Robust statistics for deterministic and stochastic gravitational waves in non-Gaussian noise: Frequentist analyses

Bruce Allen and Jolien D. E. Creighton

Department of Physics, University of Wisconsin–Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201

Éanna É. Flanagan

Newman Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14853-5001

Joseph D. Romano

Department of Physical Sciences, University of Texas at Brownsville, Brownsville, Texas 78520

Gravitational wave detectors will need optimal signal-processing algorithms to extract weak signals from the detector noise. Most algorithms designed to date are based on the unrealistic assumption that the detector noise may be modeled as a stationary Gaussian process. However most experiments exhibit a non-Gaussian "tail" in the probability distribution. This "excess" of large signals can be a troublesome source of false alarms. This article derives an optimal (in the Neyman-Pearson sense, for weak signals) signal processing strategy when the detector noise is non-Gaussian and exhibits tail terms. This strategy is robust, meaning that it is close to optimal for Gaussian noise but far less sensitive than conventional methods to the excess large events that form the tail of the distribution. The method is analyzed for two different signal analysis problems: (i) a known waveform (e.g., a binary inspiral chirp) and (ii) a stochastic background, which requires a multi-detector signal processing algorithm. The methods should be easy to implement: they amount to truncation or clipping of sample values which lie in the outlier part of the probability distribution.

I. INTRODUCTION

The construction of several new detectors of gravitational radiation is currently approaching completion. These instruments are of a different design and have significantly better sensitivity and broader bandwidth than previous detectors. They include the LIGO detector being built in the United States by a joint Caltech-MIT Collaboration [1,2], the VIRGO detector being built near Pisa by an Italian-French Collaboration [3], the GEO-600 detector being built in Hannover by an Anglo-German Collaboration [4], and the TAMA-300 detector near Tokyo [5]. There are also several resonant bar detectors currently in operation [6], and several more refined bar and interferometric detectors presently in the planning and proposal stages [7]. These instruments search for very weak signals. For the most likely sources, the signals will be buried in the noise of the detectors, and need to be extracted with sophisticated optimal signal-processing strategies [8].

The standard assumption made in the literature is that the detector noise has multivariate Gaussian statistics. This assumption is certainly incorrect: every sensitive gravitational wave detector operated to date has been characterized by noise that is both non-stationary and non-Gaussian. Some experimentation has shown that this is a serious matter [9]: existing detection strategies for both deterministic and stochastic signals do not perform nearly as well when non-Gaussian noise is present. Roughly speaking, if the non-Gaussian fluctuations are large, they bias the statistics and make it more difficult to achieve a given level of statistical confidence.

In this paper, we develop a new set of statistical signal-processing techniques to search for deterministic and stochastic gravitational waves. These techniques are *robust*, meaning that they will work well even if the detector noise is not Gaussian but falls into a broader statistical class that we expect includes realistic detectors. In large part, these new methods are similar to the older ones: one constructs matched filters to search for known waveforms or cross-correlates the instrument outputs at the different detector sites to search for a stochastic background. The essential difference is that by using locally optimal methods [10] these statistical measures are modified. The effect is to *truncate* the statistics: detector samples that fall outside the central Gaussian-like part of the sample distribution (i.e., the outliers) are excluded from (or saturated when constructing) the measurement statistic. For both cases, a robust statistic is found which performs better than the optimal linear filter in the case where the detector noise is non-Gaussian, and almost as well in the Gaussian-noise case.

The paper is organized as follows. In Sec. II we derive the locally most powerful signal-processing tests for deterministic signals. We begin in Sec. II A with a derivation of the Neyman-Pearson criteria for optimality, in the case where a known waveform is hidden in white noise. We define the power function of a test and derive a criteria for the locally optimal test in the weak-signal regime. The locally optimal test is analyzed for a number of different types of non-Gaussian noise, and we show that the locally optimal decision statistic is a matched filter where the non-Gaussian sample values are truncated or excluded. In Sec. II B the results are generalized to the case where a known waveform is hidden in colored noise, and we introduce models for non-Gaussian colored noise. In Sec. III, we turn to the detection

of a stochastic background. Section III A considers the case of a stochastic signal (i.e., where the waveform is not known) and derives the locally optimal statistic which can be used to correlate two identical detectors, where we assume that each detector has independent white noise and is co-aligned and coincident. In Sec. III B, these results are generalized to the case where the noise is colored, and the detectors are in different locations, and not aligned in parallel. In Sec. IV, we discuss an implementation of these statistics, and we illustrate how one can compare the performance of different statistics using Monte Carlo simulations. Section V contains a short conclusion and summary.

II. DETERMINISTIC SIGNALS

A. Single detector, white noise

In order to describe the idea in a simple way we first discuss the case where we are searching for a known signal in the data stream of a single detector, where the time-domain detector noise samples are independent in the time domain.

Denote the data stream of the first detector by $\mathbf{x}_1 = x_{1,j}$ for $j = 0, \dots, N-1$. In this section, since we are going to only consider this single detector, we will drop the subscript "1." Imagine that we are looking for a signal of known waveform but unknown amplitude ϵ, which we will denote by ϵs_j. Our primary interest is in the case where the amplitude ϵ is either small, or zero. For convenience, imagine for the moment that this parameter can have only two possible values, either $\epsilon = 0$ or $\epsilon = \bar{\epsilon} \neq 0$.

The detection problem that we need to solve is to partition the space of possible observations R^N into two disjoint subsets. When the observation $\mathbf{x}$ falls into one of these, we conclude that $\epsilon = 0$ and that the null hypothesis is true. When the observation falls into the other set we conclude that the signal has been observed with $\epsilon \neq 0$. To describe the partition of R^N into two regions, define a function $\delta(\mathbf{x} \in R^N)$ which is zero in the null hypothesis region and unity elsewhere. This function is called a *test*. Our goal is to find the "best" choice of a test δ.

To help characterize tests δ, it is helpful to define the *power* function of a test:

$$F(\delta|\epsilon) = \int_{R^N} \delta(\mathbf{x}) p(\mathbf{x}|\epsilon) d^N x. \qquad (2.1)$$

Here $p(\mathbf{x}|\epsilon)$ is the probability distribution of the measurement $\mathbf{x}$ given signal amplitude ϵ. For example, for additive white, stationary Gaussian noise of unit variance and vanishing mean

$$p(\mathbf{x}|\epsilon) = \prod_{i=0}^{N-1} (2\pi)^{-1/2} e^{-(x_i - \epsilon s_i)^2/2}. \qquad (2.2)$$

The quality of the test can be expressed in terms of the power function.

We characterize the quality of the test by the false alarm and the false dismissal probabilities. The false alarm probability is the probability with which we conclude that $\epsilon \neq 0$

when in fact ϵ vanishes. This is given by $F(\delta|0)$. The false dismissal probability is the probability with which we conclude that $\epsilon = 0$ when in fact it is $\epsilon = \bar{\epsilon} \neq 0$. This is given by $1 - F(\delta|\bar{\epsilon})$.

One standard definition of the "best" test δ is that it minimizes the false dismissal probability for a given false alarm probability. This is called the Neyman-Pearson test. One can find this test using calculus of variations, with a Lagrange multiplier Λ to enforce the constraint that the false alarm probability is fixed. The best test is obtained by partitioning R^N as follows. Choose a constant $\Lambda_0 > 0$. Then, set $\delta = 1$ in regions where the likelihood ratio

$$\Lambda = \frac{p(\mathbf{x}|\bar{\epsilon})}{p(\mathbf{x}|0)} \qquad (2.3)$$

is greater than Λ_0. Set $\delta = 0$ elsewhere. (We assume that the boundary between these two regions is a set of probability measure zero.) The value of the constant Λ_0 determines the false alarm probability. Thus, the likelihood ratio is a "decision statistic:" a number that can be calculated from the observed data. If the statistic is less than some value, we conclude that the null hypothesis holds. If the statistic is greater than this value, we conclude the opposite. The decision statistic provides a partition of the space of observations into two disjoint regions.

In the case where the noise is Gaussian (2.2) this criteria is easily understood. The optimal Neyman-Pearson test divides the space of observation along an $(N-1)$-dimensional plane. On one side of this plane $\delta = 1$ and on the other side δ vanishes. The plane is defined by setting the likelihood ratio (2.3) to a constant. For the Gaussian probability distribution (2.2) the plane is defined by

$$\text{const} = \prod_{i=0}^{N-1} e^{-(x_i - \epsilon s_i)^2/2 + x_i^2/2}$$

$$\Rightarrow \text{const} = \epsilon \sum_{i=0}^{N-1} x_i s_i. \qquad (2.4)$$

This plane is perpendicular to the vector $\mathbf{s}$. Different choices of this plane correspond to different false alarm rates.

In the case where the noise is not Gaussian, the problem becomes more challenging. In the Gaussian test, the decision statistic is independent of the signal amplitude ϵ. However when the noise is not Gaussian, the choice of decision statistic depends upon ϵ. Consider the graph in Fig. 1 showing the power function $F(\delta|\epsilon)$ as a function of ϵ for several different tests. All the tests have the same false alarm rate, but the optimal test depends upon the value of ϵ.

For the case of weak signals in non-Gaussian noise, there is a useful test called the "locally optimal" test. For a given noise probability distribution, the locally optimal test is easy to describe, and leads to a simple decision statistic which can be calculated from the observed data [10]. To define this test, it is useful to again consider the set of all tests with a given false alarm rate, as shown for example in Fig. 1. The locally optimal test is the one that maximizes $dF(\delta|\epsilon)/d\epsilon$ at $\epsilon = 0$
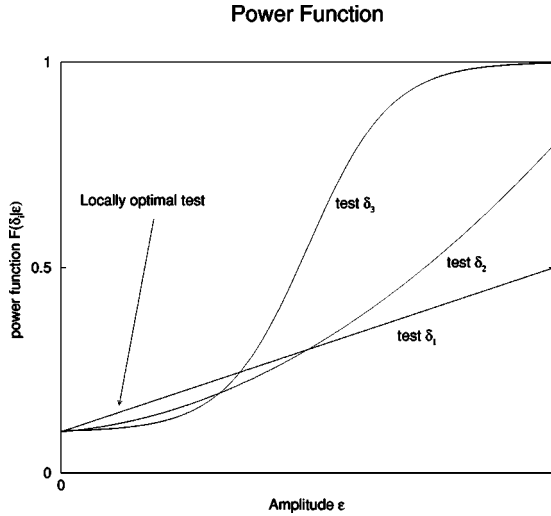
FIG. 1. The power function $F(\delta|\epsilon)$ is shown for three different tests. All have the same false alarm probability $F(\delta|0)$. Test δ_3 has the best performance for large ϵ. Test δ_2 is not the best test for any value of ϵ. Test δ_1 is the best test for small ϵ. The *locally optimal* test δ_1 is the one for which $dF(\delta|\epsilon)/d\epsilon$ is largest at $\epsilon=0$. If the first derivative of the power function with respect to ϵ vanishes for all tests, then the locally optimal test is the one with the largest second derivative (and so on, if additional derivatives vanish).

for a fixed false alarm probability. As above, one can show that the locally optimal test sets $\delta=1$ inside the region where

$$\Lambda_{(1)}=[d\ln p(\mathbf{x}|\epsilon)/d\epsilon]_{\epsilon=0}>\text{const} \qquad (2.5)$$

for some constant, and $\delta=0$ elsewhere (see Fig. 1). The value of the constant determines (or is determined by) the false alarm probability. More generally, if the first derivative vanishes, the locally optimal test is determined by the first non-zero

$$\Lambda_{(n)}=\frac{1}{p(\mathbf{x}|0)}\left.\frac{d^n p(\mathbf{x}|\epsilon)}{d\epsilon^n}\right|_{\epsilon=0}. \qquad (2.6)$$

To understand the implications of this, it is helpful to consider several examples.

The examples here are for the case where the (additive) detector noise is independent for each sample value (so the noise spectrum is white) but has an arbitrary probability distribution. For convenience, we write

$$p(\mathbf{x}|\epsilon)=\prod_{i=0}^{N-1}e^{-f(x_i-\epsilon s_i)} \qquad (2.7)$$

where the function f is a quadratic function of its argument for the case where the probability distribution of the noise is Gaussian. [Note: any probability distribution for stationary additive noise where the sample values are independent can be written in this way. If the noise is not stationary but is still additive and independent, then each function f appearing in Eq. (2.7) may be different $f(x_i-\epsilon s_i)\rightarrow f_i(x_i-\epsilon s_i)$.] The first derivative of the PDF (2.7) with respect to ϵ is

$$\frac{d\ln p(\mathbf{x}|\epsilon)}{d\epsilon}=\sum_{i=0}^{N-1}s_i f'(x_i-\epsilon s_i) \qquad (2.8)$$

where f' denotes the derivative of f with respect to its argument. Setting $\epsilon=0$ in this expression one can easily find the locally optimal test (2.5). This is defined by setting $\delta=1$ in the region

$$\Lambda_{(1)}=\sum_{i=0}^{N-1}s_i f'(x_i)>\text{const} \qquad (2.9)$$

and setting $\delta=0$ elsewhere. [Note: if ϵ can take either sign $\pm\bar{\epsilon}$ then an absolute value sign should enclose the LHS of the inequality in Eq. (2.9).] As before, the value of the constant determines the false alarm probability. Here are several examples:

Gaussian noise: $f(x)=x^2/2+\ln(2\pi)/2$, so $f'(x)=x$. For this case the locally optimal test (2.9) and the optimal test (2.4) both give the same statistic: $\Sigma_{i=0}^{N-1}s_i x_i$. This is the standard optimal linear filter.

