Optimal detection of burst events in gravitational wave interferometric observatories

Andrea Viceré
LIGO Laboratory, California Institute of Technology, M/S 18-34 1200 East California Blvd, Pasadena, CA 91125, USA[*]
(Dated: 8th December 2001)

We consider the problem of detecting a burst signal of unknown shape in the data from gravitational wave interferometric detectors. We introduce a statistic which generalizes the *excess power* statistic proposed first by Flanagan and Hughes, and then extended by Anderson *et al.* also to a multiple detector case. The statistic that we propose is shown to be optimal for arbitrary noise spectral characteristic, under the two hypotheses that the noise is Gaussian, albeit colored, and that the prior for the signal is uniform.

The statistic derivation is based on the assumption that a signal affects only affects $N_{\|}$ samples in the data stream, but that no other information is a priori available, and that the value of the signal at each sample can be arbitrary. This is the main difference from previous works, where different assumptions were made, like a signal distribution uniform with respect to the metric induced by the (inverse) noise correlation matrix. The two choices are equivalent if the noise is *white*, and in that limit the two statistics do indeed coincide. In the general case, we believe that the statistic we propose may be more appropriate, because it does not reflect the characteristics of the noise affecting the detector on the supposed distribution of the gravitational wave signal.

Moreover we show that the proposed statistic can be easily implemented in its exact form, combining standard time-series analysis tools which can be efficiently implemented, and the resulting computational cost is still compatible with an *on-line* analysis of interferometric data.

We generalize this version of an excess power statistic to the multiple detector case, considering first the noise uncorrelated among the different instruments, and then including the effect of correlated noise: we show that this can be done either perturbatively, or in exact form.

We give full details about the implementation of the algorithm, both for the single and the multiple detector case, and we discuss exact and approximate forms; the choice among them depends on the specific characteristics of the noise and on the assumed length of the burst event.

As a example, we show what would be the sensitivity of the network of interferometers to a δ-function burst.

PACS numbers: 04.80.Nn, 05.45.Tp, 07.05.Kf

I. INTRODUCTION AND SUMMARY

Several large scale interferometric detectors [1, 2, 3, 4] are currently under commissioning and are expected to start data acquisition and reach their design sensitivity in a few years. Some of the candidate sources, like the coalescing binaries in their inspiral phase, can be modeled with reasonable accuracy and the gravitational waveforms can be predicted, thus allowing a matched-filter detection strategy; see [5] and references therein for a review. On the other hand, as argued in [6] it is conceivable that the uncertainty on the waveform will remain high for sources like the Type II supernova explosions or the merger phase in the coalescence of black holes or neutron stars: in this context, the issue of detecting events poorly modeled or not modeled at all remains crucial.

The problem has already been faced from different point of views: some authors [6, 7, 8] aim at devising several simple and computational inexpensive algorithms, to be run in parallel after having been tested and optimized against model waveforms [9]. Others start from general hypotheses on the distribution of the signals and the noise and derive statistics optimal under those assumptions [10, 11, 12, 13].

We consider of particular interest strategies, like the *excess power* statistic proposed by Flanagan and Hughes [11], or the *norm filter* studied by Arnaud *et al.* [10], which try to make minimal assumptions on the nature of the signal, like time duration and bandwidth only: in particular the excess power statistic has been recently analyzed by Anderson *et al.* [13] and extended to the "blind" search of burst events from a network of interferometers. However, as pointed out in [13, note 8], the authors have actually made the assumption that the signal distribution is flat with respect to the inner product defined by the inverse of the noise correlation matrix. They recognize that this is an approximation, and correctly claim that it is legitimate when the noise spectrum does not vary rapidly in the band of interest: however we will argue in Section II A that this may be not true for real detector noise, and we will show that in view of the current models for burst signals from the core collapse of supernovae [9, 14] the correlation length of the detector noise cannot be assumed short with respect to the event duration, even assuming the design noise.

The choice of the signal distribution in [13] has the advantage of making easy to incorporate a priori informations, when available, about the absolute scale of the expected signals: we shall see that this is in general more complicated with our statistic, if detection thresholds were to be set using a Bayesian criterion.

The analysis method that we propose consists essentially of two steps:

A filter the input data with a matched filter for δ functions.

B Compute a statistic similar to the *energy* of the data within each time window we are willing to test for the pres-

[*]Permanent address: Istituto di Fisica dell'Universitá di Urbino, via S.Chiara 27, I-61029 Urbino, Italy; Electronic address: vicere@fis.uniurb.it

LIGO-P010019-01-D

ence of a burst, using a particular scalar product which can be conveniently computed either using the discrete Karhunen Loève transform (DKLT), or (approximately) using a discrete Fourier transform (DFT).

The algorithm can be easily generalized to the multiple detector case, resulting in a optimal statistic which depends on the direction in the sky the signal is supposed to come from. It also turns out that possible correlations among the detector noises can be taken into account in a natural way by modifying the δ-filtering step, and that these modifications are a natural consequence of the likelihood maximization procedure.

The paper is organized in two main sections: in Section II we derive the statistic for the case of a single detector: then in Section III we extend it to the multiple detector case.

In greater detail, the plan of the work is as follows: in Section II A we motivate our study, introducing our hypotheses on the signal and the noise and discussing them in view of the current models for supernovae signals. In Section II B we briefly recall the Bayesian framework we follow in deriving the optimal statistic. In Section II C we consider the simple but illuminating case in which the burst event is a δ function, that is with a duration affecting only one data sample: we deduce a detection strategy that corresponds to one of the data analysis methods used in resonant bar experiments [15]. In Section II D we derive the exact expression for the likelihood ratio, for a burst affecting $N_{\|}$ samples in the data stream, and we show that the calculational steps are essentially the two **A,B** mentioned above. In Section II E we make a digression on the Karhunen-Loève transform, a tool well known in statistics [16, 17] and already applied in the analysis of data from bar detectors [18], and proposed for the study of narrow resonances [19]: in Section II F we apply the DKLT expansion to our problem and we show that it is a convenient way to implement step **B**. When the supposed burst length $N_{\|}$ is large, an approximate formula using the DFT can be used, as discussed in Section II F 2.

The very meaning of the word "optimal" used for defining the proposed statistic is discussed in Section II G, where we clarify the limits of the method, in particular with respect to the possible inclusion of prior information on the strength of the signal.

We elaborate in Section II H on the details of the detection algorithm, and we show that step **A** is easily implemented using standard tools of time series analysis, including the whitening transformation, although the required operation is different from a whitening.

The distribution of our statistic in absence and presence of a signal is considered in Section II I: we show that it corresponds exactly to a χ^2 variable, respectively central or not central.

The multiple detector case is treated first in the approximation of uncorrelated Gaussian noise across the detector network in Section III B, where we show that the techniques used in the single detector case can be easily extended, in a way not much different from what has been done in [13], but including the above mentioned step **A** implementing the filtering for δ functions. We show in Section III B 1 that the step **B** of the resulting algorithm can be written in exact form using a vector DKLT, while a simpler form, similar to the one derived in [13], is valid in the long burst limit, as discussed in Section III B 2.

The case in which the noise of different detectors displays some degree of correlation is considered in Section III C: we start from the same hypotheses as Finn [20] and we show how the effect of the cross-detector terms in the noise correlation matrix can be taken into account in our algorithm, either perturbatively, if the cross terms are small, or exactly if they are not: not surprisingly, the needed modifications turn out to affect only the δ-filtering step, and in a simple way.

We finally give an example of application of the algorithm to the detection of bursts of unit duration, for a network of detectors comprising either the three LIGOs [1] or including also GEO600 [3], TAMA [4] and Virgo [2]. We compute the resulting SNR, which is a function of the direction in the sky; this allows us to pictorially show to which extent the network analysis strategy is advantageous, at least in the ideal situation in which the detector noise is Gaussian.

Throughout all the paper we adopt a discrete-time, discrete-frequency notation: the conventions followed for the Discrete Fourier Transform require some care and are detailed in App. A, while the characteristics of the detectors in the network are detailed in App. B.

