  $\Lambda_{\text{ML}}^{\text{NG}}$  and labeled ‘‘analytic’’ in Figs. 1, 2, 9, and 10.

### B. Description of the Monte Carlo simulation algorithm

Next we describe our Monte Carlo simulations of the performances of the various statistics. We numerically estimate the false dismissal and false alarm probabilities  $P_{\text{FD}}$  and  $P_{\text{FA}}$  by conducting an ensemble of  $N_E$  simulated experiments. For each experiment we simulate a detector output matrix, half of which have a signal present, and half of which do not. Since we know in advance whether or not a signal is present, we can easily estimate  $P_{\text{FA}}$  and  $P_{\text{FD}}$ . More specifically, our algorithm for simulating false dismissal versus false alarm curves, for an arbitrary statistic  $\Gamma$ , is as follows:

- (1) Choose values for  $\xi$ ,  $\alpha$ ,  $\sigma_1$ ,  $\sigma_2$ , and  $N$ .
- (2) Choose the total number of trials  $N_E$ .
- (3) For  $r=1, 2, \dots, N_E/2$ :
  - (a) Generate a data train  $h(\sigma_1, \sigma_2, N)$  of noise only.
  - (b) Compute  $\Gamma$  and store result as  $\Gamma_{r0}$ .
  - (c) Generate a data train  $h(\xi, \alpha, \sigma_1, \sigma_2, N)$  which has a signal present.
  - (d) Compute  $\Gamma$  and store result as  $\Gamma_{r1}$ .
- (4) Choose a discretization  $\Gamma_{*j}$  of the set of thresholds, where  $j=1, 2, \dots, M$ .
- (5) Set  $P_{\text{FA}}(\Gamma_{*j}) = P_{\text{FD}}(\Gamma_{*j}) = 0$ , for each  $j$ .
- (6) For  $r=1, 2, \dots, N_E/2$ :
  - (a) For each  $j$ , if  $\Gamma_{r0} > \Gamma_{*j}$ , increment  $P_{\text{FA}}(\Gamma_{*j})$  by  $2/N_E$ .
  - (b) For each  $j$ , if  $\Gamma_{r1} \leq \Gamma_{*j}$ , increment  $P_{\text{FD}}(\Gamma_{*j})$  by  $2/N_E$ .
- (7) Repeat steps (3)–(6) above several times to estimate the fluctuations in  $P_{\text{FA}}(\Gamma_{*j})$  and  $P_{\text{FD}}(\Gamma_{*j})$ .

We use the above algorithm to simulate false dismissal versus false alarm curves for the three statistics  $\Lambda_{\text{CC}}$ ,  $\Lambda_{\text{B}}$  and  $\Lambda_{\text{ML}}^{\text{NG}}$ . The analytical expressions (4.6) and (4.9), (4.10) for the cross-correlation and burst statistics are used as a check of the numerical method.

### C. Simulation results

A family of simulated false dismissal versus false alarm curves for the cross correlation statistic  $\Lambda_{\text{CC}}$  and the maximum likelihood statistic  $\Lambda_{\text{ML}}^{\text{NG}}$  is shown in Fig. 6. We see that at fixed  $\rho$ , as the Gaussianity  $\xi$  of the signal decreases,  $\Lambda_{\text{ML}}^{\text{NG}}$  performs increasingly better than  $\Lambda_{\text{CC}}$ . The curves for  $\Lambda_{\text{CC}}$  are almost indistinguishable from each other because  $\rho$  is fixed, and the curves depend only on  $\rho$  and not on  $\xi$  for this detection statistic in the large  $N$  limit [cf. Eq. (4.8) above].

If we maintain the same value for  $\rho$  as in Fig. 6, but take  $\xi \geq 0.03$ , the curves for  $\Lambda_{\text{CC}}$  and  $\Lambda_{\text{ML}}^{\text{NG}}$  cannot be distinguished from each other. We find in general that for *any* values of  $N$ ,  $\sigma_1$ ,  $\sigma_2$ , and  $\rho$ , as  $\xi \rightarrow 1$ , the false dismissal versus false alarm curves for  $\Lambda_{\text{CC}}$  and  $\Lambda_{\text{ML}}^{\text{NG}}$  cannot be distinguished from each other. Thus, the two statistics are nearly equivalent for Gaussian signals, as expected. However, for  $\xi \ll 1$ , Fig. 6 demonstrates that  $\Lambda_{\text{ML}}^{\text{NG}}$  performs noticeably better than  $\Lambda_{\text{CC}}$ .

We now discuss a comparison of the two statistics in terms of the minimum gravitational wave energy density necessary for detection, instead of in terms of the false dismissal versus false alarm curves. For a stochastic background with rms strain amplitude  $h_{\text{rms}}$ , we have  $\Omega \propto h_{\text{rms}}^2$  [5], where  $\Omega$  is the gravitational wave energy density. For our model signal (3.19) we have  $h_{\text{rms}}^2 \propto \xi \alpha^2$ , and comparing this with the formula  $\rho \propto \xi \alpha^2$  from Eq. (3.27) shows that we can interpret the signal to noise ratio  $\rho$  as the energy density in the stochastic background, even for non-Gaussian signals.

We compute the minimum detectable energy density or signal-to-noise ratio  $\rho_{\text{detectable}}$  as follows. First, we choose thresholds  $P_{\text{FA}*}$  and  $P_{\text{FD}*}$  for the false alarm and false dismissal probabilities. We refer to the pair  $(P_{\text{FA}*}, P_{\text{FD}*})$  as the *detection point*. For any statistic  $\Gamma$ , the choice of detection point determines the detection threshold  $\Gamma_{*}$ , and inverting

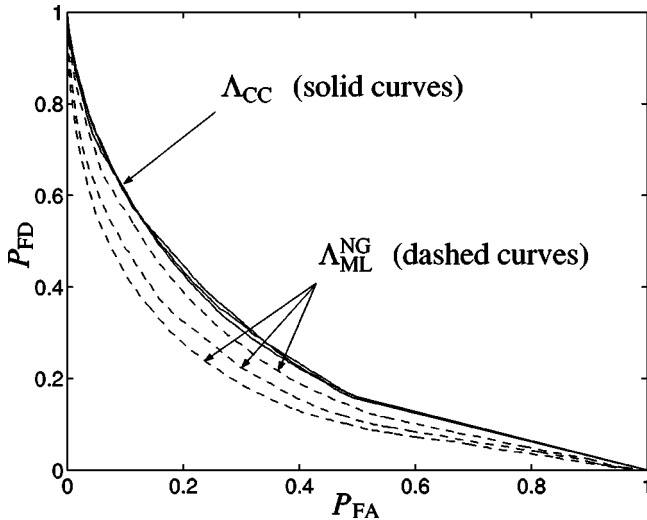


FIG. 6. Plots of false dismissal probability ( $P_{FD}$ ) versus false alarm probability ( $P_{FA}$ ) for the standard cross-correlation statistic  $\Lambda_{CC}$  and our maximum likelihood statistic  $\Lambda_{ML}^{NG}$ . Each of these curves is characterized by a total number of trials  $N_E = 2 \times 10^4$ , number of data points  $N = 5 \times 10^4$ , noise variances  $\sigma_1 = \sigma_2 = 1$ , and by the signal-to-noise ratio  $\rho = 1$ . The values of the Gaussianity parameter  $\xi$  are 0.02, 0.012, and 0.01. The solid curves are the results for  $\Lambda_{ML}^G$ ; these curves are bunched together because  $\rho$  is fixed. The dashed curves are the results for  $\Lambda_{ML}^{NG}$ . For the dashed curves, the lowest curve is for  $\xi = 0.01$ , while the highest curve is for  $\xi = 0.02$ . We estimate error bars for each of these curves by separating the  $2 \times 10^4$  runs into 10 bins of  $2 \times 10^3$ , and generating 10 separate plots; the resulting fluctuations are  $\leq 10^{-3}$ . The curves for the cross correlation statistic  $\Lambda_{ML}^G$  agree with the analytic prediction (4.6) to within  $\sim 10^{-3}$ . This plot shows that  $\Lambda_{ML}^{NG}$  can perform significantly better than  $\Lambda_{CC}$ .

Eq. (4.3) gives the minimum detectable signal-to-noise ratio

$$\rho = \rho_{\text{detectable}}(P_{FA*}, P_{FD*}, \xi, N), \quad (4.20)$$

as illustrated in Fig. 7. For the cross-correlation statistic  $\Lambda_{CC}$  the result is, from Eq. (4.6),

$$\rho_{\text{detectable}}^{CC} = \frac{2\sqrt{2}\gamma[1 + \gamma\sqrt{2/N}]}{1 + 2\gamma^2\left(1 - \frac{3}{\xi}\right)/N} \left[1 + O\left(\frac{1}{\sqrt{N}}\right)\right] \quad (4.21)$$

$$= 2\sqrt{2}\gamma + O\left(\frac{1}{\sqrt{N}}\right) + O\left(\frac{\gamma}{\sqrt{N}}\right), \quad (4.22)$$

where  $\gamma = \text{erfc}^{-1}(2P_{FA*})$  and we have assumed that  $P_{FA*} = P_{FD*}$ . This relation is plotted in Fig. 8.