Exponential noise: $f(x)=a|x|-\ln(a/2)$, so $f'(x)=a\,\text{sgn}(x)$. Here the locally optimal statistic is given by Eq. (2.9) as

$$\sum_{i=0}^{N-1}s_i\,\text{sgn}(x_i) \qquad (2.10)$$

where the $\text{sgn}(x)$ function is $+1$ for $x\geq 0$ and -1 for $x<0$.

Sum of distinct Gaussian processes: This is a white-noise version of the model given in [11]:

$$e^{-f(x)}=(1-P)(2\pi)^{-1/2}\sigma^{-1}e^{-x^2/2\sigma^2}$$
$$+P(2\pi)^{-1/2}\bar{\sigma}^{-1}e^{-x^2/2\bar{\sigma}^2}, \qquad (2.11)$$

where $0<\sigma<\bar{\sigma}$ and $P\in(0,1)$. Usually one also has $P\ll 1$. This noise model is discussed in more detail later in this paper. It often arises when the most common source of noise is Gaussian, but there is also a "tail" of "outlier" events which dominates the wings of the distribution. Here the locally optimal statistic is defined by Eq. (2.9) where

$$f'(x)=x\sigma^{-2}\left[\frac{(1-P)+P(\sigma/\bar{\sigma})^3 e^{x^2(\sigma^{-2}-\bar{\sigma}^{-2})/2}}{(1-P)+P(\sigma/\bar{\sigma})e^{x^2(\sigma^{-2}-\bar{\sigma}^{-2})/2}}\right].$$

This function is shown (for the case $\sigma=1$, $\bar{\sigma}=4$, $P=1\%$) in Fig. 2. Roughly speaking, for $|x|$ small compared to σ one has $f'(x)\approx x/\sigma^2$. For large $|x|$ one has $f'(x)\approx x/\bar{\sigma}^2$.

Gaussian noise plus uniform background: Here, we have a (small) uniform background superposed on Gaussian noise of zero mean and unit variance. This is defined for (small) $P>0$ by

$$e^{-f(x)}=\begin{cases}(1-P)(2\pi)^{-1/2}e^{-x^2/2}+P/2L, & |x|\leq L,\\ 0, & |x|>L.\end{cases} \qquad (2.12)$$
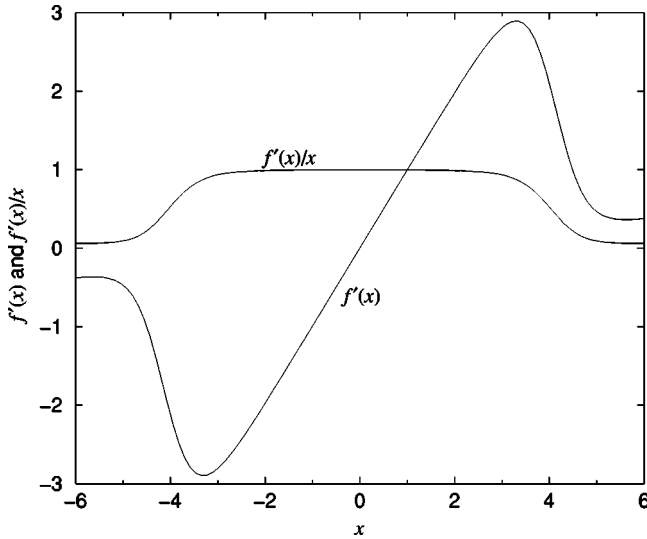
FIG. 2. The functions $f'(x)$ and $f'(x)/x$ are shown for the sum of distinct Gaussian processes, defined by Eq. (2.11) with parameters $\sigma=1$, $\bar{\sigma}=4$, and $P=1\%$. For small $|x|$ one has $f'(x)\approx x$. Outside the central Gaussian region (which dominates the probability density), i.e., for large $|x|$, $f'(x)$ falls off. This effectively "clips" the correlation statistic for outlier data samples.

Here we assume that $L\gg1$ is the scale size of the uniform background (the probability distribution is correctly normalized only in the limit $L\to\infty$). In this case one finds that $f'(x)\approx x$ for $|x|\lesssim1$ and $f'(x)=0$ for $1\lesssim|x|\lesssim L$.

While the results for the different probability distributions are technically different, they all carry the same message, which is the central result of this paper: *If the distribution of sample values has a central Gaussian region, then sample values falling in this region should be correlated exactly as they would be in the Gaussian case. If a sample value falls outside this region, its value should be truncated (or clipped) to the largest allowed value in the central region, or even dropped from any correlation statistic, depending upon the shape of the probability distribution.*

Let us repeat this central point one more time. The results show that when the noise is not Gaussian, the normal optimal filter used to construct a decision statistic is replaced by a somewhat different sum. The values of the expected signal s_i are multiplied, not by the observed data x_i but by some nonlinear function of that data, then summed. In the event that the probability distribution of the noise has a non-Gaussian tail, the effect of this non-linear function is to "clip" or truncate sample values which fall outside the central bulge of the probability distribution function.

B. Single detector, colored noise

If the detector's noise spectrum is colored rather than white, then the previous analysis does not apply: the assumption that the different sample values are uncorrelated no longer holds. However the analysis can be generalized to the colored case if we make assumptions that are motivated by the properties of stationary detector noise.

In explaining this, it helps to begin by describing the stationary Gaussian case. For a colored Gaussian process, the probability density function (PDF) of the detector samples may be expressed as

$$p(\mathbf{x})=(2\pi)^{-N/2}(\det\mathbf{R})^{-1/2}\exp\left(-\frac{1}{2}\mathbf{x}^{\dagger}\cdot\mathbf{Q}\cdot\mathbf{x}\right)\quad(2.13)$$

where the $N\times N$ correlation matrix $\mathbf{R}=\langle\mathbf{x}\otimes\mathbf{x}^{\dagger}\rangle$ is a positive-definite real symmetric matrix with $N(N+1)/2$ real degrees of freedom and $\mathbf{Q}=\mathbf{R}^{-1}$. We have assumed that the process has zero mean. The volume element associated with this PDF is $d\mathbf{x}=\Pi_{j=0}^{N-1}dx_j$. In the time domain, $\mathbf{x}$ is a vector of real numbers so $\mathbf{x}^{\dagger}=\mathbf{x}^{\mathrm{T}}$.

In the case where the random process is stationary, the matrix $\mathbf{R}$ is a Toeplitz matrix, which depends only upon $|i-j|$. Such a process is defined by the first row or first column of the matrix and has only N real degrees of freedom. Thus stationary Gaussian processes are a tiny subset of *all* Gaussian processes.

Now consider the PDF of new random variables that are linear combinations of the old ones: $\widetilde{\mathbf{x}}=\mathbf{U}\cdot\mathbf{x}$. Take $\mathbf{U}$ to be an arbitrary unitary matrix. Clearly the PDF of these new variables $\widetilde{\mathbf{x}}$ is still Gaussian. The matrix $\mathbf{U}$ can be chosen to diagonalize the correlation matrix: this is called a Karhunen-Loeve transformation. In the limit where the time interval occupied by the N samples is much larger than the correlation time of the noise, the linear combinations of random variables that diagonalize the correlation matrix asymptotically approach the discrete Fourier transform (DFT). This is given by

$$U_{jk}=N^{-1/2}e^{2\pi ijk/N}.\quad(2.14)$$

Thus, if N is sufficiently large, to a good approximation the PDF of the new variables in the Gaussian case may be written as

$$p(\widetilde{\mathbf{x}})=\prod_{k=1}^{[(N-1)/2]}2\pi^{-1}P_k^{-1}\exp(-2|\widetilde{x}_k|^2/P_k)\quad(2.15)$$

where P_k is the (real, positive) mean spectral amplitude in the kth frequency bin:

$$\langle\widetilde{x}_k\widetilde{x}_{k'}^*\rangle=\frac{1}{2}\delta_{kk'}P_k\quad(2.16)$$

for $1\leq k,k'\leq[(N-1)/2]$; thus $\widetilde{\mathbf{R}}=\mathbf{U}\cdot\mathbf{R}\cdot\mathbf{U}^{-1}\simeq\frac{1}{2}\mathrm{diag}[P_k]$. In other words, it is a good approximation to express the PDF of a stationary colored Gaussian process as a diagonal process in frequency space.

The limits of the product in Eq. (2.15) appear strange because $\widetilde{\mathbf{x}}$ cannot take on arbitrary values since $\mathbf{x}$ is real. The consequences include:

$\widetilde{x}_k=\widetilde{x}_{N-k}^*$. Hence the amplitudes of $\widetilde{x}_k$ for $k=[N/2]+1,\ldots,N-1$ are completely determined by $\widetilde{x}_k$ for $k=1,\ldots,[(N-1)/2]$.

$\widetilde{x}_0$ and, for even N, $\widetilde{x}_{N/2}$ are real. However, we assume that the data set has had the mean value (DC term) removed: $\Sigma_{j=0}^{N-1}x_j=0$. Since gravitational wave detectors are AC-

coupled and have no useful low-frequency response, this is a valid assumption. It implies that $\tilde{x}_0$ is identically zero. Second, when N is even, we assume that there is no energy in the Nyquist frequency bin: $\tilde{x}_{N/2}$ also vanishes identically. This is a very reasonable assumption, since an experiment will include an anti-aliasing filter whose response (as a function of frequency) falls off rapidly as the Nyquist frequency is approached.

The volume element associated with this PDF is therefore $\Pi_{k=1}^{[(N-1)/2]} d(\Re\tilde{x}_k) d(\Im\tilde{x}_k)$.

The likelihood ratio in the case of colored, stationary, Gaussian noise is

$$\ln \Lambda = \ln p(\mathbf{x} - \epsilon\mathbf{s}) - \ln p(\mathbf{x})$$

$$= \epsilon\mathbf{s}^\dagger \cdot \mathbf{Q} \cdot \mathbf{x} - \frac{1}{2}\epsilon^2\mathbf{s}^\dagger \cdot \mathbf{Q} \cdot \mathbf{s} \qquad (2.17)$$

or, in the frequency domain,

$$\ln \Lambda = (\text{const}) + \epsilon 4\Re \sum_{k=1}^{[(N-1)/2]} \tilde{s}_k^* \tilde{x}_k / P_k. \qquad (2.18)$$

Thus, the matched filter statistic, with a weighting equal to the inverse of the noise spectrum, is the optimal detection statistic.

This motivates a more general model for the statistical distribution of colored non-Gaussian detector noise, assuming that it is still stationary. In this case, to good approximation, the two-point correlation matrix $\langle\tilde{x}_k\tilde{x}_{k'}^*\rangle$ is diagonal. There may be higher-order correlations present between the Fourier amplitudes at different frequencies, but we will assume that this additional correlation is negligible, and that to a reasonable approximation the probability distribution of the noise in the non-Gaussian case is described by a PDF in which the different frequency components are independent:

$$p(\tilde{\mathbf{x}}) = \prod_{k=1}^{[(N-1)/2]} 2\pi^{-1} P_k^{-1} \exp[-g_k(2|\tilde{x}_k|^2/P_k)],$$
$$(2.19)$$

with volume element $\Pi_{k=1}^{[(N-1)/2]} d(\Re\tilde{x}_k) d(\Im\tilde{x}_k)$. The functions $g_k(u)$ depend upon the frequency bin index k, so that the statistical distribution can depend upon the frequency. For the colored Gaussian case the functions are $g_k(u) = u$. In order that the PDF be properly normalized, and that $\langle\tilde{x}_k\tilde{x}_{k'}^*\rangle = \frac{1}{2}\delta_{kk'}P_k$, the functions $g_k(u)$ must obey

$$\int_0^\infty e^{-g_k(u)} du = \int_0^\infty u e^{-g_k(u)} du = 1. \qquad (2.20)$$

Respectively, these constrain the additive constant in the definition of g_k, and the multiplicative scale of the argument of g_k. This is *not* the most general possible form of the probability distribution of a stationary random process, but in many situations it should be a reasonable approximation, particularly if the quantities of interest are dominated by the second moments.

The locally optimal statistic may now be easily derived. Letting $\tilde{s}_k$ denote the DFT of the expected waveform, and as before zeroing its DC and Nyquist components, the conditional probability distribution of the detector output is given by

$$p(\tilde{\mathbf{x}}|\epsilon) = \prod_{k=1}^{[(N-1)/2]} 2\pi^{-1} P_k^{-1} \exp[-g_k(2|\tilde{x}_k - \epsilon\tilde{s}_k|^2/P_k)].$$

The locally optimal test can then be obtained from the first derivative:

$$\Lambda_{(1)} = 4 \sum_{k=1}^{[(N-1)/2]} \Re(\tilde{s}_k^*\tilde{x}_k/P_k) g_k'(2|\tilde{x}_k|^2/P_k). \qquad (2.21)$$

In the colored Gaussian case $g_k'(u) = 1$ this is the ordinary optimal linear matched filter. The contributions of the different frequency bins are weighted by the inverse noise power spectrum in that bin. In the non-Gaussian case, just as for the case of uncolored white noise, the correlation in frequency space is clipped or truncated for (frequency-bin) samples that lie outside the central Gaussian part of the probability distribution, where $|g'(u)| \ll 1$. An example of this may be seen in Fig. 2: for the illustrated case $g'(x^2/2) = f'(x)/x$.