II. SINGLE DETECTOR ANALYSIS

A. Noise and signal statistics

We will keep consistently a discrete time, discrete frequency notation, assuming a sampling rate f_s and a finite observation time $T = N/f_s$. We assume that the detector noise has zero mean and is Gaussian, albeit colored, characterized by a correlation matrix

$$(\mathbf{R}_n)[i,j] = E[n_i n_j] \tag{1}$$

where $n_i \equiv n[i] \equiv n(i/f_s)$, and $i \in [1,N]$. We further assume that the noise is stationary, hence $\mathbf{R}$ is a symmetric Toeplitz matrix [17], whose entries depend only on the difference of indices. In terms of this correlation matrix, the probability of observing a certain set of noise data $\mathbf{n}$ (of total length N) is given by the joint distribution

$$P(\mathbf{n}) = \frac{1}{\sqrt{(2\pi)^N \det R}} \exp\left[-\frac{1}{2} n_i (\mathbf{R}_n^{-1})_{ij} n_j\right]; \tag{2}$$

if a certain signal $\mathbf{s}$ is also present, the conditioned probability of observing a set of data $\mathbf{x}$ is

$$P(\mathbf{x}|\mathbf{s}) = \frac{1}{\sqrt{(2\pi)^N \det R}} e^{\left[-\frac{1}{2}(\mathbf{x}-\mathbf{s})\cdot\mathbf{R}_n^{-1}\cdot(\mathbf{x}-\mathbf{s})\right]}. \tag{3}$$

We should stress that formulas in Eqs. (2,3) make use of the information available in the finite data sequence: we observe only N data and with the expression of $P(\mathbf{n})$ and $P(\mathbf{x}|\mathbf{s})$ we

cannot take into account the effect of the past data points, which fall outside our observation window[45]. We shall see later that when considering a shorter analysis window (of size $N_{\|}$) contained in a longer data train we should actually exploit also the information contained outside the $N_{\|}$ window. This information is of little relevance if the analysis window is much longer than the largest correlation times in the noise: but this is not generally the case when considering burst events.

What matters in deciding what is the relevance of this boundary effect is the noise spectrum: one knows indeed (see Section A) that

$$R^{-1}\left(\frac{a-b}{f_s}\right) \equiv \left(\mathbf{R}_n^{-1}\right)_{ab} = \frac{2}{f_s N}\sum_{k=1}^{N-2}\frac{e^{i2\pi k(a-b)}}{S_n[k]}\,, \qquad (4)$$

and considering a noise model as given in Eq. (B16), which summarizes the best sensitivity reachable in the first generation interferometers, one deduces that

$$R^{-1}(\tau) \simeq A_0 e^{-|t|/\tau_0}\cos 2\pi f_0 \tau + \ldots \qquad (5)$$

neglecting faster decaying terms. The decay time(s) $\tau_{0,1,\ldots}$ characterize how much the matrix $\mathbf{R}^{-1}$ (and therefore $\mathbf{R}$ itself) differs from a diagonal matrix. In Section B 2 and in Table III we show that for the current models of the baseline interferometer noise in the first generation detectors the values of τ_0 range from 1.4 ms in TAMA to 6ms in Virgo (corresponding to O(100) samples).

These time scales should be compared with the expected duration of the bursts: for instance Zwerger and Müller [9, Figs. 5,6] have shown several examples of gravitational waveforms emitted in axisymmetric core collapse events, displaying large variations on scales of a few ms, and narrow large amplitude peaks even shorter than 1 ms. The same features are found in more recent simulations, which include relativistic effects, by Dimmelmeier *et al.*[14, Fig. 2 (Model A)]: we conclude that it is generally not justified from the physical point of view to surmise that the correlation decay times of the noise are short compared to the burst duration.

Moreover, the values for the decay times quoted in Table III refer to a ideal detector noise, free of narrow resonances: real detectors may well exhibit richer spectral features [22]. For instance a thermal resonance with proper frequency f_0 and quality factor Q would contribute to the noise correlation a term with a characteristic decay time $\tau_D = \frac{Q}{\pi f_0}$: with violin modes easily having $Q > 10^5$, and frequencies $f_0 = O(1000)$ Hz, the decay time can easily reach tens of seconds. Although it is foreseen to subtract the effect of these resonances from the data, using for instance Kalman filters [23] it is fair to say that any residual effect, due for instance to a imperfect cancellation of a very high Q resonance, will contribute to increase the noise correlation length above the values deduced from the baseline noise.

There are at least two important consequences of the presence of a non-zero correlation length: (1) statistics built using maximum likelihood criteria must be modified to take into account noise outside the window affected by the burst event; (2) instances of these statistics, derived from these data, will exhibit a correlation in time which will have to be kept into account when computing false alarm and false dismissal probabilities [24]. In the present paper we will concentrate on the first issue.

One general way, other than subtracting the narrow spectral components, to attack this problem would be to assume that data have been pre-whitened [6, 7, 8, 10], for instance using a time-domain filter estimated from the data themselves [25, 26]: however this strategy requires to take into account the effect of the whitening filter throughout the whole detection chain, in particular the alteration of the signal waveform and consequently of the signal distribution. In other words, when integrating over the space of possible signals, we are not allowed to ignore that the measure is changed by the transformation.

To render our statements more precise, we need to discuss in detail the detection framework adopted.

B. Detection framework

We suppose that, when present, the burst affects only a interval $T_{\|} = N_{\|}/f_s$ starting at absolute time $t_{\rm burst}$, that is a number $N_{\|}$ of samples in the data stream. We are unable to prescribe any prior for the signal amplitudes, so we treat it, at any given instant, as *nuisance* parameters, to be integrated out. Anderson *et al.* [11, 13] have made similar assumptions, but as we have already anticipated with a different hypothesis on the measure for this integration.

Given a data vector $\mathbf{x}$, the *a posteriori* probability of having observed those values can be written as[27]

$$P(\mathbf{x}) = P(\mathbf{x}|1)P(1) + P(\mathbf{x}|0)P(0) \qquad (6)$$

where as usual $P(1), P(0)$ are the *a priori* probabilities for a signal of unspecified form being present in the data, and $P(\mathbf{x}|1)$ is the probability of observing the dataset $\mathbf{x}$ given that *some* signal $\mathbf{s}$ is present, while $P(\mathbf{x}|0)$ is the probability of observing the same dataset $\mathbf{x}$ in absence of signals: it was defined in Eq. (2). In turn we have that

$$P(\mathbf{x}|1) \equiv \int P(\mathbf{x}|\mathbf{s})P(\mathbf{s})\,d\mathbf{s} \qquad (7)$$

where $P(\mathbf{s})$ is the *a-priori* probability of having a signal $\mathbf{s}$ present, while $P(\mathbf{x}|\mathbf{s})$ has already been defined in Eq. (3). We may or may not have a good guess for the distribution $P(\mathbf{s})$: in this paper we assume to have no *a priori* information.

Given our complete ignorance of $P(1), P(0)$, we resort to defining the (integrated) likelihood ratio that *any* signal is present as

$$\Lambda(\mathbf{x}) \equiv \frac{P(\mathbf{x}|1)}{P(\mathbf{x}|0)} = \int e^{-\frac{1}{2}\mathbf{s}\cdot\mathbf{R}_n^{-1}\cdot\mathbf{s}+\mathbf{s}\cdot\mathbf{R}_n^{-1}\cdot\mathbf{x}}P(\mathbf{s})d\mathbf{s}; \qquad (8)$$

in terms of Λ and using the Bayes rule $P(1|\mathbf{x})P(\mathbf{x}) = P(\mathbf{x}|1)P(1)$ we can write

$$P(1|\mathbf{x}) = \frac{\Lambda(\mathbf{x})}{\Lambda(\mathbf{x}) + P(0)/P(1)}\,; \qquad (9)$$

for the probability of having observed a signal, conditioned by the particular instance of data $\mathbf{x}$ that we have received. Although we have no idea of the priors $P(0), P(1)$, this probability is a monotonic function of the likelihood $\Lambda(\mathbf{x})$, which is therefore the quantity to be estimated, in dependence on the assumptions (or lack of assumptions) on $P(\mathbf{s})$: to implement our complete ignorance on the waveform we will assume that $P(\mathbf{s})$ is flat in the space $\mathbb{R}^{N_\parallel}$ of possible signals of length $N_\parallel$.