From the results of our simulations, we determine  $\rho_{\text{detectable}}(P_{FA*}, P_{FD*}, \xi, N)$  by numerically solving the equation

$$P_{FD}(P_{FA*}, \xi, \rho, N) - P_{FD*} = 0 \quad (4.23)$$

for  $\rho$ . Unfortunately, evaluating the function on the left hand side of Eq. (4.23) is computationally expensive. Each evaluation involves simulating the false dismissal versus false

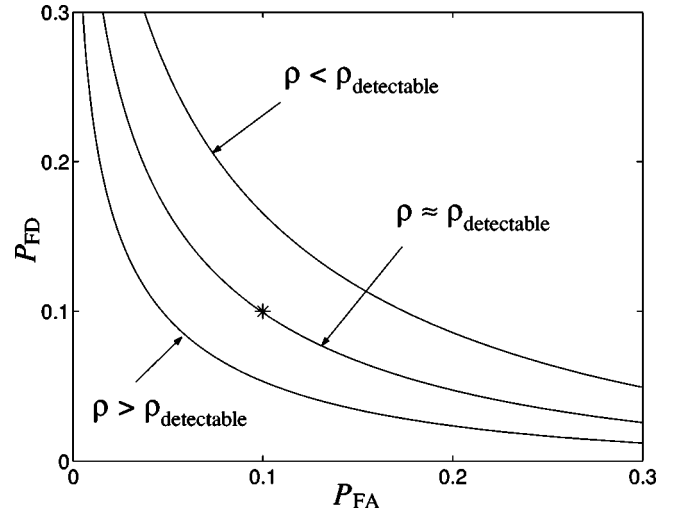


FIG. 7. A family of false dismissal versus false alarm curves for fixed  $\xi$ . Here the detection point, at  $P_{FD*} = P_{FA*} = 0.1$ , is marked with an \*.

alarm curve which is itself a computationally intensive task. Moreover, it is only feasible for us to solve Eq. (4.23) for values of  $N \leq 10^4$  while a realistic detection scenario for ground based detectors would involve a year's worth of data sampled at  $\sim 100$  Hz for which  $N \sim 10^9$ . Therefore our conclusions about the applicability of the method to ground based detectors are based on our analytic results, as discussed in the Introduction.

Figure 9 shows the results obtained from numerically solving Eq. (4.23) for  $\rho_{\text{detectable}}$  for the parameter values  $\xi = 0.02$ ,  $P_{FA*} = P_{FD*} = 0.1$ , and Fig. 10 shows the corresponding results for  $\xi = 4.3 \times 10^{-3}$ . For the cross-correlation statistic, the results are in good agreement with the analytic prediction (4.21).

Figure 1 shows the minimum detectable energy density as

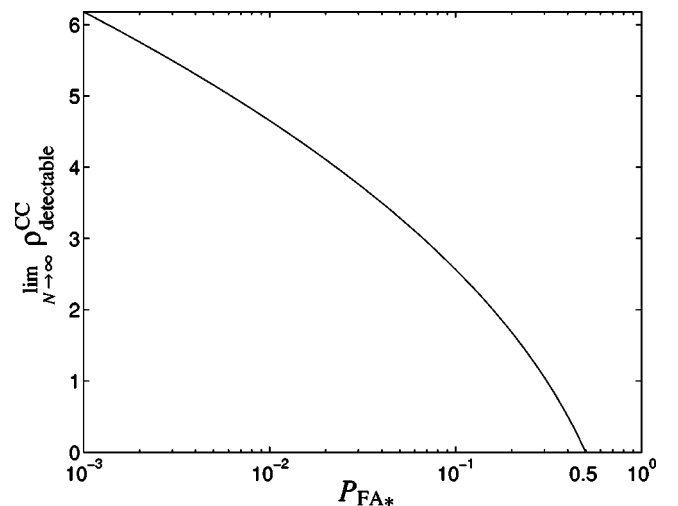


FIG. 8. The minimum detectable signal-to-noise ratio  $\rho_{\text{detectable}}^{CC}$  for the cross-correlation statistic  $\Lambda_{CC}$  as a function of the false alarm probability threshold  $P_{FA*}$ . Note that we assume the false dismissal probability threshold  $P_{FD*} = P_{FA*}$ .

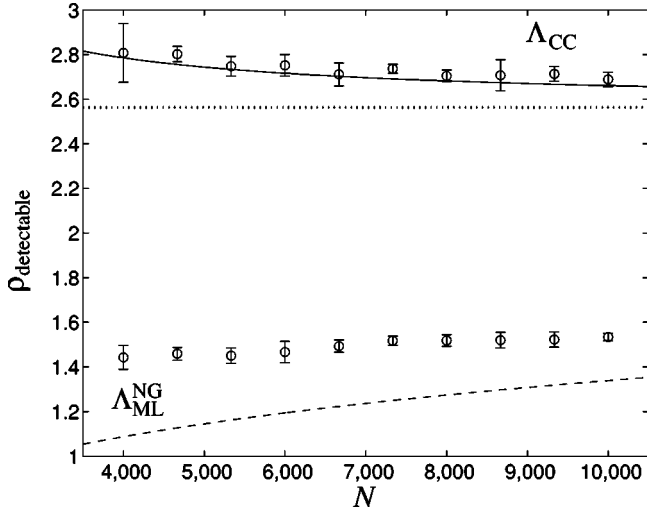


FIG. 9. The minimum detectable signal strength  $\rho_{\text{detectable}}$  as a function of the number of data points  $N$ , for the false alarm probability threshold  $P_{\text{FA}^*}=0.1$ , false dismissal probability threshold  $P_{\text{FD}^*}=0.1$ , and Gaussianity parameter  $\xi=0.02$ . The circles are the simulation results, and the error bars are estimated from ten different runs. The solid curve is the analytical prediction (4.21) for  $\Lambda_{\text{CC}}$ , and the dotted line is the  $N \rightarrow \infty$  limit (4.22). The dashed line is the analytic prediction for  $\Lambda_{\text{ML}}^{\text{NG}}$  given by Eqs. (C46) and (C55).

a function of the Gaussianity parameter  $\xi$  for  $N=10^4$  (corresponding to space based detectors), for the cross-correlation and maximum likelihood statistics and also for the burst statistic (1.9). We again use the values  $P_{\text{FA}^*}=P_{\text{FD}^*}=0.1$ . The figure shows that the maximum likelihood statistic performs better than the other statistics by a factor which is roughly 3 for  $\xi$  of order 1%. For smaller values of  $\xi$ , the maximum likelihood performs increasingly better than the cross-correlation statistic, but is eventually comparable to the burst statistic. Thus the maximum likelihood statistic gives an improvement in sensitivity to backgrounds composed of roughly 10 to  $10^3$  events per year.

Figure 2 is a similar plot, without the Monte Carlo simulation results, for  $N=10^9$  (corresponding to ground based

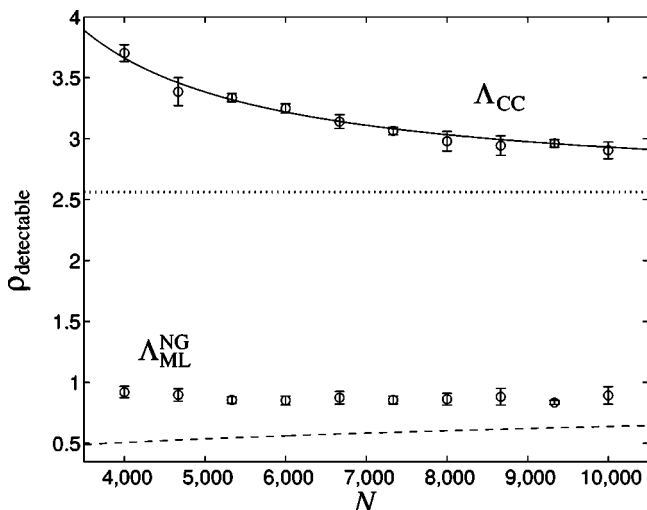


FIG. 10. Same as Fig. 9 but with  $\xi=4.3 \times 10^{-3}$ .

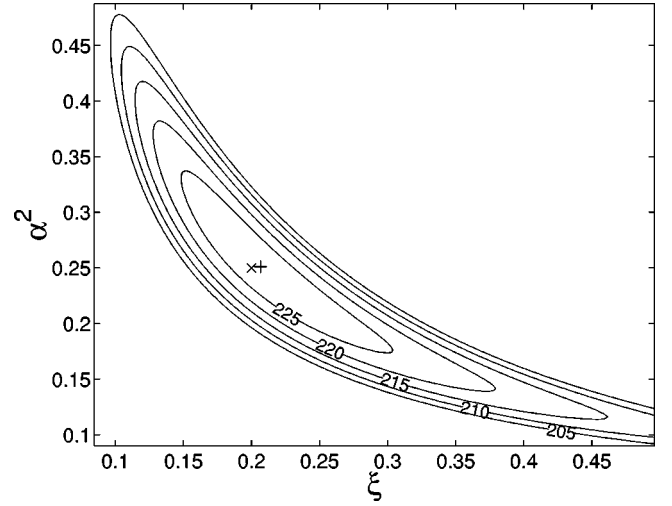


FIG. 11. Representative contours of  $\ln \lambda(\xi, \alpha^2, \hat{\sigma}_1^2, \hat{\sigma}_2^2)$ . Here  $\rho=20$  and  $N=1.6 \times 10^5$ . The simulated signal is characterized by  $\xi=0.2$  and  $\alpha^2=0.25$ , marked with an  $\times$ . The noise is characterized by  $\sigma_1^2=\sigma_2^2=1$ . The maximum, marked with a  $+$ , is found at  $\ln \lambda(0.207, 0.251, 0.993, 0.993)=229$ , while  $\ln \lambda(0.2, 0.25, 1, 1)=227$ .

detectors). Here we use  $P_{\text{FA}^*}=P_{\text{FD}^*}=0.01$ . The results are similar to those in Fig. 1, except that here the gain in sensitivity occurs in the band  $10^{-5} < \xi < 10^{-3}$ . This band corresponds to  $10^4$ – $10^6$  events per year.

#### D. Parameter estimation

The computation of the maximum likelihood statistic also serves to measure the parameters of the signal. The statistic  $\Lambda_{\text{ML}}^{\text{NG}}$ , from Eq. (3.22), can be written as

$$\Lambda_{\text{ML}}^{\text{NG}} = \max_{0 < \xi \leq 1} \max_{\alpha^2 > 0} \max_{\sigma_1^2 \geq 0} \max_{\sigma_2^2 \geq 0} \lambda(\xi, \alpha^2, \sigma_1^2, \sigma_2^2). \quad (4.24)$$

The point  $(\hat{\xi}, \hat{\alpha}^2, \hat{\sigma}_1^2, \hat{\sigma}_2^2)$  where this maximum is achieved is the maximum likelihood estimator for  $(\xi, \alpha^2, \sigma_1^2, \sigma_2^2)$ . In Fig. 11 we show contours of the function  $\ln \lambda$  for a strong ( $\rho=20$ ) signal. This figure shows that both  $\xi$  and  $\alpha^2$  can be measured with good accuracy.