Let us consider another form of non-Gaussian noise that describes a process in which there is an ambient Gaussian noise background interrupted occasionally by a large noise burst, which we will model a second component of Gaussian noise with a much larger variance. The probability distribution we adopt is [11]

$$p(\mathbf{x}) = (1-P)(2\pi)^{-N/2}(\det\mathbf{R})^{-1/2}\exp\left(-\frac{1}{2}\mathbf{x}^\dagger \cdot \mathbf{Q} \cdot \mathbf{x}\right)$$

$$+ P(2\pi)^{-N/2}(\det\bar{\mathbf{R}})^{-1/2}\exp\left(-\frac{1}{2}\mathbf{x}^\dagger \cdot \bar{\mathbf{Q}} \cdot \mathbf{x}\right) \quad (2.22)$$

where $\mathbf{R}$ is the autocorrelation matrix for the normal ambient detector noise and $\bar{\mathbf{R}}$ is the composite autocorrelation matrix for the detector noise when a noise burst is present. The noise bursts occur with probability P in this model. Also, $\mathbf{Q} = \mathbf{R}^{-1}$ and $\bar{\mathbf{Q}} = \bar{\mathbf{R}}^{-1}$. We assume that $\bar{\mathbf{Q}}$ is much smaller than $\mathbf{Q}$, meaning that $\mathbf{x}^\dagger \cdot \mathbf{Q} \cdot \mathbf{x} \gg \mathbf{x}^\dagger \cdot \bar{\mathbf{Q}} \cdot \mathbf{x}$ for all vectors $\mathbf{x}$. The locally optimal statistic is

$$\Lambda_{(1)} = \frac{d\ln p(\mathbf{x} - \epsilon\mathbf{s})}{d\epsilon}\bigg|_{\epsilon=0} = \frac{\Re(\mathbf{s}^\dagger \cdot \mathbf{Q} \cdot \mathbf{x})}{1+\alpha} \qquad (2.23)$$

where

$$\alpha = \frac{P}{1-P}\sqrt{\frac{\det\mathbf{R}}{\det\bar{\mathbf{R}}}}\exp\left[\frac{1}{2}\mathbf{x}^\dagger \cdot (\mathbf{Q} - \bar{\mathbf{Q}}) \cdot \mathbf{x}\right] \qquad (2.24)$$

is a detector of possible bursts. When a burst is absent, α is typically small and the locally optimal statistic reduces to the matched filter. However, when a burst is present, α is typically large and the matched filter is suppressed. Thus the locally optimal statistic is nearly equivalent to the matched

filter statistic with a veto if a segment of data has a large amount of excess power as measured by

$$\mathcal{E} = \mathbf{x}^\dagger \cdot \mathbf{Q} \cdot \mathbf{x} \qquad (2.25)$$

or, in the frequency domain,

$$\mathcal{E} = 4 \sum_{k=1}^{[(N-1)/2]} |\tilde{x}_k|^2 / P_k. \qquad (2.26)$$

The lengths (in time) of the data chunks used to estimate the autocorrelation matrices should be chosen to be significantly longer than the characteristic time of the signals being searched for, but still short enough that the detector behavior is quasi-stationary. For inspiral signals, typical signals are in the detector band for tens of seconds, so the matrix estimation time should be at least of the order of tens of minutes. For stochastic background detection, the correlation time between the two instruments is tens of milliseconds, so that the matrix estimation time should be at least a few seconds.

Based on the two forms of non-Gaussian noise considered in this section, it seems reasonable to adopt the following detection rules: (i) veto immediately any segment of data that has an excess of power as measured by the excess power statistic; (ii) for segments of data without an excess of power, construct the matched filter in the frequency domain, but exclude those frequency bins in which the detector power is too large. The resulting (truncated) matched filter is a good approximation of the locally optimal statistic for a wide variety of possible non-Gaussian noise distribution. In this sense, it is a robust, nearly optimal detection statistic.

III. STOCHASTIC SIGNALS

Observational limits from nucleosynthesis demonstrate that the stochastic background of gravitational radiation has such small amplitude that it would not be detectable with a single instrument [12]. In a single instrument, there would be no practical way to discriminate between intrinsic detector noise and the small additional noise-like output arising from a stochastic background. However, one can correlate the outputs of two different instruments and search for a common signal in this way. If the instrumental noise is not Gaussian, then the previous single-detector analysis can be easily generalized.

A. Two coincident co-aligned detectors, white noise

We begin by considering the simple case in which the two detectors are coincident and co-aligned, so that they have identical output contributions from the stochastic background but independent intrinsic noise. We also assume that the intrinsic noise samples in each detector are independent, and hence white.

If the signal were deterministic (known) then the joint probability distribution for the samples in the two detectors could be written as

$$p(\mathbf{x}_1, \mathbf{x}_2 | \epsilon) = \prod_{i=0}^{N-1} e^{-f_1(x_{1,i} - \epsilon s_i)} e^{-f_2(x_{2,i} - \epsilon s_i)}. \qquad (3.1)$$

This system can be analyzed in exactly the same way as in Sec. II A. However the stochastic background does not produce a known (deterministic) signal, so that the probability distribution needs to be averaged over its expected distribution $p_{\mathrm{sb}}(s_0, \ldots, s_{N-1})$ (which, by reason of the central limit theorem, is almost certainly a multivariate Gaussian). This leads to a joint probability distribution which is given by

$$p(\mathbf{x}_1, \mathbf{x}_2 | \epsilon) = \int ds_0 \cdots \int ds_{N-1} p_{\mathrm{sb}}(s_0, \ldots, s_{N-1})$$

$$\times \prod_{i=0}^{N-1} e^{-f_1(x_{1,i} - \epsilon s_i) - f_2(x_{2,i} - \epsilon s_i)}$$

$$= \int d\mathbf{s} \, p_{\mathrm{sb}}(\mathbf{s}) \prod_{i=0}^{N-1} e^{-f_1(x_{1,i} - \epsilon s_i) - f_2(x_{2,i} - \epsilon s_i)}. \qquad (3.2)$$

Here ϵ may be thought of as the coupling of the detector. The case of small ϵ corresponds to a detector that is only weakly coupled to the signal. For this non-deterministic signal, it is still straightforward to construct a locally optimal test, and a corresponding decision statistic or threshold criterion.

The locally optimal statistic is obtained from the derivative of the probability distribution with respect to ϵ. This is given by

$$\frac{dp(\mathbf{x}_1, \mathbf{x}_2 | \epsilon)}{d\epsilon}$$

$$= \int d\mathbf{s} \, p_{\mathrm{sb}}(\mathbf{s}) \left(\sum_{j=0}^{N-1} s_j [f_1'(x_{1,j} - \epsilon s_j) \right.$$

$$\left. + f_2'(x_{2,j} - \epsilon s_j)] \right) \prod_{i=0}^{N-1} e^{-f_1(x_{1,i} - \epsilon s_i) - f_2(x_{2,i} - \epsilon s_i)}. \qquad (3.3)$$

Setting $\epsilon = 0$ and dividing by $p(\mathbf{x}_1, \mathbf{x}_2 | 0)$ yields the locally optimal statistic:

$$\Lambda_{(1)} = \sum_{j=0}^{N-1} [f_1'(x_{1,j}) + f_2'(x_{2,j})] \int s_j p_{\mathrm{sb}}(\mathbf{s}) d\mathbf{s}. \qquad (3.4)$$

Unfortunately this vanishes if the random process described by $p_{\mathrm{sb}}(\mathbf{s})$ has vanishing mean, since in this case $\int s_j p_{\mathrm{sb}}(\mathbf{s}) d\mathbf{s} = 0$. This is indeed the case for the gravitational-wave stochastic background.

When the *first* derivative vanishes, the locally optimal statistic is defined by having the largest *second* derivative at $\epsilon = 0$. See Fig. 1 for example. Taking another derivative of Eq. (3.3) and setting $\epsilon = 0$ yields

$$\Lambda_{(2)} = \int d\mathbf{s} \, p_{\mathrm{sb}}(\mathbf{s}) \left\{ \left(\sum_{j=0}^{N-1} s_j [f_1'(x_{1,j}) + f_2'(x_{2,j})] \right)^2 \right.$$

$$\left. - \sum_{j=0}^{N-1} s_j^2 [f_1''(x_{1,j}) + f_2''(x_{2,j})] \right\}. \qquad (3.5)$$

The terms that appear in this statistic have different character, and before moving on, some discussion is required.

The locally optimal statistic depends upon the statistical character of the stochastic background radiation through the second-order moments. We will assume that the stochastic background is a stationary process, so that the second order correlation $\langle s_i s_j \rangle$ is a function of the lag $|i-j|$ only:

$$C(|i-j|) = \langle s_i s_j \rangle = \int ds\, p_{\rm sb}(\mathbf{s}) s_i s_j. \qquad (3.6)$$

In a stochastic background search, the "signal model" only requires an assumption about the form of the spectrum. This is (roughly) the Fourier transform of $C(\Delta)$. Without loss of generality we normalize $C(\Delta)$ so that $C(0)=1$ (this simply scales the value of ϵ). Expressing the locally optimal statistic in terms of the correlation function C then gives

$$\Lambda_{(2)} = -\sum_{i=0}^{N-1} [f_1''(x_{1,i}) + f_2''(x_{2,i})] + \sum_{j,k=0}^{N-1} C(|j-k|)$$
$$\times [f_1'(x_{1,j}) f_1'(x_{1,k}) + f_2'(x_{2,j}) f_2'(x_{2,k})$$
$$+ 2 f_1'(x_{1,j}) f_2'(x_{2,k})]. \qquad (3.7)$$

Each of the five terms that appear in Eq. (3.7) has a specific interpretation. The first four terms that appear in the locally optimal estimator $\Lambda_{(2)}$ are generalized "single-detector" statistics which do not cross-correlate the two detectors. They are generalized measures of the "energy" received by each individual detector, and provide useful information only if the stochastic background contributes substantially more to the measured signal than the detector output does, or if the detector's intrinsic noise contributions can somehow be separated from the noise contribution arising from the stochastic background. (This will not be the case for the first few generations of gravitational wave detectors.) The last term in Eq. (3.7) is a generalized cross-correlation (GCC) statistic that provides useful information even if the detector noise dominates the signal: the expected case for gravitational-wave stochastic background. To quote from Kassam [following Eqs. (7)–(24) in Ref. [10]]:

"It is important to note that the increase in power level occurs whenever random signals are present at the individual receivers of the array regardless of whether the signals *across* the array are one common signal or are completely uncorrelated. The GCC part of the locally optimal (LO) statistic responds only to a *common* signal or at least to signals which are spatially correlated across the array elements. This is a major reason why it is useful to employ only the GCC part of the LO statistic in applications involving detection as well as location of signal sources."

For this reason (and others [19]) we drop the single-detector terms from the statistic, and define the GCC statistic as

$$\Lambda_{\rm GCC} = 2 \sum_{j,k=0}^{N-1} C(|j-k|) f_1'(x_{1,j}) f_2'(x_{2,k}). \qquad (3.8)$$

This generalized cross-correlation statistic reduces to the ordinary cross-correlation statistic in the case where the detector noise is Gaussian: $f_1(x) = f_2(x) = x^2/2 + \log(2\pi)/2$. It can be easily generalized to the case of three or more detectors [10].

In practical work C will vanish for lags greater than the light travel time between the two detectors (i.e., 10 ms for the LIGO detectors). This means that even if N is chosen to be very large, $\Lambda_{\rm GCC}$ only correlates samples from the two detectors taken within this time window. (Note: if the detector noise is colored, then the time window may be larger, as will be seen shortly.)

B. Two non-coincident non-co-aligned detectors, colored noise

In this section, we generalize the work of Sec. III A to the case where the two detectors are not coincident or co-aligned, and their noise power spectrum is not white. We assume that the intrinsic detector noise of the two detectors is independent. If the two detectors are widely separated and subject to different environmental influences, this assumption should hold.

Let us start by assuming that the two detectors each have internal (instrumental) colored Gaussian noise with known autocorrelation matrices $\mathbf{R}_{\rm in,1}$ and $\mathbf{R}_{\rm in,2}$, and that the instrumental noise of the two detectors is independent. The stochastic background produces an additional source of colored Gaussian noise that is correlated between the two detectors. The stochastic background noise is measured by the autocorrelation matrices $\mathbf{S}_{11} = \langle \mathbf{s}_1 \otimes \mathbf{s}_1^\dagger \rangle$, $\mathbf{S}_{22} = \langle \mathbf{s}_2 \otimes \mathbf{s}_2^\dagger \rangle$, and the cross-correlation matrices $\mathbf{S}_{12} = \langle \mathbf{s}_1 \otimes \mathbf{s}_2^\dagger \rangle$, $\mathbf{S}_{21} = \langle \mathbf{s}_2 \otimes \mathbf{s}_1^\dagger \rangle$. Since the stochastic background is isotropic, $\mathbf{S}_{11} = \mathbf{S}_{22} = \mathbf{R}_{\rm sb}$ (the stochastic background contribution to the detector's autocorrelation matrices) and $\mathbf{S}_{12} = \mathbf{S}_{21} = \mathbf{S}$ (the cross-correlated noise between the detectors due to the stochastic background). The total autocorrelation noise of the two detectors is $\mathbf{R}_1 = \mathbf{R}_{\rm in,1} + \epsilon^2 \mathbf{R}_{\rm sb}$ and $\mathbf{R}_2 = \mathbf{R}_{\rm in,2} + \epsilon^2 \mathbf{R}_{\rm sb}$. In the presence of the stochastic background, the likelihood ratio is

$$p(\mathbf{x}_1, \mathbf{x}_2 | \epsilon) = (2\pi)^{-N} (\det \mathbf{\Sigma})^{-1} \exp\left(-\frac{1}{2} \boldsymbol{\xi}^\dagger \cdot \mathbf{\Sigma}^{-1} \cdot \boldsymbol{\xi} \right) \qquad (3.9)$$

where

$$\boldsymbol{\xi} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \quad \text{and} \quad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{R}_1 & \epsilon^2 \mathbf{S} \\ \epsilon^2 \mathbf{S} & \mathbf{R}_2 \end{bmatrix}.$$