For the sake of comparison, Anderson *et al.* [13, Eq. 3.1, and the following discussion and note] have assumed that signals $\mathbf{s}$ are a priori distributed in a uniform way with respect to the metric induced by the scalar product $\langle x, y\rangle \equiv \mathbf{x}\cdot\mathbf{R}_n^{-1}\cdot\mathbf{y}$: this is the essential reason why our results are different from theirs.

C. Burst of unit duration (δ function)

Let us first consider a extreme example, namely a burst affecting only one of our data samples: its amplitude is unknown, and we want to test the hypothesis that the event occurs at the sample arbitrarily labeled a. Then the integrated likelihood ratio in Eq. (8) is

$$\begin{aligned}\Lambda(\mathbf{x}) &= \int_{-\infty}^{+\infty} e^{-\frac{1}{2}s_a(\mathbf{R}_n^{-1})_{aa}s_a + s_a(\mathbf{R}_n^{-1}\cdot\mathbf{x})_a}\, ds_a \\ &= \exp\left[\frac{1}{2}\frac{\left[(\mathbf{R}_n^{-1}\cdot\mathbf{x})_a\right]^2}{(\mathbf{R}_n^{-1})_{aa}}\right], \end{aligned} \tag{10}$$

where no summation over a is understood. The interesting statistic for a certain possible arrival index a is

$$L_a(\mathbf{x}) \equiv 2\ln\Lambda(\mathbf{x}) = \frac{f_s}{N}\frac{\left|\sum_k e^{i2\pi ak/N}\frac{\tilde{x}[k]}{S_n[k]}\right|^2}{\sum_k \frac{1}{S_n[k]}}, \tag{11}$$

and is readily identified as the Wiener filter for a δ function[15]. In absence of signals its distribution is

$$d_0(L) = \frac{1}{\sqrt{2\pi L}}e^{-\frac{1}{2}L} \tag{12}$$

as expected for the square of a Gaussian variable; note that the statistic L is *not* equivalent to taking the energy of the signal.

Nor is it equivalent to first whitening the data and then taking the energy: indeed whitening would mean, in matrix notation, to perform a lower-upper (LU) factorization of the matrix $\mathbf{R}^{-1}$ in the form[46]

$$\mathbf{R}_n^{-1} = \mathbf{W}_n^t\cdot\mathbf{W}_n; \tag{13}$$

where $\mathbf{W}_n$ is a lower triangular matrix which defines a causal transformation, dependent on the noise statistics: the matrix $\mathbf{W}_n$ defines a whitening transformation, namely the random vector $\mathbf{z} = \mathbf{W}_n\cdot\mathbf{x}$ is distributed as white gaussian noise. The statistic $L_a(\mathbf{x})$ can also be written in terms of the whitened data $\mathbf{z}$

$$L_a(\mathbf{x}) = \frac{\left[(\mathbf{W}_n^t\cdot\mathbf{z})_a\right]^2}{(\mathbf{W}_n^t\cdot\mathbf{W}_n)_{aa}}, \tag{14}$$

which can be regarded as a en*ergy*, but takes into account the altered signal shape (a δ function in this case) under the application of the transformation $\mathbf{W}_n$.

In passim it is worth noticing that we wrote the distribution $d_0(L)$ assuming implicitly to compute the moments of our statistic over a ensemble of noise realizations; in practice however one takes samples from a single time-series, at different locations in the data stream. This means that successive instances of the statistic L, for different indices a, will belong to a joint distribution, which does not factor in a product of terms like the χ^2 in Eq. 12: for instance we can consider the cumulant

$$\frac{E\left[L_a(\mathbf{x})L_b(\mathbf{x})\right]}{E\left[L_a(\mathbf{x})\right]E\left[L_b(\mathbf{x})\right]} - 1 = 2\left[\frac{(\mathbf{R}_n^{-1})_{ab}}{(\mathbf{R}_n^{-1})_{aa}}\right]^2, \tag{15}$$

which exposes that the statistic L_a as a function of the supposed burst event location a has a non-negligible cross correlation in colored noise. This should be always kept in mind when post-processing the results of the analysis, for instance when deriving limits on false alarm or false dismissal rates[24]: we will not elaborate further on this.

We now proceed to generalize the statistic L to a burst of arbitrary length.

D. General burst

It is useful to introduce a vector space notation: let us call $\mathcal{V}_N$ the vector space of all possible data vectors, having length N; we are willing to test for the hypothesis of presence of a burst signal in a certain subspace $\mathcal{V}_\parallel$, defined by taking $N_\parallel \ll N$ consecutive samples, from a certain starting position (say, l) in the original vector. The orthogonal subspace, of dimension $N - N_\parallel \sim N$, will be called $\mathcal{V}_\perp$.

We have again the likelihood ratio

$$\Lambda(\mathbf{x}) = \int e^{-\frac{1}{2}s_i(\mathbf{R}_n^{-1})_{ij}s_j + s_i(\mathbf{R}_n^{-1}\cdot\mathbf{x})_i}\prod_i^{N_\parallel} ds_i \tag{16}$$

where the indices i, j run only over elements of $\mathcal{V}_\parallel$; however the noise correlation matrix $\mathbf{R}_n$ is defined for an arbitrary index difference. Let us introduce the matrix

$$(\mathbf{R}_n^{-1})_\parallel \equiv (\mathbf{R}_n^{-1})_{[l:l+N_{burst},\, l:l+N_{burst}]} \tag{17}$$

obtained restricting the indices of the $\mathbf{R}_n^{-1}$ matrix on the $[l, l+N_{burst}]$ interval: it acts only on the $\mathcal{V}_\parallel$ subspace.

Performing the Gaussian integrals over amplitudes s_i of the signal, which we treat as *nuisance* parameters, we obtain

$$\Lambda(\mathbf{x}) \propto \exp\left[\frac{1}{2}(\mathbf{R}_n^{-1}\cdot\mathbf{x})_\parallel \cdot \left(\left(\mathbf{R}_n^{-1}\right)_\parallel\right)^{-1}\cdot(\mathbf{R}_n^{-1}\cdot\mathbf{x})_\parallel\right] = \exp\left[\frac{1}{2}x_\alpha(\mathbf{R}_n^{-1})_{\alpha i}\left(\left(\left(\mathbf{R}_n^{-1}\right)_\parallel\right)^{-1}\right)_{ij}(\mathbf{R}_n^{-1})_{j\beta}x_\beta\right] \tag{18}$$

where indices α,β run in $\mathcal{V}_\parallel+\mathcal{V}_\perp$, and indices i,j run in $\mathcal{V}_\parallel$; the overall normalization does not depend on $\mathbf{x}$ and can be ignored.

We are forced to this somewhat cumbersome notation because the operations of projecting over the "burst" subspace $\mathcal{V}_\parallel$ and of inverting a matrix do not commute. They do only when $N=N_\parallel$: in that case one would have

$$\Lambda(\mathbf{x}) \propto e^{\frac{1}{2}\mathbf{x}\cdot\mathbf{R}_n^{-1}\cdot\mathbf{x}} = e^{\frac{1}{2}(\mathbf{W}_n\cdot\mathbf{x})^2} \tag{19}$$

where we used Eq. (13), and $\mathbf{W}_n\cdot\mathbf{x}$ is distributed as white gaussian noise. This means that in the case $N=N_\parallel$ the likelihood would be just a monotonic function of the *energy* $(\mathbf{W}_n\cdot\mathbf{x})^2$ of the whitened sample. Note however that this example is very different from the previous one: there we had a unit burst in a long data train, here we would have a burst as long as the data train: both are extreme cases.

We will from now on instead take as optimal statistic for a unknown burst, with flat prior for the sample amplitudes, the expression

$$L(\mathbf{x}) \equiv \sum_{i,j\in\mathcal{V}_\parallel}(\mathbf{R}_n^{-1}\mathbf{x})_i\left(\left(\left(\mathbf{R}_n^{-1}\right)_\parallel\right)^{-1}\right)_{ij}(\mathbf{R}_n^{-1}\mathbf{x})_j\,. \tag{20}$$

The reader might wonder if this is of any practical use. In particular, the estimation of the matrix $\left(\left(\mathbf{R}_n^{-1}\right)_\parallel\right)^{-1}$ looks awkward: we should estimate $\mathbf{R}_n$, then compute $\mathbf{R}_n^{-1}$, then take a minor $N_\parallel\times N_\parallel$ along the matrix diagonal and finally invert it[47].