Note that the main benefit of using  $\Lambda_{\text{ML}}^{\text{NG}}$  is that it allows us to detect signals that are too weak to be seen using  $\Lambda_{\text{CC}}$ . Using  $\Lambda_{\text{ML}}^{\text{NG}}$  also allows one to test if a detected signal is Gaussian, as obtained above, but this is not the main benefit of the method, as there are other, simpler, methods to test for non-Gaussianity.

#### V. CONCLUSIONS

The use of our maximum likelihood statistic in searches for a non-Gaussian background gives a gain in sensitivity over the standard cross-correlation statistic. Figures 1 and 2 show that the gain factor can be significant for sufficiently non-Gaussian signals. However, computing the maximum likelihood statistic requires significantly more computational



power than the cross-correlation statistic.

The analysis presented here must be generalized in several ways before being usable in gravitational wave detectors. These generalizations, listed in order of importance, are:

(1) Our signal model (3.19) assumes a Gaussian distribution of amplitudes of the burst events. This assumption simplified our analysis and resulted in a statistic with the useful property of being nearly equivalent to the cross-correlation statistic in the Gaussian signal limit. In practice however, the distribution of the events should instead be based on the candidate sources. For example, a popcorn-like stochastic background produced by a spatially uniform distribution of standard-candle sources out to some maximum redshift would have a signal distribution of the form (3.19) with the Gaussian term replaced by a term proportional to  $s^{-4}\theta(s-s_{\min})$ , where  $\theta$  is the step function and  $s_{\min}$  is a cutoff signal strength.

(2) One should allow the burst durations to be longer than the detector resolution time. For this situation one possibility would be to preprocess the data with a lowpass filter, and then apply the techniques developed here. Another possibility would be to try to combine the analysis of this paper with the excess power detection method of Ref. [25].

(3) Real detector noise always contains non-Gaussian components, so one needs to generalize the analysis to allow for this. Such a generalization for a Gaussian stochastic background can be found in Refs. [21,22].

(4) It would be useful to consider a more general signal model which consists of a superposition of a Gaussian background and a non-Gaussian background, since the true gravitational wave background might consist of such a superposition.

(5) The analysis needs to be generalized to allow for colored detector noise, and separated, misaligned detectors. This generalization should be fairly straightforward.

### ACKNOWLEDGMENTS

We thank Wolfgang Tichy, Tom Lored, Teviet Creighton, and Bernard Whiting for helpful discussions. The analytic computations in Appendix C were carried out using the software package MATHEMATICA. This work was supported in part by National Science Foundation awards PHY-9722189 and PHY-0140209, the Alfred P. Sloan Foundation, the Radcliffe Institute for Advanced Study, and the NASA/New York Space Grant Consortium.

### APPENDIX A: GENERAL FORM OF THE LIKELIHOOD RATIO

In this appendix we give two derivations of the general formula (2.15) for the likelihood ratio. The first derivation is based on Eq. (2.13) while the second is based on Eq. (2.14). We also derive the formula (2.18) for the posterior probability density  $p_{\mathcal{V}_s, \mathcal{V}_n | \mathcal{T}}^{(1)}(\mathbf{v}_s, \mathbf{v}_n | 1)$ .

#### 1. First derivation

We can derive Eq. (2.15) by using the total probability theorem to expand the distributions in the numerator and denominator of Eq. (2.13). Note that all distributions in this derivation are priors.

First expand  $p_{\mathcal{H}}(h)$  just in terms of the random variable  $\mathcal{T}$ :

$$p_{\mathcal{H}}(h) = P_{\mathcal{T}}(1)p_{\mathcal{H}|\mathcal{T}}(h|1) + P_{\mathcal{T}}(0)p_{\mathcal{H}|\mathcal{T}}(h|0). \quad (\text{A1})$$

Expanding  $p_{\mathcal{H}}(h)$  in terms of all the degrees of freedom yields

$$\begin{aligned} p_{\mathcal{H}}(h) &= \sum_{t=0}^1 \int_{\Theta_s} d^{\mathcal{Q}_s} v_s \int d^{ND} s \\ &\quad \times \int_{\Theta_n} d^{\mathcal{Q}_n} v_n p_{\mathcal{H}|\mathcal{T}, \mathcal{V}_s, \mathcal{S}, \mathcal{V}_n}(h|t, \mathbf{v}_s, s, \mathbf{v}_n) \\ &\quad \times p_{\mathcal{T}, \mathcal{V}_s, \mathcal{S}, \mathcal{V}_n}(t, \mathbf{v}_s, s, \mathbf{v}_n). \end{aligned} \quad (\text{A2})$$

The ratio of the coefficients of  $P_{\mathcal{T}}(1)$  and  $P_{\mathcal{T}}(0)$  in Eq. (A2) will give the general expression for the likelihood ratio by Eq. (2.13).

The conditional distribution for  $\mathcal{H}$  in Eq. (A2) can be translated into a conditional distribution for  $\mathcal{N}$ . From Eq. (2.1) it follows that

$$p_{\mathcal{H}|\mathcal{S}}(h|s) = p_{\mathcal{N}+\mathcal{S}|\mathcal{S}}(h|s) = p_{\mathcal{N}|\mathcal{S}}(h-s|s), \quad (\text{A3})$$

and since  $\mathcal{S}$  and  $\mathcal{N}$  are statistically independent we obtain

$$P_{\mathcal{H}|\mathcal{S}}(h|s) = P_{\mathcal{N}}(h-s). \quad (\text{A4})$$

Generalizing this argument gives

$$p_{\mathcal{H}|\mathcal{T}, \mathcal{V}_s, \mathcal{S}, \mathcal{V}_n}(h|t, \mathbf{v}_s, s, \mathbf{v}_n) = p_{\mathcal{N}|\mathcal{V}_n}(h-s|\mathbf{v}_n), \quad (\text{A5})$$

since *a priori*  $\mathcal{T}$ ,  $\mathcal{V}_s$ , and  $\mathcal{S}$  are statistically independent of  $\mathcal{N}$  and  $\mathcal{V}_n$ . For the same reason we can write the joint distribution that appears in Eq. (A2) as

$$p_{\mathcal{T}, \mathcal{V}_s, \mathcal{S}, \mathcal{V}_n}(t, \mathbf{v}_s, s, \mathbf{v}_n) = p_{\mathcal{T}, \mathcal{V}_s, \mathcal{S}}(t, \mathbf{v}_s, s) p_{\mathcal{V}_n}(\mathbf{v}_n). \quad (\text{A6})$$

Substituting Eqs. (A5) and (A6) into Eq. (A2) yields

$$\begin{aligned} p_{\mathcal{H}}(h) &= \sum_{t=0}^1 \int_{\Theta_s} d^{\mathcal{Q}_s} v_s \int d^{ND} s \int_{\Theta_n} d^{\mathcal{Q}_n} v_n \\ &\quad \times p_{\mathcal{N}|\mathcal{V}_n}(h-s|\mathbf{v}_n) p_{\mathcal{T}, \mathcal{V}_s, \mathcal{S}}(t, \mathbf{v}_s, s) p_{\mathcal{V}_n}(\mathbf{v}_n). \end{aligned} \quad (\text{A7})$$

We can also rewrite the distribution  $p_{\mathcal{T}, \mathcal{V}_s, \mathcal{S}}(t, \mathbf{v}_s, s)$  as

$$p_{\mathcal{T}, \mathcal{V}_s, \mathcal{S}}(t, \mathbf{v}_s, s) = p_{\mathcal{S}|\mathcal{V}_s, \mathcal{T}}(s|\mathbf{v}_s, t) p_{\mathcal{V}_s|\mathcal{T}}(\mathbf{v}_s, t) P_{\mathcal{T}}(t), \quad (\text{A8})$$

by Eq. (2.5). Substituting Eq. (A8) into Eq. (A7) and explicitly evaluating the sum over  $t$  yields

$$\begin{aligned} p_{\mathcal{H}}(h) &= P_{\mathcal{T}}(1) \int_{\Theta_{s1}} d^{\mathcal{Q}_s} v_s \int d^{ND} s \int_{\Theta_n} d^{\mathcal{Q}_n} v_n \\ &\quad \times p_{\mathcal{N}|\mathcal{V}_n}(h-s|\mathbf{v}_n) p_{\mathcal{V}_n}(\mathbf{v}_n) p_{\mathcal{S}|\mathcal{V}_s, \mathcal{T}}(s|\mathbf{v}_s, 1) \\ &\quad \times p_{\mathcal{V}_s|\mathcal{T}}(\mathbf{v}_s|1) + P_{\mathcal{T}}(0) \int_{\Theta_n} d^{\mathcal{Q}_n} v_n p_{\mathcal{N}|\mathcal{V}_s}(h|\mathbf{v}_n). \end{aligned} \quad (\text{A9})$$

Here we have used the following relations:

$$p_{S|\mathcal{V}_s, \mathcal{T}}(s|\mathbf{v}_s \in \Theta_{s1}, 0) = \delta^{ND}(s) \quad (\text{A10})$$

$$p_{\mathcal{V}_s|\mathcal{T}}(\mathbf{v}_s \in \Theta_{s0}|1) = 0 \quad (\text{A11})$$

$$p_{\mathcal{V}_s|\mathcal{T}}(\mathbf{v}_s \in \Theta_{s1}|0) = 0 \quad (\text{A12})$$

$$\int_{\Theta_{s0}} d^{\mathcal{Q}_s} v_s p_{\mathcal{V}_s|\mathcal{T}}(\mathbf{v}_s|0) = 1. \quad (\text{A13})$$

By comparing Eqs. (A1) and (A9) we can read off the distributions  $p_{\mathcal{H}|\mathcal{T}}(h|t)$  and construct Eq. (2.15) from Eq. (2.13). Note that the expression (2.13) is independent of the space  $\Theta_{s0}$  of signal parameters corresponding to “no signal present.”