In the weak signal approximation,

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} \mathbf{Q}_{\rm in,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{\rm in,2} \end{bmatrix}$$
$$- \epsilon^2 \begin{bmatrix} \mathbf{Q}_{\rm in,1} \cdot \mathbf{R}_{\rm sb} \cdot \mathbf{Q}_{\rm in,1} & \mathbf{Q}_{\rm in,1} \cdot \mathbf{S} \cdot \mathbf{Q}_{\rm in,2} \\ \mathbf{Q}_{\rm in,2} \cdot \mathbf{S} \cdot \mathbf{Q}_{\rm in,1} & \mathbf{Q}_{\rm in,2} \cdot \mathbf{R}_{\rm sb} \cdot \mathbf{Q}_{\rm in,2} \end{bmatrix}$$
$$+ \epsilon^4 \left\{ \begin{bmatrix} \mathbf{Q}_{\rm in,1} \cdot \mathbf{R}_{\rm sb} \cdot \mathbf{Q}_{\rm in,1} \cdot \mathbf{R}_{\rm sb} \cdot \mathbf{Q}_{\rm in,1}, \mathbf{0} \\ \mathbf{0}, \mathbf{Q}_{\rm in,2} \cdot \mathbf{R}_{\rm sb} \cdot \mathbf{Q}_{\rm in,2} \cdot \mathbf{R}_{\rm sb} \cdot \mathbf{Q}_{\rm in,2} \end{bmatrix} \right.$$
$$\left. + \begin{bmatrix} \mathbf{Q}_{\rm in,1} \cdot \mathbf{S} \cdot \mathbf{Q}_{\rm in,2} \cdot \mathbf{S} \cdot \mathbf{Q}_{\rm in,1}, \mathbf{0} \\ \mathbf{0}, \mathbf{Q}_{\rm in,2} \cdot \mathbf{S} \cdot \mathbf{Q}_{\rm in,1} \cdot \mathbf{S} \cdot \mathbf{Q}_{\rm in,2} \end{bmatrix} \right\} + O(\epsilon^6), \qquad (3.10)$$

$$\ln \det \boldsymbol{\Sigma} = \ln \det \mathbf{R}_{\text{in},1} + \ln \det \mathbf{R}_{\text{in},2} + \epsilon^2 [\text{tr}(\mathbf{Q}_{\text{in},1} \cdot \mathbf{R}_{\text{sb}})$$

$$+ \text{tr}(\mathbf{Q}_{\text{in},2} \cdot \mathbf{R}_{\text{sb}})] - \epsilon^4 \left[\frac{1}{2} \text{tr}(\mathbf{Q}_{\text{in},1} \cdot \mathbf{R}_{\text{sb}} \cdot \mathbf{Q}_{\text{in},1} \cdot \mathbf{R}_{\text{sb}}) \right.$$

$$+ \frac{1}{2} \text{tr}(\mathbf{Q}_{\text{in},2} \cdot \mathbf{R}_{\text{sb}} \cdot \mathbf{Q}_{\text{in},2} \cdot \mathbf{R}_{\text{sb}})$$

$$\left. + \text{tr}(\mathbf{Q}_{\text{in},2} \cdot \mathbf{S} \cdot \mathbf{Q}_{\text{in},1} \cdot \mathbf{S}) \right] + O(\epsilon^6), \qquad (3.11)$$

and

$$\ln \Lambda = \ln p(\mathbf{x}_1, \mathbf{x}_2 | \epsilon) - \ln p(\mathbf{x}_1, \mathbf{x}_2 | 0)$$

$$= \epsilon^2 \left\{ -\frac{1}{2} \text{tr}(\mathbf{Q}_{\text{in},1} \cdot \mathbf{R}_{\text{sb}}) - \frac{1}{2} \text{tr}(\mathbf{Q}_{\text{in},2} \cdot \mathbf{R}_{\text{sb}}) \right.$$

$$+ \Re(\mathbf{x}_2^\dagger \cdot \mathbf{Q}_{\text{in},2} \cdot \mathbf{S} \cdot \mathbf{Q}_{\text{in},1} \cdot \mathbf{x}_1) + \frac{1}{2} \mathbf{x}_1^\dagger \cdot \mathbf{Q}_{\text{in},1} \cdot \mathbf{R}_{\text{sb}} \cdot \mathbf{Q}_{\text{in},1} \cdot \mathbf{x}_1$$

$$\left. + \frac{1}{2} \mathbf{x}_2^\dagger \cdot \mathbf{Q}_{\text{in},2} \cdot \mathbf{R}_{\text{sb}} \cdot \mathbf{Q}_{\text{in},2} \cdot \mathbf{x}_2 \right\} + O(\epsilon^4) \qquad (3.12)$$

where $\mathbf{Q}_{\text{in},1} = \mathbf{R}_{\text{in},1}^{-1}$ and $\mathbf{Q}_{\text{in},2} = \mathbf{R}_{\text{in},2}^{-1}$. The last two terms represent the autocorrelation "energy" detectors.

The following question now becomes important: how does one obtain the quantities $\mathbf{R}_{\text{in},1}$ and $\mathbf{R}_{\text{in},2}$? There are two possible methods: (i) by a theoretical understanding of the detector, or (ii) by shielding the instrument from the stochastic background and measuring the noise autocorrelation. For gravitational wave searches, method (ii) is not available as there is no way to shield the detector from a stochastic background of gravitational waves. Method (i) holds more promise, but if the stochastic background is expected to be weak, it is unlikely that our understanding of the detector will be sufficient to distinguish between the noise autocorrelations $\mathbf{R}_{\text{in}}$ and $\mathbf{R}_{\text{in}} + \epsilon \mathbf{R}_{\text{sb}}$. We expect that the noise matrices that should be used are the *measured* noise matrices $\mathbf{R}_1 = \langle \mathbf{x}_1 \otimes \mathbf{x}_1^\dagger \rangle$ and $\mathbf{R}_2 = \langle \mathbf{x}_2 \otimes \mathbf{x}_2^\dagger \rangle$, which contain both the internal, instrumental noise as well as the stochastic background "noise." Since it is these quantities rather than $\mathbf{R}_{\text{in},1}$ and $\mathbf{R}_{\text{in},2}$ that are known, the previous analysis must be modified. We now have

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2 \end{bmatrix}$$

$$- \epsilon^2 \begin{bmatrix} \mathbf{0} & \mathbf{Q}_1 \cdot \mathbf{S} \cdot \mathbf{Q}_2 \\ \mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_1 & \mathbf{0} \end{bmatrix}$$

$$+ \epsilon^4 \begin{bmatrix} \mathbf{Q}_1 \cdot \mathbf{S} \cdot \mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_1, & \mathbf{0} \\ \mathbf{0}, & \mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_1 \cdot \mathbf{S} \cdot \mathbf{Q}_2 \end{bmatrix}$$

$$+ O(\epsilon^6), \qquad (3.13)$$

$$\ln \det \boldsymbol{\Sigma} = \ln \det \mathbf{R}_1 + \ln \det \mathbf{R}_2 - \epsilon^4 \text{tr}(\mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_1 \cdot \mathbf{S})$$

$$+ O(\epsilon^6), \qquad (3.14)$$

and

$$\ln \Lambda = \epsilon^2 \Re(\mathbf{x}_2^\dagger \cdot \mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_1 \cdot \mathbf{x}_1) + O(\epsilon^4) \qquad (3.15)$$

where $\mathbf{Q}_1 = \mathbf{R}_1^{-1}$ and $\mathbf{Q}_2 = \mathbf{R}_2^{-1}$. The locally optimal detection statistic (which is appropriate for weak signals) is the cross-correlation statistic.

To generalize to non-Gaussian noise, it is helpful to use moment generating functions. Suppose the vector $\mathbf{n}_1$ represents the internal (instrumental) noise in the first detector. The moment generating function for $\mathbf{n}_1$ is

$$\Phi_{\text{in},1}(\mathbf{w}_1) = \langle e^{i \mathbf{w}^T \cdot \mathbf{n}_1} \rangle \qquad (3.16)$$

and the probability distribution for $\mathbf{n}_1$ is the Fourier transform of the moment generating function:

$$p_{\text{in},1}(\mathbf{n}_1) = \int d\mathbf{w}_1 e^{-i \mathbf{n}_1^T \cdot \mathbf{w}} \Phi_{\text{in},1}. \qquad (3.17)$$

The moment generating function $\Phi_{\text{in},2}(\mathbf{w}_2)$ for the internal noise in detector 2 is defined similarly. We assume that the stochastic background is a multivariate Gaussian with a moment generating function

$$\Phi_{\text{sb}}(\mathbf{w}_1, \mathbf{w}_2) = \exp\left(-\frac{1}{2} \epsilon^2 \boldsymbol{\omega}^T \cdot \boldsymbol{\Sigma}_{\text{sb}} \cdot \boldsymbol{\omega} \right) \qquad (3.18)$$

with

$$\boldsymbol{\omega} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\text{sb}} = \begin{bmatrix} \mathbf{R}_{\text{sb}} & \mathbf{S} \\ \mathbf{S} & \mathbf{R}_{\text{sb}} \end{bmatrix}.$$

Then the moment generating function for the detectors' output is

$$\Phi(\mathbf{w}_1, \mathbf{w}_2) = \langle e^{i \mathbf{w}_1^T \cdot \mathbf{x}_1} e^{i \mathbf{w}_2^T \cdot \mathbf{x}_2} \rangle \qquad (3.19)$$

$$= \Phi_{\text{in},1}(\mathbf{w}_1) \Phi_{\text{in},2}(\mathbf{w}_2) \Phi_{\text{sb}}(\mathbf{w}_1, \mathbf{w}_2)$$

and the joint probability distribution is

$$p(\mathbf{x}_1,\mathbf{x}_2|\epsilon) = \int d\mathbf{w}_1 d\mathbf{w}_2 e^{-i(\mathbf{x}_1^{\mathrm{T}}\cdot\mathbf{w}_1 + \mathbf{x}_2^{\mathrm{T}}\cdot\mathbf{w}_2)}\Phi(\mathbf{w}_1,\mathbf{w}_2)$$

$$= \int d\boldsymbol{\omega}\exp(-i\boldsymbol{\xi}^{\mathrm{T}}\cdot\boldsymbol{\omega})\Phi_{\mathrm{in},1}(\mathbf{w}_1)\Phi_{\mathrm{in},2}(\mathbf{w}_2)\left\{1 - \frac{1}{2}\epsilon^2\boldsymbol{\omega}^{\mathrm{T}}\cdot\boldsymbol{\Sigma}_{\mathrm{sb}}\cdot\boldsymbol{\omega} + O(\epsilon^4)\right\}$$

$$= p(\mathbf{x}_1,\mathbf{x}_2|0) + \epsilon^2\left\{(\boldsymbol{\nabla}^{\mathrm{T}}p_{\mathrm{in},2})(\mathbf{x}_2)\cdot\mathbf{S}\cdot(\boldsymbol{\nabla}p_{\mathrm{in},1})(\mathbf{x}_1) + \frac{1}{2}(\boldsymbol{\nabla}^{\mathrm{T}}p_{\mathrm{in},1})(\mathbf{x}_1)\cdot\mathbf{R}_{\mathrm{sb}}\cdot(\boldsymbol{\nabla}p_{\mathrm{in},1})(\mathbf{x}_1)\right.$$

$$\left. + \frac{1}{2}(\boldsymbol{\nabla}^{\mathrm{T}}p_{\mathrm{in},2})(\mathbf{x}_2)\cdot\mathbf{R}_{\mathrm{sb}}\cdot(\boldsymbol{\nabla}p_{\mathrm{in},2})(\mathbf{x}_2)\right\} + O(\epsilon^4). \tag{3.20}$$

Thus, if we ignore the autocorrelation terms, the locally optimal statistic is

$$\Lambda_{(2)} = (\boldsymbol{\nabla}^{\mathrm{T}}\ln p_{\mathrm{in},2})(\mathbf{x}_2)\cdot\mathbf{S}\cdot(\boldsymbol{\nabla}\ln p_{\mathrm{in},1})(\mathbf{x}_1). \tag{3.21}$$

This equation for the locally optimal statistic is good for the time domain, in which the detectors' output vectors are real and so the derivative is meaningful.

To extend the result to complex vectors, and thus to a frequency-domain representation, we use the following formal replacement: replace every complex number $x = a + ib$ and derivative ∇ with the matrices

$$x \to \underline{x} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \quad \text{and} \quad \nabla \to \underline{\nabla} = \frac{1}{2}\begin{bmatrix} \partial/\partial a & -\partial/\partial b \\ \partial/\partial b & \partial/\partial a \end{bmatrix}.$$

Note that this means x^* is represented by $\underline{x}^{\mathrm{T}}$ and $|x|^2$ by $\underline{x}^{\mathrm{T}}\cdot\underline{x}$. Also, the meaning of $\nabla|x|^2$ is $\underline{\nabla}(\underline{x}^{\mathrm{T}}\cdot\underline{x}) = \underline{x}^{\mathrm{T}}$. The locally optimal statistic is

$$\Lambda_{(2)} = \tfrac{1}{2}(\underline{\boldsymbol{\nabla}}^{\mathrm{T}}\ln p_{\mathrm{in},2})(\underline{\mathbf{x}}_2)\cdot\underline{\mathbf{S}}\cdot(\underline{\boldsymbol{\nabla}}\ln p_{\mathrm{in},1})(\underline{\mathbf{x}}_1) + \tfrac{1}{2}(\underline{\boldsymbol{\nabla}}\ln p_{\mathrm{in},2})$$

$$\times(\underline{\mathbf{x}}_2)\cdot\underline{\mathbf{S}}\cdot(\underline{\boldsymbol{\nabla}}^{\mathrm{T}}\ln p_{\mathrm{in},1})(\underline{\mathbf{x}}_1). \tag{3.22}$$

For example, for the noise model in which $\ln p_{\mathrm{in}}(\widetilde{\mathbf{x}}) \propto \Sigma_{k=1}^{[(N-1)/2]}g_k(|\widetilde{x}_k|^2/P_k)$ and $\widetilde{\mathbf{S}} = \mathrm{diag}[\gamma_k\sigma_k^2]$, the locally optimal statistic is

$$\Lambda_{(2)} = \Re\sum_{k=1}^{[(N-1)/2]}\frac{\gamma_k\sigma_k^2\widetilde{x}_{1,k}^*\widetilde{x}_{2,k}}{P_{1,k}P_{2,k}}$$

$$\times g_{1,k}'(|\widetilde{x}_{1,k}|^2/P_{1,k})g_{2,k}'(|\widetilde{x}_{2,k}|^2/P_{2,k}). \tag{3.23}$$

Before we examine specific non-Gaussian noise models, we will describe the form of the matrices $\mathbf{R}_{\mathrm{sb}}$ and $\mathbf{S}$. A stochastic background, if present, contributes to the signal amplitude at each detector. To simplify the analysis, in Sec. III A, we assumed that the detectors were coincident and co-aligned, so that the amplitude contribution in each individual detector is identical. Here, we drop that assumption.