Let us however notice the trivial identity

$$\begin{aligned} E\left[\left(\mathbf{R}_n^{-1}\mathbf{n}\right)_i\left(\mathbf{R}_n^{-1}\mathbf{n}\right)_j\right] &= \sum_{\alpha\beta}\left(\mathbf{R}_n^{-1}\right)_{i\alpha}\left(\mathbf{R}_n^{-1}\right)_{j\beta}E\left[n_\alpha n_\beta\right] \\ &= \left(\mathbf{R}_n^{-1}\right)_{ij} \end{aligned} \tag{21}$$

in other words, considering the time series $\mathbf{R}_n^{-1}\cdot\mathbf{n}$, which is $\mathbf{R}_n^{-1}\cdot\mathbf{x}$ in absence of signal, and restricting it to the $\mathcal{V}_\parallel$ subspace, we obtain an autocorrelation matrix that is just $\mathbf{R}_n^{-1}$. Barring boundary effects on the matrix $\mathbf{R}_n$, also the process $\mathbf{R}_n^{-1}\cdot\mathbf{n}$ is shift invariant, and $\mathbf{R}_n^{-1}$ is a Toeplitz matrix when considering diagonal minors sufficiently far from the borders.

We are therefore able to easily compute $\mathbf{R}_n^{-1}$ just from the analysis of the series $\mathbf{R}_n^{-1}\cdot\mathbf{n}$: the last step, in order to write down the statistic for a burst in a certain subspace $\mathcal{V}_\parallel$, is to restrict the matrix to that interval and invert it: a efficient tool to accomplish this task is the Karhunen-Loève expansion[17], which we find useful to recall briefly in the next section.

E. Karhunen-Loève expansion

In Section II C we have exploited the LU factorization of a correlation matrix $\mathbf{R}$. It is also well known [16, 17, 21] that, being $\mathbf{R}$ real symmetric and positive definite, an expansion in terms of its eigenvalues and eigenvectors exists, namely

$$R_{\alpha\beta} = \sum_{k=1}^{K}\sigma_k\psi_\alpha^k\psi_\beta^k \tag{22}$$

where K is the dimension of the matrix and

$$R_{\alpha\beta}\psi_\beta^k = \sigma_k\psi_\beta^k \tag{23}$$

with eigenvalues $\sigma_k>0$. The $\left\{\boldsymbol{\psi}^k, k\in[1,K]\right\}$ eigen-vectors are chosen orthonormal

$$\sum_\alpha\psi_\alpha^k\psi_\alpha^l = \delta^{kl}; \tag{24}$$

and define a basis in the space $\mathbb{R}^K$: any data vector $\mathbf{x}$ can be written as

$$\mathbf{x} = \sum_{k=0}^{K-1}c_k\boldsymbol{\psi}^k,\ \text{where } c_k\equiv\mathbf{x}\cdot\boldsymbol{\psi}^k; \tag{25}$$

this decomposition is called discrete Karhunen-Loève transform (DKLT). As with the Fourier transform, the Parseval's theorem holds

$$\mathbf{x}\cdot\mathbf{x} = \sum_{k=1}^{K}c_k^2 \tag{26}$$

and it is immediate to show that

$$E\left[c_k c_l\right] = \sigma_k\delta_{kl} \tag{27a}$$

$$E\left[\mathbf{x}\cdot\mathbf{x}\right] = \sum_{\alpha=1}^{K}R_{\alpha\alpha} = \sum_{k=1}^{K}\sigma_k\,. \tag{27b}$$

The similarity with the Fourier transform goes further: it can be shown (see for instance [17, sec. 4.7.2]) that in the limit of large K the basis elements converge to sines and cosines, and the eigenvalues converge to the corresponding bins of the spectral density.

However, for finite K the DKLT is a better representation for the noise because it takes into account the finite-size effects [18]: recall that we are interested in $N_\parallel$ not necessarily large. In particular, the coefficients c_k are uncorrelated random variables, thus making the statistical analysis easier.

We are naturally led to apply the DKLT to the problem at hand.

F. Exact expression for the burst statistic

In Section II D we have shown that the exact statistic, defined in Eq. (20), can be expressed a

$$L = \mathbf{y}_{\parallel} \cdot \left[(\mathbf{R}_y)_{\parallel} \right]^{-1} \cdot \mathbf{y}_{\parallel} \tag{28}$$

where $\mathbf{y}_{\parallel} \equiv \left(\mathbf{R}_n^{-1} \cdot \mathbf{x} \right)_{\parallel}$.

The vector $\mathbf{y}$ itself can be easily computed: recall that we have assumed N large, hence we can write

$$y[l] = \frac{1}{N} \sum_{k=0}^{N-1} \frac{2}{S_n[k]} e^{-i2\pi k l/N} \tilde{x}[k]\,; \tag{29}$$

where S_n is the one-sided spectrum corresponding to the correlation matrix $\mathbf{R}_n$.

Thanks to the KL transform we are now also able to write down explicitly the inverse of the correlation matrix $\mathbf{R}_y$ of the $\mathbf{y}$ time series, keeping into account the restriction to the $\mathcal{V}_{\parallel}$ space, in terms of a appropriate DKL basis $\boldsymbol{\psi}_{\parallel}^k$, $k \in \left[1, N_{\parallel}\right]$, as

$$\left[(\mathbf{R}_y)_{\parallel} \right]^{-1} = \sum_{k=1}^{N_{\parallel}} \frac{1}{\sigma_k} \boldsymbol{\psi}_{\parallel}^k \otimes \boldsymbol{\psi}_{\parallel}^k; \tag{30}$$

hence we finally have

$$L = \sum_{k=1}^{N_{\parallel}} \frac{1}{\sigma_k} \left(\boldsymbol{\psi}_{\parallel}^k \cdot \mathbf{y}_{\parallel} \right)^2. \tag{31}$$

Note that the matrix $\mathbf{R}_y$, as we have seen in Section II D, is approximately a Toeplitz matrix: this means that $(\mathbf{R}_y)_{\parallel}$ does not depend on the segment, in the data train, chosen to test for the presence of bursts[48].

The reader might wonder why the DKLT is at all necessary: our statistic in Eq. (28) could be computed just inverting the matrix $(\mathbf{R}_y)_{\parallel}$, and then applying it to each successive data chunk; what is the advantage of the expression in Eq. (31)? The answer is that the computational cost is the same, but the DKLT decomposition gives us more flexibility. We are not forced to sum over all the elements: we can decide for instance that some of the basis elements correspond to large noise components, and can be left out without significantly affecting the detector performance. The fact that the coefficients $\boldsymbol{\psi}_{\parallel}^k \cdot \mathbf{y}_{\parallel}$, for different values of k, are by definition statistically uncorrelated renders this procedure sound and does not complicate the statistical analysis .

We will elaborate more on the practical implementation in Section II H: we now turn to consider two special cases which help building a better understanding.

1. *A special case: $N = N_{\parallel}$*

This is unrealistic: in this case there is no orthogonal space, $\left(\mathbf{R}_n^{-1}\right)_{\parallel} = \mathbf{R}_n^{-1}$ and the likelihood ratio becomes

$$\Lambda(\mathbf{x}) = \exp\left[\frac{1}{2} \mathbf{x} \cdot \mathbf{R}_n^{-1} \cdot \mathbf{x} \right]\,; \tag{32}$$

if $\left\{ \boldsymbol{\phi}^k, k = 1, N \right\}$ is the appropriate DKL basis for the noise correlation matrix $\mathbf{R}_n$, that is

$$\mathbf{R}_n = \sum_{k=1}^{N} \sigma_k \boldsymbol{\phi}^k \otimes \boldsymbol{\phi}^k\,; \tag{33}$$

one would therefore use as a statistic for burst searches

$$L = \sum_k \frac{1}{\sigma_k} \left(\boldsymbol{\phi}^k \cdot \mathbf{x} \right)^2. \tag{34}$$

This expression corresponds closely to the excess energy statistic defined in[13, Eq. (1.4)], with two differences: (1) by using the DKL expansion it takes into account the finite size of the sample; (2) it is shown to be appropriate, given our assumptions on the signal distribution, only in this very special case when no information is available *before* and *after* the data segment we are searching for a burst. The correspondence makes more evident why our result differs from theirs: having they chosen a uniform prior with the noise metric, they have effectively decoupled the $\mathcal{V}_{\perp}$, $\mathcal{V}_{\parallel}$ subspaces, which is equivalent in our context to neglecting the presence of correlated noise.