## 2. Second derivation

Here we derive Eq. (2.15), and also Eq. (2.18), from Eq. (2.14). Consider the distribution

$$p_{\mathcal{T}, \mathcal{V}_s, \mathcal{V}_n | \mathcal{H}}(1, \mathbf{v}_s, \mathbf{v}_n | h) = \frac{p_{\mathcal{T}, \mathcal{V}_s, \mathcal{V}_n, \mathcal{H}}(1, \mathbf{v}_s, \mathbf{v}_n, h)}{p_{\mathcal{H}}(h)}. \quad (\text{A14})$$

We will justify Eq. (2.15) by the defining relation Eq. (2.14), which explicitly refers to priors and posteriors. Therefore we now append the appropriate superscripts as bookkeeping devices. Equation (A14) then reads

$$p_{\mathcal{T}, \mathcal{V}_s, \mathcal{V}_n}^{(1)}(1, \mathbf{v}_s, \mathbf{v}_n) = \frac{p_{\mathcal{T}, \mathcal{V}_s, \mathcal{V}_n, \mathcal{H}}^{(0)}(1, \mathbf{v}_s, \mathbf{v}_n, h)}{p_{\mathcal{H}}^{(0)}(h)}. \quad (\text{A15})$$

Using the expansion of  $p_{\mathcal{H}}(h)$  given by Eq. (A9), and what we will justify is the likelihood ratio  $\Lambda$  given by Eq. (2.15), we have

$$p_{\mathcal{T}, \mathcal{V}_s, \mathcal{V}_n}^{(1)}(1, \mathbf{v}_s, \mathbf{v}_n) = \frac{\left[ \frac{p_{\mathcal{T}, \mathcal{V}_s, \mathcal{V}_n, \mathcal{H}}^{(0)}(1, \mathbf{v}_s, \mathbf{v}_n, h)}{\int_{\Theta_n} d^{\mathcal{Q}_n} v'_n p_{\mathcal{H}|\mathcal{V}_n}^{(0)}(h|\mathbf{v}_n) p_{\mathcal{V}_n}^{(0)}(\mathbf{v}_n)} \right]}{\Lambda P^{(0)} + 1 - P^{(0)}}. \quad (\text{A16})$$

Expanding the uppermost numerator in Eq. (A16) over  $\mathcal{S}$  by the total probability theorem gives

$$p_{\mathcal{T}, \mathcal{V}_s, \mathcal{V}_n, \mathcal{H}}^{(0)}(1, \mathbf{v}_s, \mathbf{v}_n, h) = \int d^{ND} s p_{\mathcal{T}, \mathcal{V}_s, \mathcal{V}_n, \mathcal{H}, \mathcal{S}}^{(0)}(1, \mathbf{v}_s, \mathbf{v}_n, h, s), \quad (\text{A17})$$

and rewriting this gives

$$\begin{aligned} & p_{\mathcal{T}, \mathcal{V}_s, \mathcal{V}_n, \mathcal{H}, \mathcal{S}}^{(0)}(1, \mathbf{v}_s, \mathbf{v}_n, h, s) \\ &= p_{\mathcal{N}|\mathcal{V}_s}^{(0)}(h-s|\mathbf{v}_s) p_{\mathcal{V}_n}^{(0)}(\mathbf{v}_n) p_{S|\mathcal{V}_s, \mathcal{T}}^{(0)}(s|\mathbf{v}_s, 1) \\ & \quad \times p_{\mathcal{V}_s|\mathcal{T}}^{(0)}(\mathbf{v}_s, 1) P^{(0)}. \end{aligned} \quad (\text{A18})$$

After putting Eq. (A18) into Eq. (A17), substitute the result into Eq. (A16). Using  $\Lambda(\mathbf{v}_s, \mathbf{v}_n)$  given by Eq. (2.17) then yields

$$P^{(1)} p_{\mathcal{V}_s, \mathcal{V}_n | \mathcal{T}}^{(1)}(\mathbf{v}_s, \mathbf{v}_n | 1) = \frac{\Lambda(\mathbf{v}_s, \mathbf{v}_n) P^{(0)}}{\Lambda P^{(0)} + 1 - P^{(0)}}. \quad (\text{A19})$$

On the left hand side of Eq. (A19) we have used

$$p_{\mathcal{T}, \mathcal{V}_s, \mathcal{V}_n}^{(1)}(1, \mathbf{v}_s, \mathbf{v}_n) = P^{(1)} p_{\mathcal{V}_s, \mathcal{V}_n | \mathcal{T}}^{(1)}(\mathbf{v}_s, \mathbf{v}_n | 1). \quad (\text{A20})$$

Integrate Eq. (A19) over  $\Theta_n$  and  $\Theta_{s1}$  using Eq. (2.16) and the normalization requirement

$$\int_{\Theta_{s1}} d^{\mathcal{Q}_s} v_s \int_{\Theta_n} d^{\mathcal{Q}_n} v_n p_{\mathcal{V}_s, \mathcal{V}_n | \mathcal{T}}^{(1)}(\mathbf{v}_s, \mathbf{v}_n | 1) = 1 \quad (\text{A21})$$

to get

$$P^{(1)} = \frac{\Lambda P^{(0)}}{\Lambda P^{(0)} + 1 - P^{(0)}}. \quad (\text{A22})$$

Use Eq. (A22) and Eq. (A19) to form the ratio on the left hand side of Eq. (2.18). This justifies Eq. (2.18).

Integrate Eq. (2.18) over  $\Theta_n$  and  $\Theta_{s1}$  using Eq. (2.16) and Eq. (A21) to see that the defining relation Eq. (2.14) is satisfied and thus Eq. (2.15) is justified.

## APPENDIX B: ANALYTICAL EXPRESSIONS FOR FALSE DISMISSAL VERSUS FALSE ALARM CURVES FOR CROSS-CORRELATION STATISTIC

This appendix derives the analytical form (4.6) of the false dismissal versus false alarm curves for the cross-correlation statistic  $\Lambda_{CC}$  in the large  $N$  limit, for both Gaussian and non-Gaussian signals. A derivation for Gaussian signals can be found in Sec. IV of Ref. [20].

As noted in Sec. III B, the statistics  $\Lambda_{CC}$  and  $\hat{\alpha}^2$  are equivalent in the large  $N$  limit. Thus, in this limit, the false dismissal versus false alarm curves can be found by evaluating Eqs. (4.4) and (4.5) with  $\Gamma$  replaced by  $\hat{\alpha}^2$ . The relation (3.11) between the statistics  $\bar{\alpha}^2$  and  $\hat{\alpha}^2$  implies the following relation between their probability distributions  $p_{\hat{\alpha}^2|\mathcal{T}}(x|t)$  and  $p_{\bar{\alpha}^2|\mathcal{T}}(x|t)$ :

$$p_{\hat{\alpha}^2|\mathcal{T}}(x|t) = \theta(x) p_{\bar{\alpha}^2|\mathcal{T}}(x|t) + \delta(x) \int_{-\infty}^0 dy p_{\bar{\alpha}^2|\mathcal{T}}(y|t). \quad (\text{B1})$$

Inserting this formula into Eqs. (4.4) and (4.5) gives

$$P_{\text{FA}}(\hat{\alpha}_*^2) = \begin{cases} \int_{\hat{\alpha}_*^2}^{\infty} dx p_{\bar{\alpha}^2|\mathcal{T}}(x|0) & \text{if } \hat{\alpha}_*^2 > 0, \\ 1 & \text{if } \hat{\alpha}_*^2 \leq 0, \end{cases} \quad (\text{B2})$$

$$P_{\text{FD}}(\hat{\alpha}_*^2) = \begin{cases} 1 - \int_{\hat{\alpha}_*^2}^{\infty} dx p_{\bar{\alpha}^2|T}(x|1) & \text{if } \hat{\alpha}_*^2 > 0, \\ 0 & \text{if } \hat{\alpha}_*^2 \leq 0. \end{cases} \quad (\text{B3})$$

$$\Delta(\bar{\alpha}_0^2) = \frac{\sigma_1 \sigma_2}{\sqrt{N}} \quad (\text{B5})$$

$$\langle \bar{\alpha}_1^2 \rangle = \xi \alpha^2 \quad (\text{B6})$$

In the large  $N$  limit, the distribution  $p_{\bar{\alpha}^2|T}(x|t)$  must be Gaussian by the central limit theorem, and therefore this distribution is characterized entirely by its mean  $\langle \bar{\alpha}_1^2 \rangle$  and variance  $[\Delta(\bar{\alpha}_1^2)]^2$ . From Eqs. (2.1), (3.3), (3.10), (3.11) and (3.19), these are given by

$$\langle \bar{\alpha}_0^2 \rangle = 0 \quad (\text{B4})$$

$$\Delta(\bar{\alpha}_1^2) = \sqrt{\frac{\xi \alpha^4 (3 - \xi) + \xi \alpha^2 (\sigma_1^2 + \sigma_2^2) + \sigma_1^2 \sigma_2^2}{N}}. \quad (\text{B7})$$

Substituting Gaussian distributions, with means and variances determined by Eqs. (B4)–(B7), into Eqs. (B2) and (B3) yields

$$P_{\text{FA}}(\hat{\alpha}_*^2, \sigma_1, \sigma_2, N) = \begin{cases} \frac{1}{2} \operatorname{erfc}\left(\frac{\hat{\alpha}_*^2}{\sigma_1 \sigma_2} \sqrt{\frac{N}{2}}\right) & \text{if } \hat{\alpha}_*^2 > 0, \\ 1 & \text{if } \hat{\alpha}_*^2 \leq 0, \end{cases} \quad (\text{B8})$$

$$P_{\text{FD}}(\hat{\alpha}_*^2, \xi, \alpha, \sigma_1, \sigma_2, N) = \begin{cases} 1 - \frac{1}{2} \operatorname{erfc}\left[(\hat{\alpha}_*^2 - \xi \alpha^2) \sqrt{\frac{N}{2[\xi \alpha^4 (3 - \xi) + \xi \alpha^2 (\sigma_1^2 + \sigma_2^2) + \sigma_1^2 \sigma_2^2]}}\right] & \text{if } \hat{\alpha}_*^2 > 0, \\ 0 & \text{if } \hat{\alpha}_*^2 \leq 0. \end{cases} \quad (\text{B9})$$

If we now eliminate  $\hat{\alpha}_*^2$  between Eqs. (B8) and (B9), change variables from  $\alpha$  to  $\rho$  using Eq. (3.27), and set  $\sigma_1 = \sigma_2$ , we obtain Eq. (4.6).