Because the detectors are not co-aligned, the axes of the two interferometer arms point in different directions, and are sensitive to different linear combinations of the two possible gravitational wave polarizations. This reduces the correlation between the amplitudes in the two detectors, since we will assume that the stochastic background is unpolarized. An additional loss of correlation occurs because the two detectors are separated. This loss of correlation becomes increasingly greater for shorter wavelengths. Roughly speaking, there is no significant loss of correlation for wavelengths much longer than the inter-detector distance, and there is a complete loss of correlations for wavelengths much shorter than this [12].

The loss of amplitude correlation due to the separation and non-alignment of the two detectors may be described (for an unpolarized and isotropic stochastic background) in terms of the overlap reduction function $\gamma(f)$ defined by Flanagan [13]. This quantity is the average value of the product of the detector outputs, for a stochastic background of a given frequency f, averaged over the possible directions of arrival and phases. It is given by

$$\gamma(f) = \frac{5}{8\pi}\int_{S^2}d\hat{\Omega}e^{2\pi if\hat{\Omega}\cdot\Delta\vec{x}/c}(F_1^+ F_2^+ + F_1^\times F_2^\times). \tag{3.24}$$

Here $\hat{\Omega}$ is a unit-length vector on the two-sphere, $\Delta\vec{x}$ is the separation between the two detector sites, and $F_i^{+,\times}$ is the response of detector i to the $+$ or $\times$ polarization. For the ith detector ($i = 1,2$) one has

$$F_i^{+,\times} = \frac{1}{2}(\hat{X}_i^a\hat{X}_i^b - \hat{Y}_i^a\hat{Y}_i^b)e_{ab}^{+,\times}(\hat{\Omega}), \tag{3.25}$$

where $e_{ab}^{+,\times}(\hat{\Omega})$ are the gravitational wave polarization tensors for a wave propagating in direction $\hat{\Omega}$. The normalization of $\gamma(f)$ is chosen so that for coincident and co-aligned detectors, $\gamma(f) = 1$. For co-aligned but not coincident detectors, $\gamma(f=0) = 1$. For coincident but unaligned detectors, $\gamma(f)$ is a frequency-independent constant that depends only upon the relative orientation of the two detectors, and vanishes if the two detectors are sensitive to orthogonal polarizations.

General expressions for $\gamma(f)$ for arbitrary detectors may be found in Refs. [13,14]. For the pair of Laser Interferomet-
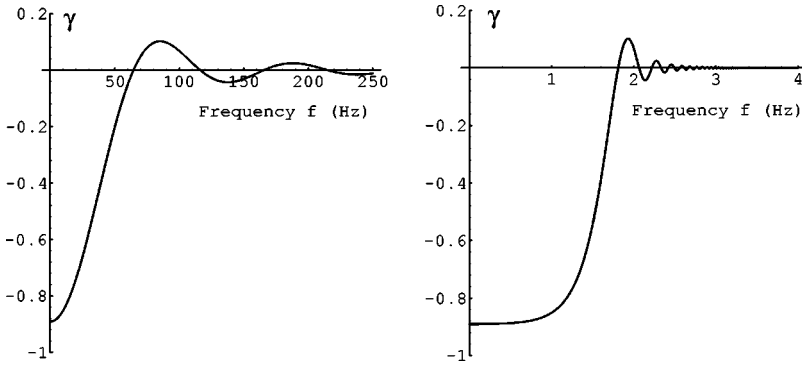
FIG. 3. The overlap reduction function $\gamma(f)$ is shown for the two LIGO detectors as a function of frequency f. The left or right graphs have linear or $\log_{10}$ frequency axes. Because the detectors are almost anti-aligned, the function is close to -1 at low frequencies. The first root is at 64 Hz.

ric Gravitational Wave Observatory (LIGO) detectors $\gamma(f)$ is shown in Fig. 3, and is given by

$$\gamma(f) \approx -0.1248 j_0(x) - 2.900 \frac{j_1(x)}{x} + 3.008 \frac{j_2(x)}{x^2} \tag{3.26}$$

where $x = 2\pi f d/c$ is a frequency variable, $d = 3010$ km is the detector separation, $c = 2.998 \times 10^5$ km/s is the speed of light, and j_n is a spherical Bessel function. It is helpful to introduce notation for the overlap reduction function's values in the frequency bins of interest. Let $f_k = k/(N\Delta t) = k/T$ denote the frequency of the kth bin, with $k = 0, \ldots, [N/2]$. Here Δt is the sample interval and $T = N\Delta t$ is the total observation time. Then

$$\gamma_k = \gamma(f_k) = \gamma(k/T) \tag{3.27}$$

are the values of the overlap reduction function in the kth bin.

The stochastic background is characterized by its dimensionless energy density

$$\Omega_{\text{gw}}(f) = \frac{1}{\rho_{\text{critical}}} \frac{d\rho_{\text{gw}}}{d \ln f}, \tag{3.28}$$

where $d\rho_{\text{gw}}$ is the energy density of the gravitational radiation contained in the frequency range f to $f+df$, and ρ_{critical} is the critical energy density required (today) to close the universe:

$$\rho_{\text{critical}} = \frac{3c^2 H_0^2}{8\pi G} \approx 1.6 \times 10^{-8} h_{100}^2 \text{ erg cm}^{-3}. \tag{3.29}$$

H_0 is the Hubble expansion rate (today):

$$H_0 = h_{100} \times 100 \text{ km s}^{-1} \text{ Mpc}^{-1} = 3.2 \times 10^{-18} h_{100} \text{ s}^{-1}, \tag{3.30}$$

and h_{100} is a dimensionless factor that we have included to account for the different values of H_0 that are quoted in the literature [20].

The probability distribution function (PDF) of the stochastic background strain can usually be expressed in closed form. The central limit theorem shows that if the stochastic background has been produced (as it is in many scenarios) by an incoherent sum of many small processes, then its sta-

tistics will be stationary [15] and Gaussian. This means that it is characterized by the single-site second moments

$$\langle \tilde{s}_{1,k} \tilde{s}_{1,k'}^* \rangle = \langle \tilde{s}_{2,k} \tilde{s}_{2,k'}^* \rangle = \sigma_k^2 \delta_{kk'}, \tag{3.31}$$

with

$$\sigma_k^2 = \frac{3 H_0^2 \Omega_{\text{gw}}(f_k)}{20\pi^2 \Delta t |f_k|^3}. \tag{3.32}$$

As before, we have assumed that N is chosen so that $N\Delta t$ is much larger than the correlation time of the stochastic background (filtered by the instrument response function), so that the RHS of Eq. (3.33) is proportional to $\delta_{kk'}$. The expectation value of the product of the strain at the two different sites is reduced by the overlap reduction function:

$$\langle \tilde{s}_{1,k} \tilde{s}_{2,k'}^* \rangle = \langle \tilde{s}_{2,k} \tilde{s}_{1,k'}^* \rangle = \gamma_k \langle \tilde{s}_{1,k} \tilde{s}_{1,k'}^* \rangle = \gamma_k \sigma_k^2 \delta_{kk'}. \tag{3.33}$$

This follows from Eq. (3.56) of Ref. [14]. In practice, since the shape of the stochastic background spectrum is not known, the dependence of the σ_k on k should be assumed to fit some simple parametrized model, such as a power law $\sigma_k^2 \propto k^\alpha$ for a reasonable range of α.

We can now express the locally optimal detection statistic for a stochastic background in colored Gaussian noise. It is

$$\ln \Lambda = \epsilon^2 \Re \sum_{k=1}^{[(N-1)/2]} \gamma_k \sigma_k^2 x_{1,k}^* x_{2,k}/(P_{1,k} P_{2,k}) \tag{3.34}$$

where $P_{1,k}$ and $P_{2,k}$ are the measured noise spectra in the two detectors.

Let us now turn to our first non-Gaussian noise model. Our starting point is a PDF for the noise in the two detectors in the absence of any stochastic background signal. We make the same assumptions about the detector noise as in Sec. II B. The PDF is given in frequency space by a product of two terms identical to Eq. (2.19),

$$p(\tilde{\mathbf{x}}_1,\tilde{\mathbf{x}}_2)=\prod_{k=1}^{[(N-1)/2]}2\pi^{-1}P_{1,k}^{-2}e^{-2g_{1,k}(|\tilde{x}_{1,k}|^2/P_{1,k})}$$

$$\times\prod_{k'=1}^{[(N-1)/2]}2\pi^{-1}P_{2,k'}^{-2}e^{-2g_{2,k'}(|\tilde{x}_{2,k'}|^2/P_{2,k'})}$$

$$=\prod_{k=1}^{[(N-1)/2]}4\pi^{-2}P_{1,k}^{-2}P_{2,k}^{-2}$$

$$\times e^{-2g_{1,k}(|\tilde{x}_{1,k}|^2/P_{1,k})-2g_{2,k}(|\tilde{x}_{2,k}|^2/P_{2,k})}.$$

The statistical distribution of the stochastic background is

$$p_{\rm sb}(\tilde{\mathbf{s}}_1,\tilde{\mathbf{s}}_2)=\prod_{k=1}^{[(N-1)/2]}(\pi\sigma_k^2)^{-2}(1-\gamma_k^2)^{-1}$$

$$\times\exp\left(-\frac{|\tilde{s}_{1,k}|^2+|\tilde{s}_{2,k}|^2-2\gamma_k\Re(\tilde{s}_{1,k}^*\tilde{s}_{2,k})}{\sigma_k^2(1-\gamma_k^2)}\right).$$

$$(3.35)$$

We can now find the locally optimal statistic. Since the detector is linear, as before, one has a joint probability distribution for the observed Fourier amplitudes:

$$p(\tilde{\mathbf{x}}_1,\tilde{\mathbf{x}}_2|\epsilon)=\int d\tilde{\mathbf{s}}_1 d\tilde{\mathbf{s}}_2 p_{\rm sb}(\tilde{\mathbf{s}}_1,\tilde{\mathbf{s}}_2)$$

$$\times p(\tilde{\mathbf{x}}_1-\epsilon\tilde{\mathbf{s}}_1,\tilde{\mathbf{x}}_2-\epsilon\tilde{\mathbf{s}}_2).$$

This corresponds to a stochastic background with a characteristic energy-density function $\epsilon\Omega_{\rm gw}(f)$ [21]. The locally optimal statistic is

$$\Lambda_{(1)}=4\Re\sum_{k=1}^{[(N-1)/2]}\left\{\frac{\langle\tilde{s}_{1,k}^*\rangle\tilde{x}_{1,k}g'_{1,k}(|\tilde{x}_{1,k}|^2/P_{1,k})}{P_{1,k}}\right.$$

$$\left.+\frac{\langle\tilde{s}_{2,k}^*\rangle\tilde{x}_{2,k}g'_{2,k}(|\tilde{x}_{2,k}|^2/P_{2,k})}{P_{2,k}}\right\}$$

$$(3.36)$$

where the quantities $\langle\tilde{s}_{1,k}^*\rangle$ and $\langle\tilde{s}_{2,k}^*\rangle$ are mean values of the stochastic background's Fourier amplitudes at each of the two detector sites. These both vanish,

$$\langle\tilde{s}_{\{1,2\},k}^*\rangle=\int d\tilde{\mathbf{s}}_1 d\tilde{\mathbf{s}}_2 p_{\rm sb}(\tilde{\mathbf{s}}_1,\tilde{\mathbf{s}}_2)s_{\{1,2\},k}^*=0,\quad(3.37)$$

since the mean values of the Fourier amplitudes are zero. Hence, as in Sec. III A one must look for the locally optimal statistic at the next order in ϵ. Taking an additional derivative, one can easily compute $\Lambda_{(2)}$. As in Sec. III A this consists of two types of terms. For the same reasons as before, we discard from this decision statistic all the *single detector* terms. This leaves us with the following generalized cross-correlation statistic:

$$\Lambda_{\rm GCC}=16\int d\tilde{\mathbf{s}}_1 d\tilde{\mathbf{s}}_2 p_{\rm sb}(\tilde{\mathbf{s}}_1,\tilde{\mathbf{s}}_2)$$

$$\times\sum_{k,k'=1}^{[(N-1)/2]}\frac{\Re(\tilde{s}_{1,k}^*\tilde{x}_{1,k})g'_{1,k}(|\tilde{x}_{1,k}|^2/P_{1,k})}{P_{1,k}}$$

$$\times\frac{\Re(\tilde{s}_{2,k'}^*\tilde{x}_{2,k'})g'_{2,k'}(|\tilde{x}_{2,k'}|^2/P_{2,k'})}{P_{2,k'}}.\quad(3.38)$$

Since the expectation value of the product of the stochastic background at the two sites is given by

$$\langle\tilde{s}_{1,k}\tilde{s}_{2,k'}\rangle=\int d\tilde{\mathbf{s}}_1 d\tilde{\mathbf{s}}_2 p_{\rm sb}(\tilde{\mathbf{s}}_1,\tilde{\mathbf{s}}_2)\tilde{s}_{1,k}\tilde{s}_{2,k}^*=\delta_{kk'}\gamma_k\sigma_k^2$$

$$(3.39)$$

one obtains the generalized cross-correlation statistic

$$\Lambda_{\rm GCC}=16\Re\sum_{k=1}^{[(N-1)/2]}\frac{\gamma_k\sigma_k^2\tilde{x}_{1,k}^*\tilde{x}_{2,k}}{P_{1,k}P_{2,k}}$$

$$\times g'_{1,k}(|\tilde{x}_{1,k}|^2/P_{1,k})g'_{2,k}(|\tilde{x}_{2,k}|^2/P_{2,k}).\quad(3.40)$$

If the functions g' are replaced by unity, this reduces to the standard result for the optimal filter for the case where the detector noise is assumed to be stationary and Gaussian. For typical non-Gaussian noise models, the effect of the g' functions is to exclude those frequency bins in which $|\tilde{x}_k|^2/P_k$ is large in either detector.