2. *Approximate expression: $N_{\parallel}$ large*

Let us go back to the case in which we search for a burst of length $N_{\parallel}$ in a much longer data train of length N: as we said, sufficiently long to resolve the narrow spectral features which give rise to long correlation times. The high sampling rate $O(20\,\mathrm{kHz})$ needed to exploit the wide spectral range available in interferometer data may lead to consider $N_{\parallel}$ of the order of several hundred samples, even for signals of a few tens of ms.

If this results in a excessive computational cost in the application of the KL transform, we can exploit its convergence to the Fourier Transform, in the limit of $N_{\parallel}$ large: namely, we can write down approximately

$$\left[(\mathbf{R}_y)_{\parallel} \right]^{-1} \simeq f_s^{-1} \sum_{k=1}^{N_{\parallel}-2} \frac{2}{S_y[k]} \mathbf{w}_k \otimes \mathbf{w}_k^H \tag{35}$$

where $\mathbf{w}_k$ are the Fourier basis vectors (see Section A) in $\mathcal{V}_{\parallel}$, approximating the DKL transform for $(\mathbf{R}_y)_{\parallel}$ with eigenvalues $\lambda_k = \frac{1}{2} S_y[k]\ f_s$.

It follows the expression for the approximate statistic

$$\begin{aligned} L(\mathbf{x}) &\simeq f_s^{-1} \sum_{k=1}^{N_{\parallel}-2} \frac{2}{S_y[k]} \left| \mathbf{w}_k \cdot \mathbf{y}_{\parallel} \right|^2 \\ &= \frac{f_s}{N_{\parallel}} \sum_{k=1}^{N_{\parallel}/2-1} \frac{1}{S_y[k]} \left| \tilde{y}_{\parallel}[k] \right|^2; \end{aligned} \tag{36}$$

$L(\mathbf{x})$ is the sum of the squares of the Fourier coefficients $\tilde{y}_{\parallel}[k] \equiv f_s^{-1} \sqrt{N_{\parallel}}\, \mathbf{w}_k \cdot \mathbf{y}_{\parallel}$ of the time series $\mathbf{y}_{\parallel} \equiv \left(\mathbf{R}_n^{-1} \cdot \mathbf{x} \right)_{\parallel}$, weighted with the corresponding spectral noise density. This expression is similar to the excess energy statistic defined in[11, 13]: the difference, apart the $N_{\parallel}$ large approximation, is that we found it necessary to start from data filtered for the occurrence of δ.

G. In which sense the statistic is "optimal"

Before moving to the practical implementation of the algorithm, it is very important to fully understand the consequences of the assumption we have made about the *a priori* distribution of the signals: to this end, let us have a second look at the likelyhood ratio

$$\Lambda(\mathbf{x}) = \int e^{-\frac{1}{2}\mathbf{s}\cdot\mathbf{R}_n^{-1}\cdot\mathbf{s}+\mathbf{s}\cdot\mathbf{R}_n^{-1}\cdot\mathbf{x}} P(\mathbf{s}) d\mathbf{s}. \tag{37}$$

Our choice has been that $P(\mathbf{s}) = 1$, which is a way to avoid introducing any scale in the problem, which might bias the analysis. There is however a drawback: larger values of $\mathbf{s}\cdot\mathbf{s}$ are favored, in fact

$$d\mathbf{s} = \rho^{N_\parallel - 1} d\rho d\Omega_{N_\parallel}(\hat{\mathbf{s}}) \tag{38}$$

where $\rho \equiv \sqrt{\mathbf{s}\cdot\mathbf{s}}$ and $d\Omega_{N_\parallel}$ is the solid angle element in $N_\parallel$ dimensions. If we have (say) *a priori* information only on the energy of the signal, or equivalently on the distribution $p(\rho)$, then we would like to follow the same approach as in [13, Sec. III] and write

$$\Lambda(\mathbf{x}) = \int p(\rho)\Lambda(\mathbf{x}|\rho) d\rho \tag{39}$$

where

$$\Lambda(\mathbf{x}|\rho) \equiv \int \delta\left(\rho - \sqrt{\mathbf{s}\cdot\mathbf{s}}\right) e^{-\frac{1}{2}\mathbf{s}\cdot\left(\mathbf{R}_n^{-1}\right)_\parallel\cdot\mathbf{s}+\mathbf{s}\cdot\mathbf{y}_\parallel} d\mathbf{s} \tag{40}$$

and $\mathbf{y} = \mathbf{R}_n^{-1}\cdot\mathbf{x}$ belongs to the parallel space $\mathcal{V}_\parallel$. Changing basis with a DKL transform $\mathbf{s} \to \boldsymbol{\varsigma}$, we obtain

$$\begin{aligned}\Lambda(\mathbf{x}|\rho) &= \int \delta(\rho - \sqrt{\boldsymbol{\varsigma}\cdot\boldsymbol{\varsigma}}) e^{-\frac{1}{2}\sum_k \sigma_k \varsigma_k^2 + \boldsymbol{\varsigma}\cdot\mathbf{c}} d\boldsymbol{\varsigma} \\ &= \int e^{-\frac{\rho^2}{2}\sum_k \sigma_k^2 \hat{\varsigma}_k^2 + \rho\sum_k \hat{\varsigma}_k c_k} d\Omega_{N_\parallel}(\hat{\varsigma})\end{aligned} \tag{41}$$

where $c_k \equiv \boldsymbol{\psi}_k\cdot\mathbf{y}_\parallel$ and σ_k are the eigenvalues of $\left(\mathbf{R}_n^{-1}\right)_\parallel$. If we had σ_k independent on k, in other words if all the directions in the $\boldsymbol{\varsigma}$ space were equivalent, we could as in [13, Sec. III] compute the integral in closed form, by aligning one of the axes with the direction of the vector $\mathbf{c}$. This was possible to Anderson *et al.* because they had chosen a signal prior function of $\mathbf{s}\cdot(\mathbf{R}_n)_\parallel^{-1}\cdot\mathbf{s}$.

In the case considered by us the expression in Eq.(41) is not in general a function solely of the statistic $L = \sum_k \sigma_k^{-1} c_k^2$ and of ρ, because the noise introduces preferential directions in the space of possible signals.

This discussion shows that the statistic L we have proposed is strictly speaking *optimal* only for a signal prior $P(\mathbf{s})$ constant[49]: we can make the *ansatz* that for a more general prior, depending on a scalar function of $\mathbf{s}$, the optimal statistic L might still be of the form

$$L = \sum_{k,l} \mathsf{L}_{kl} \frac{c_k c_l}{\sqrt{\sigma_k \sigma_l}} \tag{42}$$

where the matrix L should be determined maximizing the probability of detection while keeping the false alarm rate fixed. We were however unable to prove that this is actually the case, at least under certain restrictions on the form of $P(\mathbf{s})$.

Another implication of this discussion is that with our statistic it is difficult to set Bayesian thresholds, as derived by choosing a particular form for $p(\rho)$, and making assumptions on its parameters. This is however a somewhat less crucial issue because thresholds may be set following a frequentist approach, that is by limiting the false alarm rate, as shall be discussed in Section II I.

H. Description of the algorithm

The expressions in Eqs. (20,31,36) define the algorithm for estimating the (log)-likelihood, at different approximation levels: we find useful to detail the procedure.

We assume that the starting point is a continuous stream of data, at sampling frequency f_s, whose Gaussian noise component is assumed to be stationary. The purpose of the algorithm is to search the stream for bursts of length $N_\parallel$ and unknown shape; the choice of $N_\parallel$ is arbitrary if there is no physical hint.