### APPENDIX C: ASYMPTOTIC BEHAVIOR OF MAXIMUM LIKELIHOOD STATISTIC

In this appendix we derive the large- $N$  behavior of the maximum likelihood statistic  $\Lambda_{\text{ML}}^{\text{NG}}$ . From Eq. (3.22), we can write the statistic in the form

$$\Lambda_{\text{ML}}^{\text{NG}}(h) = \exp[N\mathcal{L}(h)] \quad (\text{C1})$$

with

$$\mathcal{L}(h) = \max_{\sigma_1, \sigma_2, \xi, \alpha} g(\sigma_1, \sigma_2, \xi, \alpha, h) \quad (\text{C2})$$

where

$$g = \frac{1}{N} \sum_{k=1}^N g_k(\sigma_1, \sigma_2, \xi, \alpha), \quad (\text{C3})$$

and the function  $g_k = g_k(\sigma_1, \sigma_2, \xi, \alpha)$  is given by

$$e^{g_k} = \xi A_k(\alpha) + (1 - \xi) A_k(0) \quad (\text{C4})$$

with

$$A_k(\alpha) = \exp \left[ \frac{\left( \frac{h_1^k}{\sigma_1^2} + \frac{h_2^k}{\sigma_2^2} \right)^2}{2 \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} + \frac{1}{\alpha^2} \right)} - \frac{(h_1^k)^2}{2\sigma_1^2} - \frac{(h_2^k)^2}{2\sigma_2^2} + 1 \right] \times \frac{\bar{\sigma}_1 \bar{\sigma}_2}{\sqrt{\sigma_1^2 \sigma_2^2 + \sigma_1^2 \alpha^2 + \sigma_2^2 \alpha^2}}. \quad (\text{C5})$$

We denote by  $\tilde{\sigma}_1$ ,  $\tilde{\sigma}_2$ ,  $\tilde{\xi}$  and  $\tilde{\alpha}$  the “true” parameters governing the distribution of the quantities  $h_1^k$  and  $h_2^k$  according to Eqs. (2.1), (3.1), (3.3), and (3.19), with untilded quantities replaced by the corresponding tilded quantities. (These “true parameters” were denoted by  $\sigma_1$ ,  $\sigma_2$ ,  $\xi$  and  $\alpha$  in the body of the paper.) We define  $\tilde{\rho}$  to be the signal-to-noise ratio (3.27) with untilded quantities replaced by tilded quantities:

$$\tilde{\rho} \equiv \frac{\tilde{\xi} \tilde{\alpha}^2 \sqrt{N}}{\tilde{\sigma}_1 \tilde{\sigma}_2}. \quad (\text{C6})$$

For simplicity, in this appendix we restrict attention to the case  $\tilde{\sigma}_1 = \tilde{\sigma}_2$ . Then, without loss of generality, we can take  $\tilde{\sigma}_1 = \tilde{\sigma}_2 = 1$  by rescaling our units of strain amplitude.

We discuss separately the computation of the false alarm and false dismissal probabilities, as different techniques are required to compute each.

### 1. False dismissal probability

The false dismissal probability for the statistic (C2) will be some function

$$P_{\text{FD}} = P_{\text{FD}}(\mathcal{L}_*, N, \tilde{\xi}, \tilde{\rho}) \quad (\text{C7})$$

of the threshold  $\mathcal{L}_*$  on  $\mathcal{L}$ , the number of data points  $N$ , the Gaussianity parameter  $\tilde{\xi}$  and signal-to-noise ratio  $\tilde{\rho}$  of the signal. For applications to ground based detectors, we will have  $\tilde{\rho} \sim$  (a few), in order that the signal be detectable,  $N \sim 10^9$ , and  $10^{-3} \lesssim \tilde{\xi} \leq 1$ . Therefore it would be useful to find approximate analytic expressions for the false alarm probability in the limit of large  $N$ . There are actually several different, large  $N$  regimes in the three dimensional parameter space  $(N, \tilde{\xi}, \tilde{\rho})$  that one might explore:

(1) The limit  $N \rightarrow \infty$  with  $\tilde{\alpha}$  and  $\tilde{\xi}$  held fixed. This corresponds to fixing the stochastic background signal and going to a limit of long observation times. In this limit we have  $\tilde{\rho} \propto \sqrt{N}$  which diverges. This is not a very realistic limit to explore.

(2) The limit  $N \rightarrow \infty$  with  $\tilde{\rho}$  and  $\tilde{\xi}$  held fixed. In this limit, the signal-to-noise ratio is held fixed, and correspondingly the amplitude  $\tilde{\alpha}$  of the stochastic background signal goes to zero, from Eq. (C6). This would be the most natural limit to explore. However, in this limit the statistical error  $\Delta\tilde{\xi}$  in our measurement of the Gaussianity parameter would diverge, from Eq. (4.18), and therefore in this limit we do not expect to be able to compute analytically the value of the parameter  $\xi$  which achieves the maximum in Eq. (C2). The analytic approximation methods which we discuss below do not work in this regime. (In addition our Monte Carlo simulations show that the maximum likelihood statistic itself does not perform any better than the cross-correlation statistic in this regime, as discussed in the Introduction.)

(3) The limit we actually explore is the limit  $N \rightarrow \infty$  with  $\tilde{\xi}$  fixed and  $\tilde{\rho}$  scaling  $\propto N^{1/4}$ , corresponding to  $\tilde{\alpha} \propto N^{-1/8}$ . The reason for our choosing to explore this particular limit is simply that it is amenable to analytic computations. Fractional corrections to our analytic results should scale like  $1/N$  or as  $1/\tilde{\rho}^4$ . Since  $\tilde{\rho} \sim$  (a few) at the threshold for detection, the approximation should be good to 10–20% or so.

We now turn to a discussion of the computational technique. We write

$$\tilde{\alpha} = \tilde{\alpha}_0 N^{-1/8}, \quad (\text{C8})$$

where  $\tilde{\alpha}_0$  is independent of  $N$ . Correspondingly, from Eq. (3.19) we can write

$$s^k = N^{-1/8} \hat{s}^k, \quad (\text{C9})$$

where the distribution of  $\hat{s}^k$  is given by Eq. (3.19) with  $\xi$  replaced by  $\tilde{\xi}$  and  $\alpha$  replaced by  $\tilde{\alpha}_0$ . In particular, the distribution of  $\hat{s}^k$  is independent of  $N$ . In computing the maximum over  $(\xi, \alpha, \sigma_1, \sigma_2)$  in Eq. (C2), it is useful change variables from  $\alpha$  to  $\kappa$  defined by

$$\kappa = \rho N^{-1/4} = \frac{\xi \alpha^2 N^{1/4}}{\sigma_1 \sigma_2}, \quad (\text{C10})$$

which we expect to be independent of  $N$  to leading order in the large  $N$  limit. The value of the variable  $\kappa$  that characterizes the signal is

$$\tilde{\kappa} = \tilde{\rho} N^{-1/4} = \frac{\tilde{\xi} \tilde{\alpha}_0^2}{\tilde{\sigma}_1 \tilde{\sigma}_2}, \quad (\text{C11})$$

cf. Eqs. (C8) and (C10).

We now consider fixed realizations of the infinite sequences of random variables  $n_1^k$ ,  $n_2^k$  and  $\hat{s}^k$ , and  $1 \leq k < \infty$ , and examine the limiting behavior of  $\mathcal{L}(h)$  as  $N \rightarrow \infty$ . We compute this limiting behavior by substituting into the right hand side of Eq. (C1) the relations

$$h_1^k = n_1^k + N^{-1/8} \hat{s}^k, \quad h_2^k = n_2^k + N^{-1/8} \hat{s}^k, \quad (\text{C12})$$

writing  $\alpha$  in terms of  $\kappa$  using Eq. (C10), and expanding in powers of  $N^{-1/8}$ . The result is an expression which can be written in terms of the sums  $Q_{abc}$  defined by

$$Q_{abc} = \frac{1}{N} \sum_{k=1}^N (\hat{s}^k)^a (n_1^k)^b (n_2^k)^c, \quad (\text{C13})$$

where  $a$ ,  $b$ , and  $c$  are non-negative integers. From the central limit theorem we can write

$$Q_{abc} = \mu_{abc} + \frac{1}{\sqrt{N}} \Delta_{abc}, \quad (\text{C14})$$

where  $\mu_{abc} = \langle Q_{abc} \rangle$  are computable functions of  $\tilde{\xi}$  and  $\tilde{\alpha}$ , and where the random variables  $(\Delta_{100}, \Delta_{010}, \dots)$  converge in distribution<sup>9</sup> as  $N \rightarrow \infty$  to a multivariate Gaussian of zero mean whose variance-covariance matrix is independent of  $N$ . Thus, in particular the joint distribution of all  $\Delta_{abc}$ 's is  $N$ -independent in limit that  $N \rightarrow \infty$ .