Our second non-Gaussian noise model is similar to the noise burst model used in Sec. II B, generalized to the two detector case. The composite PDF for this model is

$$p(\mathbf{x}_1,\mathbf{x}_2|\epsilon)=(2\pi)^{-N}\left\{(1-P_1)(1-P_2)(\det\mathbf{\Sigma})^{-1}\right.$$

$$\times\exp\left(-\frac{1}{2}\boldsymbol{\xi}^\dagger\cdot\mathbf{\Sigma}^{-1}\cdot\boldsymbol{\xi}\right)+P_1(1-P_2)(\det\mathbf{\Sigma}_1)^{-1}$$

$$\times\exp\left(-\frac{1}{2}\boldsymbol{\xi}^\dagger\cdot\mathbf{\Sigma}_1^{-1}\cdot\boldsymbol{\xi}\right)+P_2(1-P_1)(\det\mathbf{\Sigma}_2)^{-1}$$

$$\times\exp\left(-\frac{1}{2}\boldsymbol{\xi}^\dagger\cdot\mathbf{\Sigma}_2^{-1}\cdot\boldsymbol{\xi}\right)+P_1P_2(\det\mathbf{\Sigma}_{12})^{-1}$$

$$\left.\times\exp\left(-\frac{1}{2}\boldsymbol{\xi}^\dagger\cdot\mathbf{\Sigma}_{12}^{-1}\cdot\boldsymbol{\xi}\right)\right\}\quad(3.41)$$

where P_1 and P_2 are the probabilities of bursts in detectors 1 and 2. The matrices $\mathbf{\Sigma}_1$, $\mathbf{\Sigma}_2$, and $\mathbf{\Sigma}_{12}$ represent the correlation matrices when a noise burst is present. As in Sec. II B, a burst effectively changes the noise level for the detector experiencing the burst. Thus, if there is a burst in detector 1, simply replace $\mathbf{R}_1$ with $\bar{\mathbf{R}}_1$ in $\mathbf{\Sigma}$ to obtain $\mathbf{\Sigma}_1$. Then we find

$$\mathbf{\Sigma}_1^{-1}\simeq\begin{bmatrix}\bar{\mathbf{Q}}_1 & -\epsilon^2\bar{\mathbf{Q}}_1\cdot\mathbf{S}\cdot\mathbf{Q}_2 \\ -\epsilon^2\mathbf{Q}_2\cdot\mathbf{S}\cdot\bar{\mathbf{Q}}_1 & \mathbf{Q}_2\end{bmatrix}+O(\epsilon^4)$$

$$(3.42)$$

and

$$\ln \det \boldsymbol{\Sigma}_1 \simeq \ln \det \overline{\mathbf{R}}_1 + \ln \det \mathbf{R}_2 + \mathrm{tr}(\overline{\mathbf{Q}}_1 \cdot \mathbf{R}_1) + O(\epsilon^4) \tag{3.43}$$

to first order in $\overline{\mathbf{Q}}_1$ and similarly for $\boldsymbol{\Sigma}_2$. We also have

$$\boldsymbol{\Sigma}_{12}^{-1} \simeq \begin{bmatrix} \overline{\mathbf{Q}}_1 & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{Q}}_2 \end{bmatrix} \tag{3.44}$$

and

$$\ln \det \boldsymbol{\Sigma}_1 \simeq \ln \det \overline{\mathbf{R}}_1 + \ln \det \overline{\mathbf{R}}_2 + \mathrm{tr}(\overline{\mathbf{Q}}_1 \cdot \mathbf{R}_1) + \mathrm{tr}(\overline{\mathbf{Q}}_2 \cdot \mathbf{R}_2) \tag{3.45}$$

to first order in $\overline{\mathbf{Q}}_1$ and $\overline{\mathbf{Q}}_2$.

We can now compute the locally optimal statistic:

$$\Lambda_{(2)} \simeq \frac{2 \Re(\mathbf{x}_2^\dagger \cdot \mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_1 \cdot \mathbf{x})}{1 + \alpha_1 + \alpha_2 + \alpha_1 \alpha_2} \tag{3.46}$$

where

$$\alpha_1 \simeq \frac{P_1}{1 - P_1} \frac{\det \mathbf{R}_1}{\det \overline{\mathbf{R}}_1} \exp\left(\frac{1}{2} \mathbf{x}_1^\dagger \cdot \mathbf{Q}_1 \cdot \mathbf{x}\right) \tag{3.47}$$

and α_2 is given by a similar expression. Here we have neglected all $\overline{\mathbf{Q}}$ terms. The terms α_1 and α_2 detect bursts, and their role is to suppress $\Lambda_{(2)}$ when a burst is present in either detector.

C. Estimators

In analyzing experimental data, there are different possible goals. One goal might be to set an upper limit (with a certain statistical confidence) on the stochastic background energy density in a particular frequency band. Another goal might be to estimate this energy density in a particular frequency band.

For this latter purpose, there are different possible estimators that might be used. One standard estimator is the maximum likelihood estimators (MLE). In this section, we show how this estimator is related to the cross-correlation statistic.

Recall that the probability distribution for the joint detector output is

$$\ln p(\mathbf{x}_1, \mathbf{x}_2 | \epsilon) = (\text{terms that do not depend on } \epsilon)$$

$$+ \epsilon^2 \mathbf{x}_2^\dagger \cdot \mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_2 \cdot \mathbf{x}_1 + \frac{1}{2} \epsilon^4 \{ \mathrm{tr}(\mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_1 \cdot \mathbf{S})$$

$$- \mathbf{x}_1^\dagger \cdot \mathbf{Q}_1 \cdot \mathbf{S} \cdot \mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_1 \cdot \mathbf{x}_1$$

$$- \mathbf{x}_2^\dagger \cdot \mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_1 \cdot \mathbf{S} \cdot \mathbf{Q}_2 \cdot \mathbf{x}_2 \} + O(\epsilon^6). \tag{3.48}$$

Suppose we wish to estimate the strength ϵ^2 of the stochastic background. The maximum likelihood estimator is the value $\epsilon_{\mathrm{MLE}}^2$ for which this probability is maximized: $[d \ln p(\mathbf{x}_1, \mathbf{x}_2 | \epsilon)/d\epsilon^2]_{\epsilon_{\mathrm{MLE}}^2} = 0$. The result is

$$\epsilon_{\mathrm{MLE}}^2 = \eta \mathbf{x}_2^\dagger \cdot \mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_2 \cdot \mathbf{x}_1 \tag{3.49}$$

where

$$\eta^{-1} = -\mathrm{tr}(\mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_1 \cdot \mathbf{S}) + \mathbf{x}_1^\dagger \cdot \mathbf{Q}_1 \cdot \mathbf{S} \cdot \mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_1 \cdot \mathbf{x}_1$$

$$+ \mathbf{x}_2^\dagger \cdot \mathbf{Q}_2 \cdot \mathbf{S} \cdot \mathbf{Q}_1 \cdot \mathbf{S} \cdot \mathbf{Q}_2 \cdot \mathbf{x}_2 \tag{3.50}$$

is a measure of how sensitive the detectors were to the stochastic background. Normally η will be on the order of unity so $\epsilon_{\mathrm{MLE}}^2$ is approximately just the cross-correlation statistic. However, if the detectors were abnormally noisy, then η would be less than unity and the estimate of the stochastic background strength would be smaller than the cross-correlation statistic would indicate: this is a correction that compensates for artificially large values of the cross-correlation statistic due to noise fluctuations.

IV. IMPLEMENTATION AND SIMULATIONS

A. Implementation

A nice feature of these techniques is that in practice, they should be easy to implement. Work by Scott and Whiting [16] has shown that the PDFs of the Fourier amplitudes in different frequency bins can be easily obtained. Since the characteristic time scale for stochastic background correlation is ≈ 10 ms, these can be computed using data segments with lengths of seconds or tens of seconds. These PDFs can then be used to determine where to truncate or clip the correlation, frequency bin by frequency bin. Provided that the instrument's characteristics are stable over periods of minutes or hours, it should be simple to accumulate sufficient statistics to determine the PDFs and therefore the truncation or weighting functions with reasonable accuracy.

In practice, it may also be desirable to "discard" a small part of the "attainable-in-principle" correlation in exchange for obtaining more robust statistics. For example, one can arbitrarily zero the 1% of frequency bins that are the largest number of standard deviations away from the mean value (for that bin). Since the dominant contribution in any bin always comes from the detector noise, this is only very weakly correlated with the actual stochastic background signal, and the net effect is to discard just a bit more than 1% of the "in principle" attainable signal-to-noise ratio. But in exchange, the detection statistic becomes far less sensitive to non-Gaussian detector fluctuations. The precise effects of such treatment, and the appropriate truncation thresholds, can be easily determined with Monte Carlo simulations using simulated signals added into real detector noise.

In searching for a known waveform (e.g., binary inspiral) the methods are again easily implementable. Here, since the signal time scale is less than a minute, the frequency-bin by frequency-bin statistics take a bit more time to accumulate, and the detector's statistical properties have to be stable over a slightly longer time scale (an hour, perhaps). This appears likely.

Since certain non-Gaussian noise features are more likely to appear as outlier points in the time domain, and others in the frequency domain, a combination of the time- and

frequency-domain methods may be desirable. Unfortunately, if the detector noise is not white, this may require the removal (vetoing) of entire small sections of time-series data. This is easy in the stochastic background case, where only tens of milliseconds around a glitch need excision. It may be more problematic for signals like binary inspiral chirps that have longer duration.

B. Comparing different statistics

In Secs. II and III, we derived locally optimal statistics to search for deterministic and stochastic gravitational wave signals in the presence of non-Gaussian noise. These statistics reduce to the standard matched-filtering and cross-correlation statistics when the detector noise is Gaussian. But they are more *robust* (i.e., less sensitive to outliers) when the detector noise has non-Gaussian components. For both cases, the standard and robust statistics take (as input) the output of one or more detectors, and return (as output) a single real number. But the statistics also depend on the gravitational wave signal and detector noise models, which are not directly observable. Different choices for the signal and noise models correspond to different statistics, and these different statistics will in general perform differently given the same detector output. In order to compare and evaluate the statistics, we need a way to quantify their performance.

As mentioned in Sec. II, the quality of a test (i.e., a decision rule based on a particular statistic) is characterized by its *false alarm* and *false dismissal* probabilities for a given source. These are, respectively, the probability that the test leads us to conclude that a signal is present, when in fact it is absent ($\epsilon = 0$), and the probability that the test leads us to conclude that a signal is absent, when in fact it is present ($\epsilon > 0$). These two probabilities (denoted α and β_ϵ) completely specify the long-term performance of a statistic. But to rank different tests, we need to reduce these multi-dimensional error measures to a single figure of merit. How we do this depends on the problem we are trying to solve (see, e.g. [17]), but in the context of gravitational wave detection, it is common to look for a test that minimizes the false dismissal probability, keeping the false alarm probability less than or equal to some maximum tolerable value. This criterion is known as the *Neyman-Pearson* criterion, and it was used in Sec. II to define the locally optimal statistics.

Thus, to compare the performance of different statistics, we should plot false dismissal versus false alarm curves for different values of the signal amplitude ϵ. The best test (or best statistic) is the one that has the smallest false dismissal probability $\beta_\epsilon(\alpha)$, for fixed false alarm probability α and fixed signal amplitude ϵ. Note that since the false dismissal probability depends on both α and ϵ, it is possible that the best test for one choice of (α, ϵ) is not the best test for a different choice of (α, ϵ). Note also that this method of comparing statistics is different than simply comparing expected signal-to-noise ratios. What is important when determining error rates (and hence the performance of a particular test) is not the expected value of the statistic, but rather its probability *distribution*.