The steps of the algorithm are the following:

1. partition the data stream into data vectors $\mathbf{x}$ of length $N \gg N_\parallel$; each vector should overlap the following by a certain amount $2M$ which needs tuning (it is related to boundary effects on the correlation matrix, which depend on the specific noise considered). The length N must be sufficient to resolve the narrow spectral features in the data, or in other words to ensure that the correlation function $R_n(l/f_s) = E[x[i]x[i+l]]$ is sampled over a sufficiently long interval, allowing it to decrease to zero with sufficient accuracy.

2. Estimate the correlation matrix, or equivalently the *sample* spectrum $S_n[k]$, over the $\mathbf{x}$ segments: that is, with frequency resolution $df = \frac{f_s}{N}$. For instance using a Welch's overlap and save procedure to combine estimates from different vectors of length N, or averaging estimates obtained with a multi-taper spectral estimator[28]. The cost of this step, per vector, is $O(N\ln N)$.

3. For each data vector $\mathbf{x}$, estimate a new vector $\mathbf{y} \equiv \mathbf{R}_n^{-1}\cdot\mathbf{x}$; this is equivalent, as we have seen in Section II C, to filter for δ functions. It can be done at least in two ways, not necessarily equivalent from the numerical point of view:

   (a) in the frequency domain, Fourier transforming the data $\mathbf{x} \to \tilde{\mathbf{x}}$ and defining

   $$y[l] = \frac{1}{N}\sum_{k=0}^{N-1} \frac{2}{S_n[k]} e^{-i2\pi kl/N} \tilde{x}[k]; \tag{43}$$

   the computational cost is again $O(N\ln N)$.

   (b) In the time domain, first performing a spectral factorization $\mathbf{R}_n^{-1} = \mathbf{W}_n^{\mathbf{t}}\cdot\mathbf{W}_n$ which defines two matrix operators $\mathbf{W}_n$ and $\mathbf{W}_n^{\mathbf{t}}$, the first causal and

the other anti-causal. Both can be implemented as digital filters: the anti-causal one after reversing the input data. Several methods exist to perform such a factorization and to apply the resulting filters to the data [16, 17]: see [25, 26] for an application to the analysis of data from the Caltech 40m detector [22]. If the noise is stationary across several data vectors, the computational cost of the spectral factorization itself becomes negligible: otherwise, it can be shown [17] that the cost of estimating filters with P coefficients, using for instance the Levinson-Durbin recursion, grows as P^2, where the right order depends on the spectral characteristics of the data (see [26], where it has been shown that FIR filters a few hundred taps are sufficient for correctly whitening real interferometer data). Once the filters are estimated, they can be applied with a cost $O(NP)$. One possible advantage of these time-domain methods, which are generally slower than the FFT (but can be implemented very efficiently on DSP systems) is that they can follow a slow noise non-stationarity: this issue however requires more study, because the very definition of the "burst" statistic needs a revision, when the noise is not stationary.

4. Drop the first and last M points in the data vector $\mathbf{y}$, in order to reduce the boundary effects, and partition it in segments $\mathbf{y}_\parallel$ of length $N_\parallel$. The partitioning will require to overlap the $\mathbf{y}_\parallel$ segments: in any case one will have to take into account the correlation of the resulting statistics.

5. Use the $N-2M$ data in $\mathbf{y}$ to estimate the correlation matrix $\mathbf{R}_y$ over lags of at most $N_\parallel$: this operation can be done using DFT methods with a cost at most $O(N \times \ln N)$, or smaller if we can exploit the fact that only $(\mathbf{R}_y)_\parallel$ is needed.

6. Decompose of the matrix $(\mathbf{R}_y)_\parallel$, using either the Karhunen-Loève or the Fourier transform.

   (a) In the case of the DKLT, one needs the eigenvalues σ_k and eigenvectors $\boldsymbol{\psi}^k_\parallel$ of the matrix

   $$(\mathbf{R}_y)_\parallel = \sum_{k=1}^{N_\parallel} \sigma_k \, \boldsymbol{\psi}^k_\parallel \otimes \boldsymbol{\psi}^k_\parallel \tag{44}$$

   this decomposition has to be done at most once for each $\mathbf{y}$ vector, and possibly even more rarely, depending on the noise stationarity. Generally the cost of this decomposition is $O(N_\parallel^3)$, unless one is able to exploit the Toeplitz structure of the matrix $\mathbf{R}_y$. One possibly selects then only a subset of the eigenvectors, chosen for instance by setting a threshold on the value of the eigenvalues σ_k. This subset consists of $N_{KL} \leq N_\parallel$ elements, to be used in the actual evaluation of the statistic.

   (b) With the Fourier transform method (provided it is a reasonable approximation) we need instead only the spectrum of the time series $\mathbf{y}$, at a reduced frequency resolution $df_\parallel \equiv f_s/N_\parallel$: we call this spectrum $S_y[k]$, $k \in [0, N_\parallel - 1]$: it can also be computed by averaging the bins of the $(S_n)^{-1}$, if the noise is stationary across instances of the vector $\mathbf{x}$.

7. For each segment $\mathbf{y}_\parallel$ assemble the statistic for the log likelihood, following a recipe dependent on the chosen method:

   (a) using the DKLT one computes

   $$L = \sum_{k=1}^{N_{KL}} \frac{1}{\sigma_k} \left[\boldsymbol{\psi}^k \cdot \mathbf{y}_\parallel \right]^2, \tag{45}$$

   with a cost $O(N_{KL} N_\parallel) \leq O(N_\parallel^2)$ in floating-point operations: this recipe gives the exact result.

   (b) Using the Fourier transform one instead first computes the Fourier coefficients $\tilde{\mathbf{y}}_\parallel[k]$, a operation costing $O(N_\parallel \times \ln N_\parallel)$: then one combines the outputs in the statistic

   $$L = \frac{f_s}{N_\parallel} \sum_{k=1}^{N_\parallel/2-1} \frac{1}{S_y[k]} \left| \tilde{\mathbf{y}}_\parallel[k] \right|^2; \tag{46}$$

   this expression is approximate, but its accuracy increases with $N_\parallel$, and can be legitimate in some cases.

8. Perform the statistical analysis of the results: we will discuss about this step in Section III.

A few comments are in order:

- even though the procedure is designed to cope with long correlations in the data, we have implicitly assumed that no *deterministic* lines are present: in fact, the Wold Decomposition theorem[17, sec. 7.6.2] states that a general random process can be decomposed in the sum

  $$x(t) = x_r(t) + x_p(t) \tag{47}$$

  of a *regular* process[50] $x_r(t)$ and a *predictable* process[51] $x_p(t)$. The latter could correspond to a harmonic of the power line: it would contribute to the spectrum a term

  $$\sigma_p^2 \delta(\nu - \nu_p) \tag{48}$$

  where ν_p is the frequency of the line and σ_p its contribution to the RMS noise. In the sample spectrum $S_n[k]$ this feature would translate approximately into a term

  $$\frac{N}{f_s} \delta_{k,k_p} = T \, \delta_{k,k_p}, \tag{49}$$

a trend in T which is the symptom of a infinite correlation length. Such spectral features cannot be properly handled by spectral factorization methods, because they are deterministic components and not stochastic. It is advisable to subtract altogether such predictable components, because this operation can certainly improve the signal to-noise ratio : several examples are available in literature [29, 30]. Notice *in passim* that violin mode lines excited by thermal noise belong to the class of regular processes, although they can also be modeled and partially subtracted [23].

- If the noise is not stationary, we can still perform an adaptive whitening, and refresh the estimate of the DKL basis $\{\boldsymbol{\psi}^k\}$ each time it is found necessary. However as we have anticipated, further study is necessary; for instance, the very definition of $\mathbf{R}_y$ as an average becomes uncertain when one cannot trade ensemble averaging for time averaging. In case of a slow non-stationarity one may also think of using an adaptive estimator of the δ filter in Eq. (43); this is a good topic for future research work.

- Whether it is necessary to adopt the exact statistic in Eq. (45), or the approximate one in Eq. (46), can only be judged comparing the receiver performance in a definite situation.

We turn out to the analysis of the distribution of our statistic for burst detection.