We define the vector

$$\mathbf{v} = (v^1, v^2, v^3, v^4) = (\xi, \kappa, \sigma_1^2, \sigma_2^2). \quad (\text{C15})$$

We denote the value of  $\mathbf{v}$  that achieves the maximum in Eq. (C2) by  $\hat{\mathbf{v}}$ :

$$g(\hat{\mathbf{v}}) = \max_{\mathbf{v}} g(\mathbf{v}), \quad (\text{C16})$$

<sup>9</sup>See Chap. 8 of Ref. [29] for definitions of different notions of convergence for sequences of random variables.

where  $\hat{\mathbf{v}} = (\hat{\xi}, \hat{\kappa}, \hat{\sigma}_1^2, \hat{\sigma}_2^2)$ . These estimators satisfy a system of four equations<sup>10</sup>

$$\left. \frac{\partial g}{\partial v^l} \right|_{\mathbf{v}=\hat{\mathbf{v}}} = 0. \quad (\text{C17})$$

We solve Eq. (C17) perturbatively. First assume that the estimators can be expanded in the form

$$\hat{v}^l = \sum_{j=0}^{\infty} \hat{v}^{l[j]} \epsilon^j, \quad (\text{C18})$$

where for ease of notation we have defined  $\epsilon = N^{-1/8}$ . We define the expansion coefficients  $v^{l[j]}$  analogously by an expansion of the form (C18) but without the hats. Now using Eq. (C14) the function  $g$  can be expanded as a power series in  $\epsilon$  whose coefficients are functions of  $v^{l[k]}$ ,  $\mu_{abc}$ , and  $\Delta_{abc}$ :

$$g(\mathbf{v}) = \sum_{j=0}^{\infty} g^{[j]}[v^{l[k]}, \mu_{abc}, \Delta_{abc}] \epsilon^j. \quad (\text{C19})$$

Substituting the expansions (C18) and (C19) into the condition (C17) for a local extremum gives an infinite set of equations which must collectively be satisfied by the coefficients  $\hat{v}^{l[j]}$

$$\left. \frac{\partial g^{[j]}}{\partial v^{l[k]}} \right|_{v^{m[n]} = \hat{v}^{m[n]}} = 0. \quad (\text{C20})$$

We solve these equations order by order to determine the coefficients  $\hat{v}^{l[j]}$ , and thereby justify *a posteriori* the ansatz (C18).

We find that in order to compute the leading order expression for  $\mathcal{L}$ , we must obtain the expansion for  $\hat{\xi}$  to zeroth order in  $\epsilon$ , the expansion for  $\hat{\kappa}$  to fourth order in  $\epsilon$ , and the expansions of  $\hat{\sigma}_1^2$  and  $\hat{\sigma}_2^2$  to sixth order in  $\epsilon$ . The leading order results are

$$\hat{\kappa} = \tilde{\kappa} + \epsilon^2 X + O(\epsilon^3), \quad (\text{C21})$$

$$\frac{1}{\hat{\xi}} = \frac{1}{\tilde{\xi}} + \frac{Y}{\sqrt{6}\tilde{\kappa}^2} + O(\epsilon), \quad (\text{C22})$$

$$\hat{\sigma}_1^2 = 1 + O(\epsilon^2), \quad (\text{C23})$$

$$\hat{\sigma}_2^2 = 1 + O(\epsilon^2), \quad (\text{C24})$$

where

<sup>10</sup>Here we are assuming that the maximum is achieved as a local maximum in the interior of the 4 dimensional parameter space. Cases when the maximum is achieved on the boundary are discussed below.

$$X = \Delta_{011} \quad (\text{C25})$$

and

$$Y = \frac{1}{8\sqrt{6}} [4(\Delta_{031} + \Delta_{013}) - 12(\Delta_{002} + \Delta_{020}) - 24\Delta_{011} + \Delta_{040} + \Delta_{004} + 6\Delta_{022}]. \quad (\text{C26})$$

Using Eqs. (3.1), (3.19), (C13) and (C14) one can show that the random variables  $X$  and  $Y$  are independent Gaussian random variables of zero mean and unit variance.

In deriving Eqs. (C21)–(C24) we assumed that the value of  $\mathbf{v}$  which achieves the maximum in Eq. (C2) corresponds to a local maximum. However, if the right hand side of Eq. (C21) is negative, the maximum will instead be achieved on the boundary of the parameter space at  $\hat{\kappa} = 0$ , since the variable  $\kappa$  must be non-negative. Similarly, if the right hand side of Eq. (C22) is less than 1, the maximum will be achieved at  $\hat{\xi} = 1$ , since  $1/\xi$  must lie in the interval  $[1, \infty)$ .

Substituting the results (C21)–(C24) (together with the higher order corrections to those results which we have not shown) into the expansion for the statistic  $\mathcal{L}$ , and taking into account the various special cases discussed in the last paragraph, gives

$$\begin{aligned} \mathcal{L} = & \left[ \frac{1}{2} (Y + \sqrt{6}q\tilde{\kappa}^2)^2 \epsilon^8 \theta(Y + \sqrt{6}q\tilde{\kappa}^2) + \frac{1}{2} (\tilde{\kappa} + \epsilon^2 X)^2 \epsilon^4 \right. \\ & \left. - \tilde{\kappa}^3 \epsilon^6 + \frac{7}{4} \tilde{\kappa}^4 \epsilon^8 + \tilde{\kappa} U \epsilon^7 + \tilde{\kappa} V \epsilon^8 \right] \theta(\tilde{\kappa} + \epsilon^2 X) + O(\epsilon^9). \end{aligned} \quad (\text{C27})$$

Here  $\theta(x)$  is the step function and

$$q = \frac{1}{\tilde{\xi}} - 1, \quad (\text{C28})$$

$$U = \Delta_{101} + \Delta_{110}, \quad (\text{C29})$$

$$V = \Delta_{200} - \frac{1}{2} \tilde{\kappa} (\Delta_{002} + \Delta_{020}) - 2\tilde{\kappa} \Delta_{011}. \quad (\text{C30})$$

We note that the corresponding expression for the statistic  $(\ln \Lambda_{\text{ML}}^{\text{G}})/N$  [which is equivalent to the cross-correlation statistic by Eq. (3.13)] is given by Eq. (C27) with the first term in the square brackets dropped.

Next we drop all the terms in the square bracket in Eq. (C27) other than the first two terms. The reason is that these terms will give corrections that are smaller than the terms retained (both in expected value and in fluctuations) by a factor of

$$\tilde{\kappa} \epsilon^2 = \frac{\tilde{\rho}}{\sqrt{N}}, \quad (\text{C31})$$

which will be small compared to unity for all cases we are interested in. This gives for the false dismissal probability the expression

$$P_{\text{FD}} = P(\mathcal{L} < \mathcal{L}_*) = \int_{\mathcal{R}} \frac{dx dy}{2\pi} \exp\left[-\frac{(x-x_0)^2}{2} - \frac{(y-y_0)^2}{2}\right], \quad (\text{C32})$$

where

$$x_0 = \tilde{\kappa}/\epsilon^2 \quad (\text{C33})$$

$$y_0 = \sqrt{6}q\tilde{\kappa}^2 \quad (\text{C34})$$

$$r_0 = \sqrt{2N\mathcal{L}_*}. \quad (\text{C35})$$

Here the region  $\mathcal{R}$  in the  $x, y$  plane is the union of the two regions

$$\begin{aligned} x &\geq 0 \\ y &\geq 0 \\ x^2 + y^2 &\leq r_0^2 \end{aligned} \quad (\text{C36})$$

and

$$\begin{aligned} y &\leq 0 \\ 0 &\leq x \leq r_0. \end{aligned} \quad (\text{C37})$$

The integral over the region (C37) is

$$P_{\text{FD}}^{(1)} = \mathcal{P}(-y_0)[\mathcal{P}(r_0 - x_0) - \mathcal{P}(-x_0)], \quad (\text{C38})$$

where

$$\mathcal{P}(x) \equiv 1 - \frac{1}{2} \operatorname{erfc}(x/\sqrt{2}) = \int_{-\infty}^x dt \frac{1}{\sqrt{2\pi}} \exp[-t^2/2]. \quad (\text{C39})$$

The integral over the region (C36) can be written as

$$P_{\text{FD}}^{(2)} = \frac{1}{2\pi} \int_0^{\pi/2} d\theta \int_0^{r_0} dr r \exp\left[-\frac{1}{2}(r \cos \theta - x_0)^2 - \frac{1}{2}(r \sin \theta - y_0)^2\right]. \quad (\text{C40})$$

The integrand in Eq. (C40) peaks at  $r \cos \theta = x_0$ ,  $r \sin \theta = y_0$ . In order for  $P_{\text{FD}}$  to be small, it is necessary that this peak occurs outside the domain of integration, at  $r > r_0$ . So we must have

$$x_0^2 + y_0^2 \geq r_0^2. \quad (\text{C41})$$

The criterion  $x_0 \geq r_0$  is, in order of magnitude, just the usual criterion for detectability with the cross-correlation statistic. The criterion  $y_0 \geq r_0$  reduces to, in order of magnitude,

$$\xi \lesssim \frac{\rho^2}{\sqrt{N}} \quad (\text{C42})$$

which is what we claimed earlier to be the regime where the maximum likelihood statistic starts to work well, cf. Sec. IV A 3 above.