For sufficiently simple statistics with sufficiently simple signal and noise models, it may be possible to analytically calculate the corresponding false dismissal versus false alarm curves. But for most cases of interest, we must resort to *Monte Carlo simulations* to generate the curves. This consists of adding simulated signals to simulated (or real) detector noise, and then processing the resulting data with a statistic. For each stretch of data, the statistic outputs a single number which is then compared to a threshold to determine if we should claim detection. Since we know if a signal is present in the data, we can easily determine the fraction of times that the decision rule was in error. In the absence of a signal, this procedure yields the false alarm probability α as a function of the threshold Λ_0. In the presence of a signal having fixed amplitude ϵ, we obtain the false dismissal probability β_ϵ, again as a function of the threshold. If we invert $\alpha(\Lambda_0)$ for $\Lambda_0 = \Lambda_0(\alpha)$, and substitute this expression back into $\beta_\epsilon(\Lambda_0)$, we obtain the false dismissal versus false alarm curve $\beta_\epsilon(\alpha)$. We can then repeat these steps for a different signal amplitude ϵ' to produce a new curve $\beta_{\epsilon'}(\alpha)$. The final result will be a set of curves similar to those shown in Fig. 4.

Alternatively, we can plot $1 - \alpha - \beta_\epsilon$ or $\epsilon^{-2}(1 - \alpha - \beta_\epsilon)$ versus α, as shown in Figs. 5 and 6. Note that the quantity $1 - \alpha - \beta_\epsilon$ is the difference of two probabilities: $1 - \beta_\epsilon$ is the probability that the statistic exceeds some threshold in the presence of a signal ($\epsilon > 0$), while α is the probability that the statistic exceeds the same threshold in the absence of a signal (i.e., $\epsilon = 0$). Although Figs. 5 and 6 contain the same information as the false dismissal versus false alarm curves (Fig. 4), plotting $\epsilon^{-2}(1 - \alpha - \beta_\epsilon)$ versus α has the nice property that, for stochastic signals, the curves have a well-defined $\epsilon \rightarrow 0$ limit.

C. Example

To illustrate how we can compare different statistics using Monte Carlo simulations, consider the simple case of a search for a white, Gaussian stochastic background signal using two independent, identical, coincident and coaligned detectors. Statistic 1 will be the standard cross-correlation statistic defined by a white, Gaussian stochastic background signal and white, Gaussian detector noise. Statistic 2 will be a locally optimal statistic, also defined by a white, Gaussian stochastic background signal, but with white, 2-component, mixture Gaussian noise with an arbitrary knee. We will assume that we know (*a priori*) that the two detectors are identical and have uncorrelated white noise. We will not assume, however, that we know (*a priori*) the parameters describing the statistical properties of the detector noise or the overall amplitude of the stochastic background signal. Each statistic will have to internally estimate the parameters from the detector output, without any other prior knowledge.

We perform Monte Carlo simulations of the two statistics for the following three cases:

(i) Uncorrelated, white, Gaussian detector noise with zero mean and unit variance.

(ii) Uncorrelated, white, 2-component, mixture Gaussian detector noise with zero mean, unit variance, $\bar{\sigma}/\sigma = 4$, and $P = 1\%$ [see Eq. (2.11)].
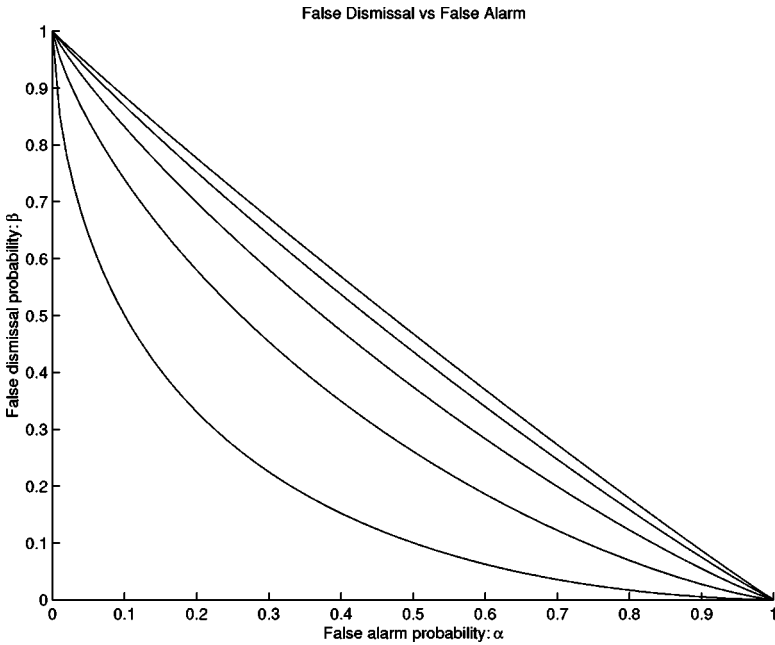
FIG. 4. False dismissal versus false alarm curves for a typical statistic. Lower curves correspond to larger values of the signal amplitude ϵ.

(iii) Uncorrelated, white, exponential detector noise with zero mean and unit variance [see Eq. (2.10)].

The first two simulations test the optimal behavior of the statistics. Statistic 1 is designed for the data of case (i), and statistic 2 is designed for the data of case (ii). The third simulation tests the two statistics in a sub-optimal situation, representative of a real search where we do not know in advance the exact statistical character of the detector noise.

Details of the Monte Carlo simulations are summarized below:

(i) A single stretch of data consists of $N = 1024$ discrete-time samples. This N is sufficiently large that the large observation time approximation is valid. Since we are considering white noise (which has zero correlation length), any $N \gtrsim 100$ would do.

(ii) The simulated stochastic gravitational-wave signal strengths are $\epsilon^2 = .0025, .005, .010, .020,$ and $.040$, where ϵ is the ratio of the rms amplitude of the stochastic background signal to the rms amplitude of the detector noise. These signal strengths correspond to signal-to-noise ratios $(\sim \epsilon^2 \sqrt{N})$ ranging from approximately 0.1 to 1 for a single stretch of data.

Note: Since a real stochastic background is expected to have a smaller value of ϵ^2 $(\sim 10^{-4})$, we would need a much longer observation time to build up similar signal-to-noise ratios in a real search. The purpose of this example, however, is to illustrate how one can compare two different statistics; it is not meant to simulate a real $(\gtrsim 4$ month) stochastic background search.
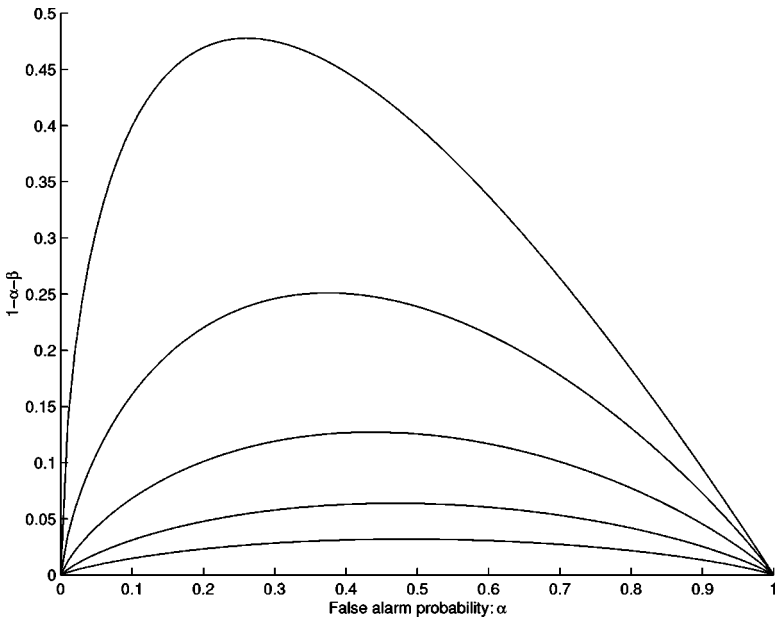

FIG. 5. $1 - \alpha - \beta_\epsilon$ versus the false alarm probability α for a typical statistic. Lower curves correspond to smaller values of the signal amplitude ϵ.
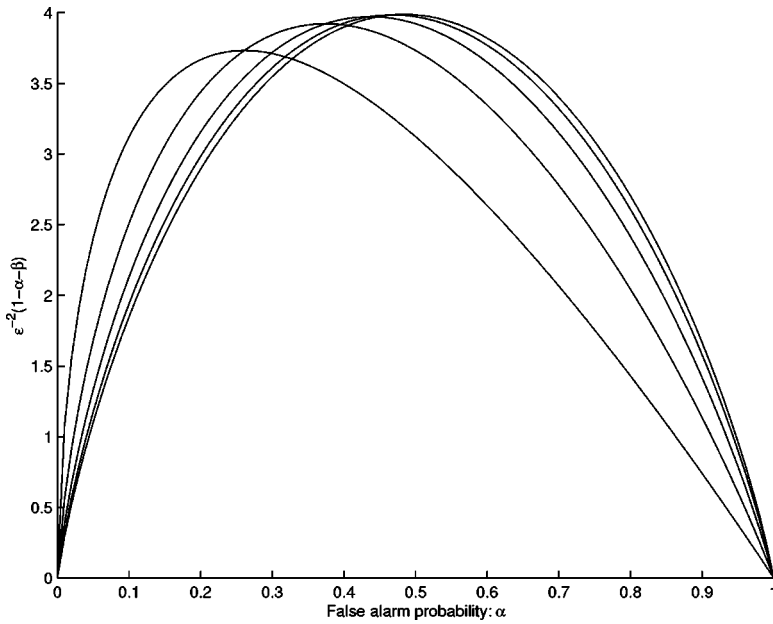
FIG. 6. $\epsilon^{-2}(1 - \alpha - \beta_\epsilon)$ versus the false alarm probability α for a typical statistic. Higher curves correspond to smaller values of the signal amplitude ϵ.

(iii) For all three types of simulated detector noise, the standard cross-correlation statistic estimates the variance of the noise by calculating the sample variance of a stretch of detector output equal to $100N$. Since the detector output consists in general of signal plus noise, the estimate of the noise variance gets worse as the signal amplitude increases. The sample variance is needed to specify the white, Gaussian noise model that enters the definition of the standard cross-correlation statistic [cf. Eq. (3.8)]:

$$^{(1)}\Lambda_{\mathrm{GCC}} = \frac{1}{N} \sum_{j=0}^{N-1} x_{1,j} x_{2,j} / \sigma_1^2 \sigma_2^2, \tag{4.1}$$

where σ_1^2 and σ_2^2 are the estimated variances of the noise in detectors 1 and 2, respectively.

False Dismissal vs False Alarm
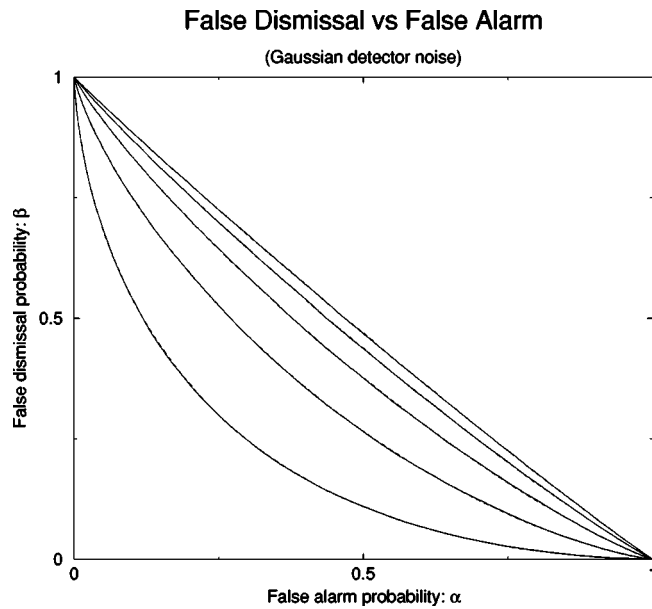
(Gaussian detector noise)

FIG. 7. False dismissal versus false alarm curves for the standard cross-correlation and locally optimal statistics for simulated white, Gaussian detector noise. The solid lines correspond to the standard cross-correlation statistic; the dashed lines correspond to the locally optimal statistic. The top curve for each statistic has $\epsilon^2 = .0025$; ϵ^2 increases by a factor of 2 as one moves to successively lower curves in the graph. As explained in the text, the two statistics perform almost identically for this case.

False Dismissal vs False Alarm

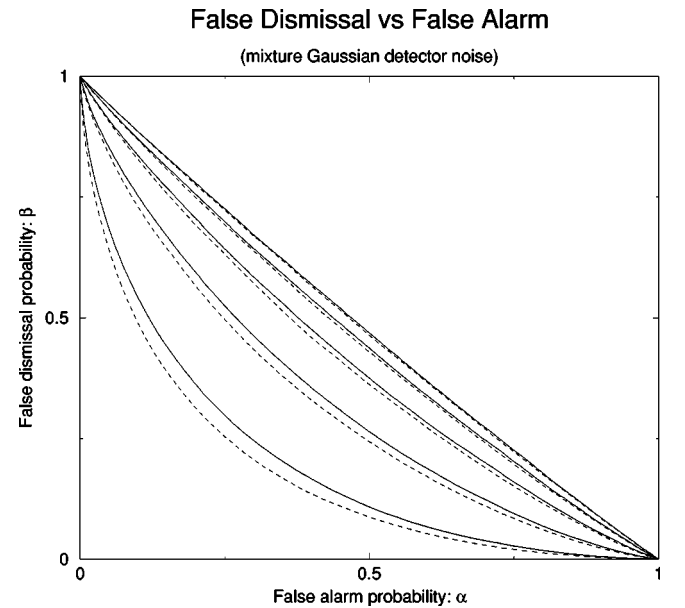### (mixture Gaussian detector noise)

FIG. 8. False dismissal versus false alarm curves for the standard cross-correlation and locally optimal statistics for simulated white, 2-component, mixture Gaussian detector noise. The solid lines correspond to the standard cross-correlation statistic; the dashed lines correspond to the locally optimal statistic. The top curve for each statistic has $\epsilon^2 = .0025$; ϵ^2 increases by a factor of 2 as one moves to successively lower curves in the graph. Since the locally optimal statistic has a lower false dismissal probability $\beta_\epsilon(\alpha)$ for each false alarm probability α and each signal amplitude ϵ, it is clearly the better test for this case, as expected.