I. Statistical analysis

Given the statistic L we need the distributions $d_0(L)$ and $d_1(L|\mathbf{s})$ under the respective hypotheses H_0 (no signal) and H_1 (a signal $\mathbf{s}$ of unspecified form), in order to set up a detection strategy, for instance based on the Neyman-Pearson criterion [31].

The distribution $d_0(L)$ is immediately recognized to be a $\chi^2(N_\parallel)$; we find illustrative to prove this directly, from the definition

$$\begin{aligned} d_0(L) &\propto \int e^{-\frac{1}{2}\mathbf{n}\cdot\mathbf{R}_n^{-1}\cdot\mathbf{n}} \\ &\times\ \delta\left(L-\left(\mathbf{R}_n^{-1}\mathbf{n}\right)_\parallel\left(\left(\mathbf{R}_n^{-1}\right)_\parallel\right)^{-1}\left(\mathbf{R}_n^{-1}\mathbf{n}\right)_\parallel\right)d\mathbf{n} \\ &\propto \int e^{-\frac{1}{2}\mathbf{y}\left(\mathbf{R}_y\right)^{-1}\mathbf{y}}\delta\left(L-\mathbf{y}_\parallel\left(\left(\mathbf{R}_y\right)_\parallel\right)^{-1}\mathbf{y}_\parallel\right)d\mathbf{y}_\parallel d\mathbf{y}_\perp \end{aligned} \tag{50}$$

and observing that

$$\mathbf{R}_y^{-1}=\begin{pmatrix}(\mathbf{R}_y)_\parallel & \mathbf{B}\\ \mathbf{B}^t & (\mathbf{R}_y)_\perp\end{pmatrix}^{-1}=\begin{pmatrix}\mathbf{A} & \mathbf{D}\\ \mathbf{D}^t & \mathbf{C}\end{pmatrix} \tag{51}$$

where

$$\begin{aligned} \mathbf{A} &= \left((\mathbf{R}_y)_\parallel\right)^{-1}+ \\ &\quad+\left[\left((\mathbf{R}_y)_\parallel\right)^{-1}\mathbf{B}\mathbf{C}\mathbf{B}^t\left((\mathbf{R}_y)_\parallel\right)^{-1}\right] && (52a)\\ \mathbf{D} &= -\left((\mathbf{R}_y)_\parallel\right)^{-1}\mathbf{B}\mathbf{C} && (52b) \end{aligned}$$

and the explicit form of $\mathbf{C}$ is not relevant for us. Integrating over $\mathbf{y}_\perp$ one obtains a factor

$$\begin{aligned} &\int d\mathbf{y}_\perp e^{-\frac{1}{2}\mathbf{y}_\perp\mathbf{C}\mathbf{y}_\perp+\mathbf{y}_\perp\mathbf{C}\mathbf{B}^t\left((\mathbf{R}_y)_\parallel\right)^{-1}\mathbf{y}_\parallel}\ \propto \\ &\qquad e^{\frac{1}{2}\left[\left((\mathbf{R}_y)_\parallel\right)^{-1}\mathbf{B}\mathbf{C}\mathbf{B}^t\left((\mathbf{R}_y)_\parallel\right)^{-1}\right]}\quad ; \end{aligned} \tag{53}$$

comparing with Eqs. (51,52a), the term in square parentheses is recognized to be $\left(\mathbf{R}_y^{-1}\right)_\parallel-\left((\mathbf{R}_y)_\parallel\right)^{-1}$, hence

$$\begin{aligned} d_0(L) &\propto \int e^{-\frac{1}{2}\mathbf{y}_\parallel\left((\mathbf{R}_y)_\parallel\right)^{-1}\mathbf{y}_\parallel} \\ &\quad\times\delta\left(L-\mathbf{y}_\parallel\left((\mathbf{R}_y)_\parallel\right)^{-1}\mathbf{y}_\parallel\right)d\mathbf{y}_\parallel \\ &= \frac{L^{N_\parallel/2-1}}{2^{N_\parallel/2}\Gamma(N_\parallel/2)}e^{-L/2} \end{aligned} \tag{54}$$

which is as expected the distribution of a $\chi^2(N_\parallel)$ [16, Sec. 4-3]: we underline the role of the cross correlation effects, in order to derive the correct result.

When on the contrary a signal is present, the distribution can be written as

$$\begin{aligned} d_1(L|\mathbf{s}) &= \mathcal{N}\int e^{-\frac{1}{2}\mathbf{n}\cdot\mathbf{R}_n^{-1}\cdot\mathbf{n}}\delta\left(L-\left(\mathbf{R}_n^{-1}\cdot(\mathbf{n}+\mathbf{s})\right)_\parallel\cdot\left(\left(\mathbf{R}_n^{-1}\right)_\parallel\right)^{-1}\cdot\left(\mathbf{R}_n^{-1}\cdot(\mathbf{n}+\mathbf{s})\right)_\parallel\right)d\mathbf{n} \\ &= \mathcal{N}\int e^{-\frac{1}{2}\left(\mathbf{y}-\mathbf{R}_n^{-1}\cdot\mathbf{s}\right)_\parallel\cdot\left((\mathbf{R}_y)_\parallel\right)^{-1}\cdot\left(\mathbf{y}-\mathbf{R}_n^{-1}\cdot\mathbf{s}\right)_\parallel}\delta\left(L-\mathbf{y}_\parallel\cdot\left((\mathbf{R}_y)_\parallel\right)^{-1}\cdot\mathbf{y}_\parallel\right)d\mathbf{y}_\parallel \end{aligned} \tag{55}$$

which immediately results in the non-central $\chi^2\left(N_\parallel\right)$: in terms of the signal-to-noise ratio (SNR) one has

$$d_1(L|\mathsf{SNR}) = \frac{L^{N_\parallel/2-1}}{2^{N_\parallel/2}\Gamma(N_\parallel/2)} e^{-\frac{1}{2}\left(L+\sqrt{2N_\parallel}\mathsf{SNR}\right)} {}_0\mathrm{F}_1\left(;\frac{N_\parallel}{2};\frac{\mathsf{SNR}\,L\sqrt{2N_\parallel}}{4}\right), \tag{56}$$

where ${}_0F_1$ is a hyper-geometric function[32, sec. 9.14].[52]

The SNR explicit form is

$$\mathsf{SNR} = \frac{\left(\mathbf{R}_n^{-1}\cdot\mathbf{s}\right)_\parallel \cdot \left(\left(\mathbf{R}_y\right)_\parallel\right)^{-1} \cdot \left(\mathbf{R}_n^{-1}\cdot\mathbf{s}\right)_\parallel}{\sqrt{2N_\parallel}} \tag{57}$$

consistently with the general definition [21, chap. 6]:

$$\mathsf{SNR} \equiv \frac{|E\left[L|H_1\right] - E\left[L|H_0\right]|}{\sqrt{E\left[\left(L-E\left[L|H_0\right]\right)^2|H_0\right]}}, \tag{58}$$

where $E\left[L|H\right]$ is the expectation value of the statistic L under the hypothesis H, and H_1, H_0 correspond respectively to the hypotheses of presence of absence of a signal. Please note that the SNR defined in Eq. (58) has nothing to do with the *intrinsic* signal-to-noise ratio which would result from a matched filter procedure, $\mathsf{SNR}_{\mathsf{intrinsic}} \propto \sqrt{\int |\tilde{s}(f)|^2/S_n(f)\,df}$; in particular, our SNR is quadratic in the signal amplitude.