Evaluating the integral (C40) using the Laplace approximation gives

$$P_{\text{FD}}^{(2)} = \frac{1}{r_0(\lambda-1)\sqrt{2\pi\lambda}} \exp\left[-\frac{1}{2}r_0^2(\lambda-1)^2\right] \left[1 + \mathcal{O}\left(\frac{1}{r_0}\right)\right], \quad (\text{C43})$$

where we define the variables  $\lambda$  and  $\gamma$  by

$$(x_0, y_0) = r_0\lambda(\cos \gamma, \sin \gamma). \quad (\text{C44})$$

However, the result (C43) is not very accurate for small  $r_0$ . Alternatively we can integrate over  $r$  in Eq. (C40) to obtain

$$\begin{aligned} P_{\text{FD}}^{(2)} = \int_0^{\pi/2} d\theta \left\{ \frac{1}{2\pi} e^{r_0^2(1+\lambda^2)/2} [e^{r_0^2/2} - e^{r_0^2\lambda \cos(\gamma-\theta)}] + \frac{r_0\lambda}{2\sqrt{2\pi}} e^{(r_0^2\lambda^2/4)[\cos(2\{\gamma-\theta\})-1]} \cos(\gamma-\theta) \right. \\ \left. \left[ \operatorname{erf}\left(\frac{r_0\lambda \cos\{\gamma-\theta\}}{\sqrt{2}}\right) + \operatorname{erf}\left(\frac{r_0\{1-\lambda \cos[\gamma-\theta]\}}{\sqrt{2}}\right) \right] \right\}, \end{aligned} \quad (\text{C45})$$

where

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x dy e^{-y^2}. \quad (\text{C46})$$

The integral (C45) can be evaluated numerically. The false dismissal probability is then given by

$$P_{\text{FD}} = P_{\text{FD}}^{(1)} + P_{\text{FD}}^{(2)}, \quad (\text{C47})$$

with  $P_{\text{FD}}^{(1)}$  given by Eq. (C38) and  $P_{\text{FD}}^{(2)}$  given by Eq. (C45).

## 2. False alarm probability

The false alarm probability is some function

$$P_{\text{FA}} = P_{\text{FA}}(\mathcal{L}_*, N) \quad (\text{C48})$$

of the threshold  $\mathcal{L}_*$  value of the detection statistic (C2) and of the number of data points  $N$ . It does not depend on the signal parameters  $\tilde{\rho}$  and  $\tilde{\xi}$  because no signal is present. We would like to evaluate this quantity in the large  $N$  limit.

We start by rewriting the statistic (C2) in the form

$$\mathcal{L} = \max_{\mathbf{v}} \left\{ \frac{1}{N} \sum_{k=1}^N \ln A_k(0) + \frac{1}{N} \sum_{k=1}^N \ln [1 + \xi \mathcal{D}_k(\alpha)] \right\}, \quad (\text{C49})$$

where

$$\mathcal{D}_k(\alpha) = \frac{A_k(\alpha)}{A_k(0)} - 1. \quad (\text{C50})$$

Consider first the first term in Eq. (C49). Using the definition (C5) of  $A_k(\alpha)$  and the definition (1.4) of  $\bar{\sigma}_1$  and  $\bar{\sigma}_2$  we can write this term as

$$\frac{1}{N} \sum_{k=1}^N \ln A_k(0) = -\frac{\Delta\sigma_1^2}{\bar{\sigma}_1^2} - \frac{\Delta\sigma_2^2}{\bar{\sigma}_2^2} + O(\Delta\sigma_1^3, \Delta\sigma_2^3), \quad (\text{C51})$$

where  $\Delta\sigma_1 = \sigma_1 - \bar{\sigma}_1$ ,  $\Delta\sigma_2 = \sigma_2 - \bar{\sigma}_2$ . Therefore the first term is maximized at  $\sigma_1 = \bar{\sigma}_1$ ,  $\sigma_2 = \bar{\sigma}_2$ . Below we shall show that the second term in Eq. (C49) is of order  $O(\epsilon^2)$ , where in this subsection we define  $\epsilon = 1/\sqrt{N}$ . Therefore the values of  $\sigma_1$  and  $\sigma_2$  that achieve the maximum are

$$\begin{aligned} \hat{\sigma}_1 &= \bar{\sigma}_1 [1 + O(\epsilon^2)] \\ \hat{\sigma}_2 &= \bar{\sigma}_2 [1 + O(\epsilon^2)]. \end{aligned} \quad (\text{C52})$$

Moreover, in analyzing the second term it suffices to take  $\sigma_1 = \bar{\sigma}_1$ ,  $\sigma_2 = \bar{\sigma}_2$  in order to obtain the statistic to the leading  $O(\epsilon^2)$  order. Lastly, since we have assumed that  $\bar{\sigma}_1 = \bar{\sigma}_2 = 1$  and no signal is present, we have  $\bar{\sigma}_{1,2} = 1 + O(\epsilon)$ . Hence, in analyzing the second term, it is sufficient to take  $\sigma_1 = \sigma_2 = 1$ .

The statistic (C49) therefore reduces to

$$\mathcal{L} = \max_{\alpha, \xi} \frac{1}{N} \sum_{k=1}^N \ln [1 + \xi \mathcal{D}_k(\alpha)] + O(\epsilon), \quad (\text{C53})$$

where from Eqs. (C5) and (C50)

$$\mathcal{D}_k(\alpha) = \frac{1}{\sqrt{1+2\alpha}} \exp \left[ \frac{w_k^2}{2 + \frac{1}{\alpha}} \right] - 1. \quad (\text{C54})$$

Here  $w_k = (n_1^k + n_2^k)/\sqrt{2}$ ,  $1 \leq k \leq N$ , are independent Gaussian random variables of zero mean and unit variance.

It is straightforward to numerically compute the distribution of the statistic (C53), by generating the Gaussian variables  $w_k$  and numerically maximizing over  $\xi$  and  $\alpha$ . The result is shown in Fig. 12. We find that at large  $N$ , the distri-

## Cumulative distribution function

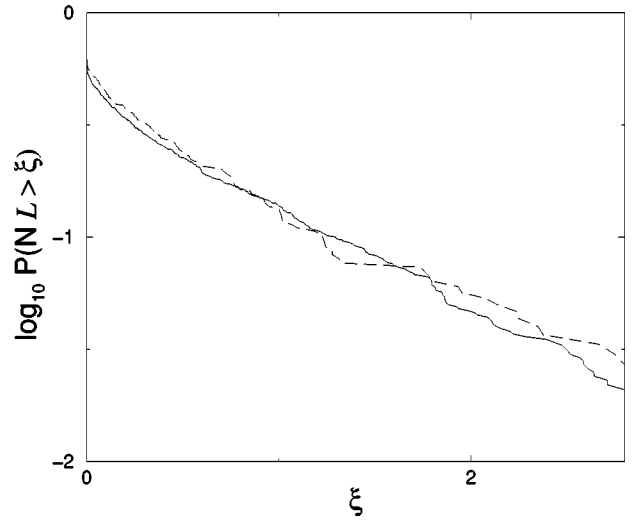


FIG. 12. The cumulative distribution function for the leading order expression (C53) for the statistic when no signal is present, obtained numerically. The solid line is for  $N=1000$ , and the dashed line for  $N=5000$ .

bution of  $N\mathcal{L}$  becomes independent of  $N$ , and is approximately given by

$$P(N\mathcal{L} > \xi) = \alpha_0 e^{-\beta_0 \xi} \quad (\text{C55})$$

for  $\xi > 0$ , where  $\alpha_0 \approx 0.42$  and  $\beta_0 \approx 1.08$ . Therefore the false alarm probability is approximately given by

$$P_{\text{FA}} = \alpha_0 \exp[-\beta_0 N\mathcal{L}_*]. \quad (\text{C56})$$

Finally, we remark why it is plausible to expect the distribution of  $N\mathcal{L}$  to be independent of  $N$  in the large  $N$  limit. The numerical maximizations over  $\xi$  and  $\alpha$  in Eq. (C53) show that the maximum is nearly always achieved at  $\alpha \ll 1$  or  $\xi \ll 1$ . In both these regimes, one can obtain some information about the  $N$  dependence of the statistic.

Consider first the regime  $\xi \ll 1$ . In this regime we can expand the expression (C53) as a power series in  $\xi$  to obtain

$$\mathcal{L} = \max_{\alpha, \xi} \frac{1}{N} \sum_{k=1}^N \left[ \xi \mathcal{D}_k(\alpha) - \frac{1}{2} \xi^2 \mathcal{D}_k(\alpha)^2 + O(\xi^3) \right] + O(\epsilon). \quad (\text{C57})$$

The generalized central limit theorem (reviewed in Appendix D) implies that

$$\frac{1}{N} \sum_{k=1}^N \mathcal{D}_k(\alpha) = N^{(1-\gamma_1)/\gamma_1} (\ln N)^{\delta_1} \mathcal{F}_N(\alpha), \quad (\text{C58})$$

where for each fixed  $\alpha$ , the distribution of the random variable  $\mathcal{F}_N(\alpha)$  becomes independent of  $N$  in the large  $N$  limit. Here

$$\gamma_1 = \begin{cases} 2 & 0 < \alpha \leq 1/2, \\ 1 + \frac{1}{2\alpha} & 1/2 \leq \alpha \end{cases} \quad (\text{C59})$$

and

$$\delta_1 = \begin{cases} 0 & 0 < \alpha \leq 1/2, \\ \frac{-\alpha}{1+2\alpha} & 1/2 \leq \alpha. \end{cases} \quad (\text{C60})$$

The limiting distribution is a Levy distribution with parameters  $p=1$  and  $\gamma=\gamma_1$ . Similarly we have

$$\frac{1}{N} \sum_{k=1}^N \mathcal{D}_k(\alpha)^2 = N^{(1-\gamma_2)/\gamma_2} (\ln N)^{\delta_2} \mathcal{G}_N(\alpha), \quad (\text{C61})$$

where as  $N \rightarrow \infty$  at each fixed  $\alpha$  the distribution of the random variable  $\mathcal{G}_N(\alpha)$  tends to a Levy distribution with parameters  $p=1$  and  $\gamma=\gamma_2$ , with

$$\gamma_2 = \begin{cases} 2 & 0 < \alpha \leq 1/6, \\ \frac{1+2\alpha}{4\alpha} & 1/6 \leq \alpha \end{cases} \quad (\text{C62})$$

and

$$\delta_2 = \begin{cases} 0 & 0 < \alpha \leq 1/6, \\ \frac{-2\alpha}{1+2\alpha} & 1/6 \leq \alpha. \end{cases} \quad (\text{C63})$$

We now substitute the results (C58) and (C61) into the expression (C57) for the statistic, and maximize analytically over the quadratic dependence on  $\xi$ . For  $\alpha \geq 1/2$ , the value of  $\xi$  which achieves the maximum goes to zero as  $N \rightarrow \infty$ , consistent with the assumption  $\xi \ll 1$ , and the result is<sup>11</sup>

$$N\mathcal{L} = \frac{1}{2} \max_{\alpha} \frac{\mathcal{F}_N(\alpha)^2}{\mathcal{G}_N(\alpha)} + O(\epsilon). \quad (\text{C64})$$

In the regime  $\alpha \ll 1$ , if we expand the expression (C53) to quadratic order in  $\alpha$ , the result is an expression which is a linear function of  $1/\xi$  at fixed  $\alpha\xi$ . Hence, when one maximizes over values of  $\xi$  in the range  $0 \leq \xi \leq 1$ , the maximum is always achieved either at  $\xi=0$  or  $\xi=1$ . One can show that the maximum to this order is always achieved at  $\xi=1$ , and the resulting expression is

$$N\mathcal{L} = \frac{1}{4} \mathcal{G}^2 + O(\epsilon), \quad (\text{C65})$$

where

$$\mathcal{G} = \sqrt{N} \left[ \frac{1}{N} \sum_{k=1}^N w_k^2 - 1 \right] \quad (\text{C66})$$

<sup>11</sup>For  $\alpha < 1/2$  this argument fails, which is why we must numerically verify that the distribution of  $N\mathcal{L}$  is asymptotically independent of  $N$ .

has a distribution that is independent of  $N$  in the large  $N$  limit.