False Dismissal vs False Alarm
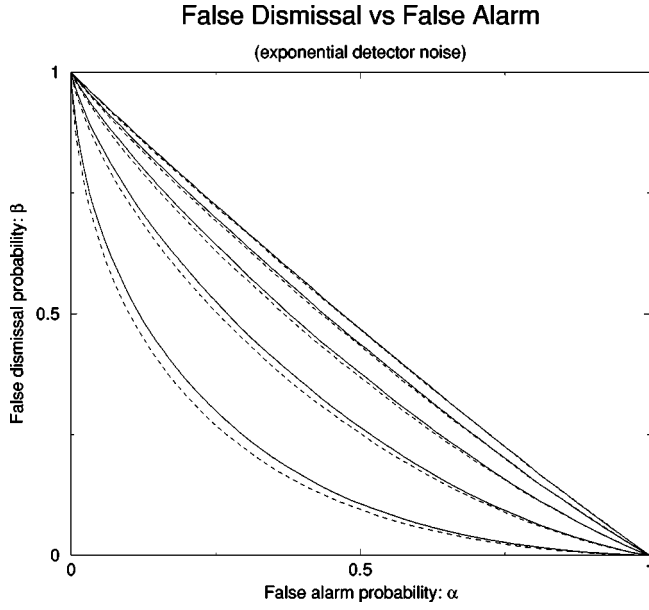### (exponential detector noise)

FIG. 9. False dismissal versus false alarm curves for the standard cross-correlation and locally optimal statistics for simulated white, exponential detector noise. The solid lines correspond to the standard cross-correlation statistic; the dashed lines correspond to the locally optimal statistic. The top curve for each statistic has $\epsilon^2 = .0025$; ϵ^2 increases by a factor of 2 as one moves to successively lower curves in the graph. Since the locally optimal statistic has a lower false dismissal probability $\beta_\epsilon(\alpha)$ for each false alarm probability α and each signal amplitude ϵ, it is the better test for this case.

(iv) In addition to estimating the variance of the detector noise, the locally optimal statistic also estimates the variances σ, $\bar{\sigma}$, and breakpoint x_b of the 2-component, mixture Gaussian model that define this statistic. It does this by fitting two straight lines to a $\ln(p(x))$ vs x^2 plot obtained from a histogram of a stretch of detector output, again equal to $100N$. Best-fit lines at small x and large x, respectively, yield estimates of σ and $\bar{\sigma}$, while the intersection of the lines yields an estimate of x_b. Actually, only the breakpoints for the detector noise are needed to define the following locally optimal statistic:

$$^{(2)}\Lambda_{\text{GCC}} = \frac{1}{N} \sum_{j=0}^{N-1} x_{1,j} \Theta(x_{1b} - |x_{1,j}|)$$

$$\times x_{2,j} \Theta(x_{2b} - |x_{2,j}|)/\sigma_1^2 \sigma_2^2, \qquad (4.2)$$

which is a truncated version of $^{(1)}\Lambda_{\text{GCC}}$. (See the discussion of truncation in the previous subsection.) Here $\Theta(x)$ is the usual step function, which equals 0 if $x < 0$, and equals 1 if $x \geq 0$.

Note: In order to handle pure Gaussian noise (which is a pathological case when one tries to model it as a 2-component, mixture Gaussian distribution), the locally optimal statistic sets the breakpoint x_b to ∞ whenever the estimated slopes at small and large values of x have a percent difference less than 10%. By doing this, the locally optimal
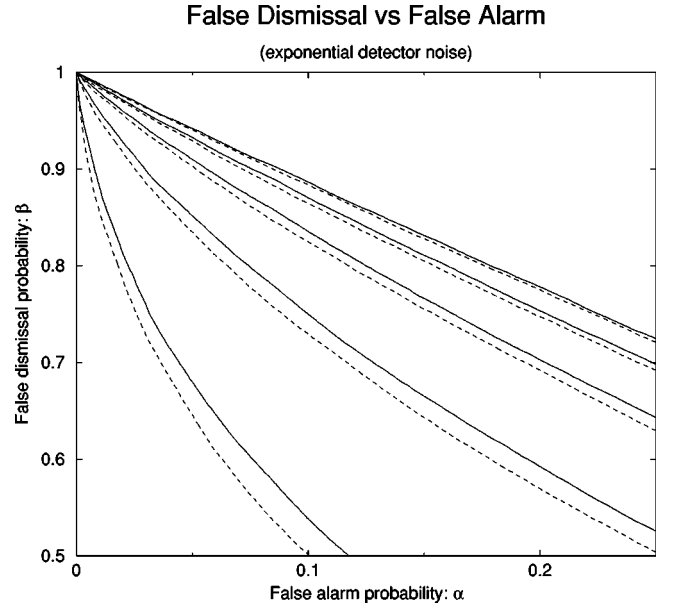
False Dismissal vs False Alarm
(exponential detector noise)

FIG. 10. A blowup of the false dismissal versus false alarm curves from Fig. 9 for small values of the false alarm probability α.

statistic $^{(2)}\Lambda_{\text{GCC}}$ effectively reduces to the standard cross-correlation statistic $^{(1)}\Lambda_{GCC}$ when the noise is pure Gaussian.

(v) We use 10^5 trials to generate each false dismissal versus false alarm curve.
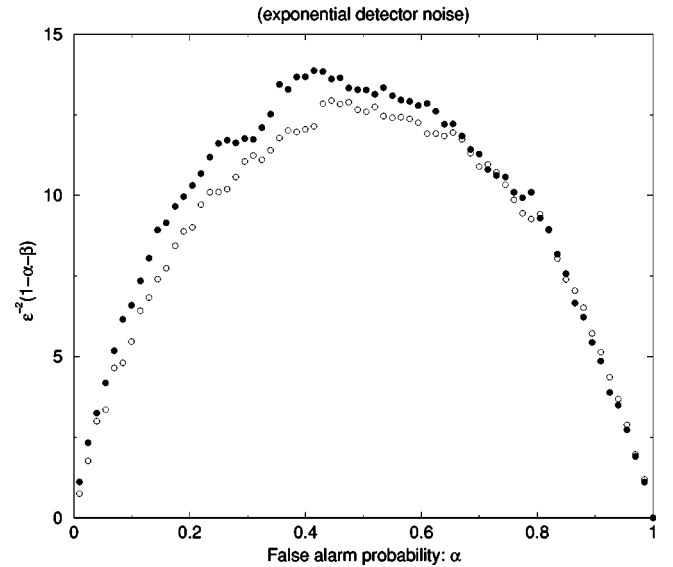
Performance of the statistics
(exponential detector noise)

FIG. 11. A plot of $\epsilon^{-2}(1 - \alpha - \beta_\epsilon)$ versus the false alarm probability α for the standard cross-correlation and locally optimal statistics for simulated white, exponential detector noise in the weak signal limit (small ϵ). The top curve (filled circles) corresponds to the locally optimal statistic; the lower curve (open circles) corresponds to the standard cross-correlation statistic. The difference between the performance of the two statistics in the small signal limit is more apparent in this plot (cf. Fig. 9). Since the locally optimal statistic has a larger value of $\epsilon^{-2}(1 - \alpha - \beta_\epsilon)$ for each false alarm probability α, it is the better test for this case.

(vi) The simulations were written in Matlab [18].

The results of the simulation are shown in Figs. 7–11.

As noted in (iv) above, our implementation of the locally optimal statistic reduces to the standard cross-correlation statistic when the detector noise is pure Gaussian. That is why the false dismissal versus false alarm curves for the two statistics are effectively identical in Fig. 7.

From Fig. 8 we see that the locally optimal statistic performs better than the standard cross-correlation statistic when the simulated detector noise is mixture Gaussian. For each value of the stochastic signal strength ϵ^2 and for each false alarm probability α, the false dismissal probability $\beta_\epsilon(\alpha)$ for the locally optimal statistic is less than that for the standard cross-correlation statistic. This is as expected, since the locally optimal statistic was constructed precisely to handle mixture Gaussian noise.

Finally, from Figs. 9–11 we see that the locally optimal statistic also performs better than the standard cross-correlation statistic when the simulated detector noise has an exponential distribution. The difference in performance between the two statistics for this case is less than that for mixture Gaussian noise, but it is still noticeable. [Figure 10 focuses attention on the false dismissal versus false alarm curves for small values of the false alarm probability, while Fig. 11 is a plot of $\epsilon^{-2}(1-\alpha-\beta_\epsilon)$ versus α, which highlights the difference between the two statistics in the small signal limit.] This behavior is again as expected, since a locally optimal statistic is constructed to be less sensitive to the tails of a non-Gaussian distribution.

V. CONCLUSION

In this paper, we have constructed a replacement for the standard linear matched filter estimators used for gravitational wave detection. The replacements are more robust because they are less susceptible to corruption by non-Gaussian detector noise.

We have explicitly illustrated the locally optimal detection strategies for a variety of different noise PDFs, and for two different detection problems (single detector known waveform, and two-detector stochastic background). In all cases, the optimal strategy is similar to the one for Gaussian noise except that data samples that lie outside the central part of the distribution (the outliers) are excluded from the sums which form the estimators.

We believe that for the future generation of sensitive gravitational wave detectors, these strategies may be easily implemented and offer an improvement on the existing matched filter algorithms.

[1] A. Abramovici, W.E. Althouse, R.W.P. Drever, Y. Gürsel, S. Kawamura, F.J. Raab, D. Shoemaker, L. Sievers, R.E. Spero, K.S. Thorne, R.E. Vogt, R. Weiss, S.E. Whitcomb, and M.E. Zucker, Science **256**, 325 (1992).

[2] Barry C. Barish and Rainer Weiss, Phys. Today **52**(10), 44 (1999).

[3] C. Bradaschia *et al.*, Nucl. Instrum. Methods Phys. Res. A **289**, 518 (1990); also in *Gravitation 1990*, Proceedings of the Banff Summer Institute, Banff, Alberta, 1990, edited by R. Mann and P. Wesson (World Scientific, Singapore, 1991).

[4] K. Danzmann *et al.*, in *Gravitational Wave Experiments*, Proceedings of the Edoardo Amaldi Conference, edited by E. Coccia, G. Pizzella, and F. Ronga (World Scientific, Singapore, 1995), p. 100.

[5] K. Tsubono, in *Gravitational Wave Experiments* (Ref. [4]), p. 112.

[6] I.S. Heng, D.G. Blair, E.N. Ivanov, and M.E. Tobar, Phys. Lett. A **218**, 190 (1996); E. Mauceli *et al.*, Phys. Rev. D **54**, 1264 (1996); P. Astone *et al.*, *ibid.* **47**, 362 (1993); M. Cerdonio *et al.*, in *Gravitational Wave Experiments* (Ref. [4]), p. 176; E. Coccia, *ibid.*, p. 161.

[7] Instruments currently in the planning and proposal stage include the second-generation LIGO detector (LIGO-II), an advanced European detector, and a 3 km scale-up of the TAMA-300 detector.

[8] K.S. Thorne, in *300 Years of Gravitation*, edited by S.W.

Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1987), pp. 330–458.

[9] Simple experiments carried out by Allen and Romano using the GRASP data analysis package http://www.lsc-group.phys.uwm.edu indicate that the large non-Gaussian fluctuations present in the data from the LIGO 40-m prototype created large biases in the cross-correlation estimators. The 1994 data stream was split into two parts, and each was assumed to come from an independent detector.

[10] Saleem A. Kassam, *Signal Detection in Non-Gaussian Noise* (Springer-Verlag, New York, 1988).

[11] Jolien D.E. Creighton, Phys. Rev. D **60**, 021101 (1999).

[12] B. Allen, in *Proceedings of the Les Houches School on Astrophysical Sources of Gravitational Waves*, edited by Jean-Alain Marck and Jean-Pierre Lasota (Cambridge University Press, Cambridge, England, 1997).

[13] É.É. Flanagan, Phys. Rev. D **48**, 2389 (1993). Note that the second term on the RHS of Eq. (B6) should read $-10j_1(\alpha)$ rather than $-2j_1(\alpha)$, that the sliding delay function shown in Fig. 2 is incorrect, and that the right-hand side of Eq. (2.8) should be divided by π.

[14] B. Allen and J. Romano, Phys. Rev. D **59**, 102001 (1999).

[15] Bruce Allen, É.É. Flanagan, and Maria Alessandra Papa, Phys. Rev. D **61**, 024024 (2000).

[16] Susan Scott and Bernard Whiting, work reported at the LIGO Scientific Collaboration Meetings in August 1999, and March

2000, and at the TAMA Meeting in October 1999.

[17] L.S. Finn (unpublished).

[18] Matlab: http://www.mathworks.com. The matlab scripts that we used to generate the Monte Carlo simulations are available at: http://feynman.utb.edu/ ∼joe/ research/ stochastic/ algorithms/ robust/ simulations/ code/ .

[19] There are other motivations for dropping the remaining terms. In particular, the remaining terms are non-zero in the absence of correlated stochastic background: the GCC term vanishes in this case. One can show that terms 2, 3, and 4 are all positive definite (since $\langle s_i s_j \rangle$ is positive definite) and that for the expected "Gaussian + tail" detector probability distribution, term 1 is also positive definite. Baysian analysis (in part II of this series of papers) also shows that the single-detector terms must be dropped. This may also be seen in the analysis at the end of Sec. III C.

[20] h_{100} is probably very close to 0.65.

[21] In fact ϵ should really be interpreted as "coupling to the detector." The locally optimal statistic is the best choice in the "weakly coupled detector" limit.