#### APPENDIX D: GENERALIZED CENTRAL LIMIT THEOREM

In this appendix we review the generalized central limit theorem that can be found on p. 574 of Ref. [37]. First we define a particular distribution function called the Levy distribution. It depends on 3 real parameters, a positive constant  $C$ , a parameter  $\gamma$  in the range  $0 < \gamma \leq 2$ , and constant  $p$  in the range  $0 \leq p \leq 1$ .<sup>12</sup> We say a random variable  $X$  has a Levy distribution with parameters  $C$ ,  $\gamma$  and  $p$  if the characteristic function of  $X$  is given by

$$\langle e^{i\xi X} \rangle = \exp \left\{ |\xi|^\gamma \frac{C\Gamma(3-\gamma)}{\gamma(\gamma-1)} [\cos(\pi\gamma/2) + i \operatorname{sgn}(\xi)(p-q)\sin(\pi\gamma/2)] \right\}, \quad (\text{D1})$$

where  $q=1-p$ . The corresponding probability distribution function is obtained by taking a Fourier transform and decays like  $x^{-(1+\gamma)}$  at large  $x$  for  $\gamma < 2$  ( $\gamma=2$  is the Gaussian case).

Consider now a random variable  $X$  with probability distribution function  $f(x)$  whose variance is infinite. Let

$$F(x) = \int_{-\infty}^x dy f(y) \quad (\text{D2})$$

be the cumulative distribution function and define

$$\mu(x) = \int_{-x}^x dy y^2 f(y). \quad (\text{D3})$$

Suppose that the distribution satisfies the following conditions: (i) As  $x \rightarrow \infty$  we have  $\mu(x) \sim x^{2-\gamma} L(x)$ , where  $0 < \gamma \leq 2$ , and  $L(x)$  varies slowly in the sense that  $L(tx)/L(x) \rightarrow 1$  as  $t \rightarrow \infty$  for all  $x > 0$ . (ii) We have

$$\frac{1-F(x)}{F(-x)+1-F(x)} \rightarrow p, \quad \frac{F(-x)}{F(-x)+1-F(x)} \rightarrow q \quad (\text{D4})$$

as  $x \rightarrow \infty$ , where  $0 \leq p \leq 1$ ,  $0 \leq q \leq 1$  and  $p+q=1$ . (iii) For  $1 < \gamma \leq 2$ , we assume that the expected value  $\int dx x f(x)$  vanishes; this can be enforced by making a transformation of the form  $X \rightarrow X + \text{constant}$ .

<sup>12</sup>The parameter  $\gamma$  is conventionally denoted by  $\alpha$ . We use  $\gamma$  here to avoid confusion with the variable  $\alpha$  defined in Eq. (1.7).



We define the sequence of random variables

$$S_N = \frac{1}{a_N} \sum_{i=1}^N X_i, \quad (\text{D5})$$

where the  $X_i$  are independent, identically distributed random variables with distribution function  $f$ , and the constants  $a_N$  are chosen to satisfy

$$\frac{N\mu(a_N)}{a_N^2} \rightarrow C \quad (\text{D6})$$

as  $N \rightarrow \infty$ , where  $C$  is a positive constant. Then, the distribution functions of the random variables  $S_N$  converge to a Levy distribution with parameters  $C$ ,  $\gamma$  and  $p$  as  $N \rightarrow \infty$ .

- 
- [1] A. Abramovici *et al.*, *Science* **256**, 325 (1992).  
 [2] C. Bradaschia *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **289**, 518 (1990).  
 [3] R. Schilling, in *Laser Interferometer Space Antenna*, edited by William M. Folkner, AIP Conf. Proc. No. 456 (AIP, Woodbury, NY, 1998), p. 217.  
 [4] M.K. Fujimoto, *Journal of the Communications Research Laboratory* **46**, 437 (1999).  
 [5] B. Allen, in *Relativistic Gravitation and Gravitational Radiation*, Proceedings of the Les Houches School of Physics, Les Houches, 1995, edited by J. A. Marck and J. P. Lasota (CNRS, Observatoire de Paris, Meudon, 1997), p. 373.  
 [6] D. Blair and L. Ju, *Mon. Not. R. Astron. Soc.* **283**, 648 (1996).  
 [7] V. Ferrari, S. Matarrese, and R. Schneider, *Mon. Not. R. Astron. Soc.* **303**, 247 (1999).  
 [8] R. Schneider *et al.*, *Mon. Not. R. Astron. Soc.* **317**, 385 (2000).  
 [9] V. Ferrari, S. Matarrese, and R. Schneider, *Mon. Not. R. Astron. Soc.* **303**, 258 (1999).  
 [10] T. Regimbau and J.A. de Freitas Pacheco, astro-ph/0105260.  
 [11] T. Damour and A. Vilenkin, *Phys. Rev. D* **64**, 064008 (2002).  
 [12] M. Kamionkowski, A. Kosowsky, and M.S. Turner, *Phys. Rev. D* **49**, 2837 (1994).  
 [13] L.P. Grishchuk, *Zh. Éksp. Teor. Fiz.* **67**, 825 (1974) [*Sov. Phys. JETP* **40**, 409 (1975)]; E.W. Kolb and M.S. Turner, *The Early Universe* (Addison-Wesley, Redwood, CA, 1990), and references therein.  
 [14] R. Schneider, V. Ferrari, S. Matarrese, and S.F. Portegies Zwart, *Mon. Not. R. Astron. Soc.* **342**, 797 (2001).  
 [15] D.M. Coward, R.R. Burman, and D.G. Blair, *Mon. Not. R. Astron. Soc.* **324**, 1015 (2001).  
 [16] C.J. Hogan, *Phys. Rev. D* **62**, 121302(R) (2000).  
 [17] P.F. Michelson, *Mon. Not. R. Astron. Soc.* **227**, 933 (1987).  
 [18] N. Christensen, *Phys. Rev. D* **46**, 5250 (1992).  
 [19] É.É. Flanagan, *Phys. Rev. D* **48**, 2389 (1993).  
 [20] B. Allen and J.D. Romano, *Phys. Rev. D* **59**, 102001 (1999). Note that the criticism in this paper of the upper limit formula given in Eq. (6.5) of Ref. [19] is incorrect as it misinterprets that formula as a frequentist upper limit rather than a Bayesian upper limit.  
 [21] B. Allen, J.D.E. Creighton, É.É. Flanagan, and J.D. Romano, *Phys. Rev. D* **65**, 122002 (2002).  
 [22] B. Allen, J.D.E. Creighton, É.É. Flanagan, and J.D. Romano, gr-qc/0205015.  
 [23] S. Klimentko and G. Mitselmakher, LIGO Technical Report LIGO-T010125-00-D, 2001; gr-qc/0208007.  
 [24] L.S. Finn, *Phys. Rev. D* **46**, 5236 (1992).  
 [25] W.G. Anderson, P.R. Brady, J.D.E. Creighton, and É.É. Flanagan, *Phys. Rev. D* **63**, 042003 (2001).  
 [26] L.S. Finn and J.D. Romano (in preparation).  
 [27] L.S. Finn (in preparation).  
 [28] P.J. Bickel and K.A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics* (Holden-Day, San Francisco, CA, 1977), Sec. 6.4.  
 [29] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed., edited by S.W. Director (McGraw-Hill, New York, 1984).  
 [30] J. Neyman and K. Pearson, *Philos. Trans. R. Soc. London* **231**, 289 (1933).  
 [31] T.S. Ferguson, *Mathematical Statistics, a Decision Theoretic Approach*, edited by Z.W. Birnbaum and E. Kukacs (Academic, New York, 1967).  
 [32] T. Bayes and R. Price, *Philos. Trans. R. Soc. London* **53**, 370 (1763).  
 [33] T.J. Loredo, in *Astronomical Data Analysis Software and Systems VIII*, edited by D.M. Mehringer, R.L. Plante, and D.A. Roberts (ASP, San Francisco, CA, 1999), Vol. 172, p. 297.  
 [34] T. Zwerger and E. Müller, *Astron. Astrophys.* **320**, 209 (1997).  
 [35] H. Dimmelmeier, J.A. Font, and E. Müller, *Astron. Astrophys.* **393**, 523 (2002).  
 [36] S. Drasco and É.É. Flanagan, in *Detecting a Non-Gaussian Stochastic Background of Gravitational Radiation*, Proceedings of the Ninth Marcel Grossmann Meeting on General Relativity, edited by V. Gurzadyan, R.T. Jantzen, and R. Ruffini (World Scientific, Singapore, 2001) (also gr-qc/0101051).  
 [37] W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1971), Vol. II.