Given the distribution $d_0\left(L\right)$ the false alarm probability can be readily computed as

$$Q_f(L_0) \equiv \int_{L_0}^{\infty} d_0\left(L\right)dL = \frac{\Gamma\left(\frac{N_\parallel}{2},\frac{L_0}{2}\right)}{\Gamma\left(\frac{N_\parallel}{2}\right)}, \tag{59}$$

while the detection probability

$$Q_d\left(L_0|\mathsf{SNR}\right) \equiv \int_{L_0}^{\infty} d_1\left(L|\mathsf{SNR}\right)dL \tag{60}$$

cannot be written in closed form. For large SNR it is possible to approximate $d_1\left(L|\mathsf{SNR}\right)$ with a Gaussian distribution having the same first and second order momenta

$$d_1(L|\mathsf{SNR}) \simeq \frac{e^{-\frac{\left(L-(\mathsf{SNR}\sqrt{2N_\parallel}+N_\parallel)\right)^2}{4(N_\parallel+2\mathsf{SNR}\sqrt{2N_\parallel}}}}{\sqrt{4\pi(N_\parallel+2\mathsf{SNR}\sqrt{2N_\parallel})}} \tag{61}$$

obtaining

$$Q_d \simeq \frac{1}{2}\left[1+\mathrm{erf}\left(\frac{N_\parallel+\mathsf{SNR}\sqrt{2N_\parallel}-L_0}{2\sqrt{N_\parallel+2\mathsf{SNR}\sqrt{2N_\parallel}}}\right)\right]. \tag{62}$$

The expressions for Q_f and Q_d are the building blocks for setting up the detection strategy: they are appropriate if we are able to compute the exact statistic L. How should however d_0 and d_1 be modified, if we have instead chosen to compute an approximate statistic?

1. *Approximate statistic distribution*

Recall that we have discussed in Section II F also the possibility of defining the statistic L, when computed using the DKL expansion, using just a subset $N_{KL} < N_\parallel$ of the basis vectors: assuming to have ordered the $\boldsymbol{\psi}^k$ vectors and taking only the first N_{KL}, we define

$$L = 2\ln\Lambda(\mathbf{n}) = \sum_{k=1}^{N_{KL}} \frac{1}{\sigma_k}\left(\boldsymbol{\psi}^k\cdot\left(\mathbf{R}_n^{-1}\cdot\mathbf{n}\right)_\parallel\right)^2, \tag{63}$$

where we dropped the suffix $\parallel$ from the basis vectors $\boldsymbol{\psi}^k$; recall that the basis is independent on the specific $N_\parallel$ segment in the data train. We know that in absence of signals the expansion coefficients $\xi_k = \frac{1}{\sqrt{\sigma_k}}\boldsymbol{\psi}^k\cdot\left(\mathbf{R}^{-1}\cdot\mathbf{n}\right)_\parallel$ are by construction uncorrelated, with zero mean and unit variance (see Eq. (27a)); they are also Gaussian variables, because they are linear combinations of Gaussian variables. Hence L is distributed as a $\chi^2\left(N_{KL}\right)$, and analogously the formula in Eq. (56) for $d_1\left(L|\mathsf{SNR}\right)$ holds, where $N_\parallel \to N_{KL}$ and

$$\mathsf{SNR} \equiv \frac{1}{\sqrt{2N_{KL}}}\sum_{k=1}^{N_{KL}}\frac{1}{\sigma_k}\left[\boldsymbol{\psi}^k\cdot\left(\mathbf{R}_n\cdot\mathbf{s}\right)_\parallel\right]^2; \tag{64}$$

again, these results are exact, despite the fact that we are using only a subset of the DKL vectors. This might be useful in order to implement χ^2 tests for non-Gaussianity, similarly to what has been done in the analysis of Caltech 40m data while searching for coalescing binaries signals [33].

The other possible approximation is the use of the Fourier transform instead of the DKL transform, as discussed in Section II F 2. In the limit in which this approximation is legitimate, namely, when $N_\parallel$ is large and the $\boldsymbol{\psi}^k$ vectors converge to the vectors $\mathbf{w}^k$ of the Fourier basis, the coefficients of the Fourier expansion behave from the statistical point of view as those of the DKL expansion, and the same results apply.

III. MULTIPLE DETECTOR CASE

A. The signal at each detector

The mathematical tools needed to compactly describe the response of interferometric detectors to a coherent GW signal have been laid out in several papers [34, 35, 36], and have been recently reviewed and applied to the problem of network detection of coalescing binary signals [37]. We collect the definitions and formulas useful to us in Section B 1 and we refer

particularly to [37], whose nomenclature we follow closely, for a complete treatment.

We describe the incoming gravitational wave by means of a *wave frame*, having the Z axis aligned with the direction of propagation; the signal is parameterized by two polarizations $h_{+,\times}$, whose definition depends on the orientation of the $\mathbf{X}$, $\mathbf{Y}$ axes; as in Section II D, we will regard $h_{+,\times}$ as independent nuisance parameters.

Another important frame is the *network frame*, defined as centered on the Earth and having the $\mathbf{Z}$ axis oriented along the North pole and the $\mathbf{X}$ axis along the Greenwich meridian: rotations between the network frame and the wave frame are accomplished by the Euler angles ϕ, θ, ψ; we can set $\psi = 0$ from now on, because it specifies a rotation of the wave frame around the $\mathbf{Z}$ axis, which is inessential. In turn we can define reference frames centered on the detectors (see Section B 1), and call $\alpha_{\mathsf{L}}, \beta_{\mathsf{L}}, \gamma_{\mathsf{L}}$ the Euler angles needed to rotate between the network frame and the *detector frame* relative to the L-th detector: see Section B 2 for the values of these angles for the interferometers under construction.

We write therefore the signal at the L-th detector as

$$s_{\mathsf{L}}(t) = h_+(t-\tau_{\mathsf{L}})\, F_{\mathsf{L}}^{+} + h_\times(t-\tau_{\mathsf{L}})\, F_{\mathsf{L}}^{\times} \tag{65}$$

where τ_{L} is the delay of the signal with respect to what would be received by a detector at the center of the Earth: it depends on the direction of the source. The two antenna patterns $F_{\mathsf{L}}^{+}, F_{\mathsf{L}}^{\times}$ are given in Section B 1 as functions of the ϕ, θ and $\alpha_{\mathsf{L}}, \beta_{\mathsf{L}}, \gamma_{\mathsf{L}}$ angles; we concentrate on burst signals and omit the dependence on time of the source location. In other words, the angles ϕ, θ are a function of time and of declination and right ascension, while the network frame rotates with the Earth.

Please note that ϕ, θ should not be confused with the elevation and azimuth angles θ_s, ϕ_s which locate the source in polar coordinates with respect to the network frame: the two sets of angles are related by

$$\phi_s = \frac{\pi}{2} + \phi,\ \theta_s = \pi - \theta; \tag{66}$$

in the following we will *always* use, unless explicity stated, the Euler angles ϕ, θ.

B. Network likelihood with uncorrelated noise

We consider first a simpler case, assuming that the Gaussian noise of the individual detectors is uncorrelated. This would be definitely true if the only noise sources were the fundamental ones, namely those dictating the baseline sensitivity (with the possible exception of the seismic noise for detectors at the same site); technical noises, like power line interferences [38], or external noises, like correlated magnetic field fluctuations [39], may spoil this assumption.

Given this simplification, the likelihood ratio is just the product of the ratios for the M individual detectors

$$\Lambda(\boldsymbol{x}|\,\mathsf{h}) \equiv \prod_{\mathsf{L}=1}^{M} \Lambda_{\mathsf{L}}(\mathbf{x}_{\mathsf{L}}|\,\mathbf{s}) \tag{67}$$

where we borrow from [20] a bold-italic notation for the direct sum $\boldsymbol{x} \equiv \mathbf{x}_1 \oplus \mathbf{x}_2 \oplus \cdots \oplus \mathbf{x}_M$ of the data vector from the individual detectors. Λ is conditioned by the presence of the signal h, described in each detector by Eq. (65) in terms of the same two polarizations (in the wave frame) $h_{+,\times}$: we have

$$\Lambda_{\mathsf{L}}(\mathbf{x}_{\mathsf{L}}|\,\mathbf{s}) \equiv e^{-\frac{1}{2}(\mathbf{s}_{\mathsf{L}})_i (\mathbf{R}_{\mathsf{LL}}^{-1})_{ij} (\mathbf{s}_{\mathsf{L}})_j + (\mathbf{s}_{\mathsf{L}})_i (\mathbf{R}_{\mathsf{LL}}^{-1}\cdot\mathbf{x}_{\mathsf{L}})_{i+d_{\mathsf{L}}}} \tag{68}$$

where we have exploited the time invariance of the correlation matrices, and introduced a shift d_{L} in the index of the data $\mathbf{x}_{\mathsf{L}}$, changing the reference time at detector L in order to compensate for the delay $\tau_{\mathsf{L}}(\theta, \phi)$; hence the burst signal appears at the same time in the individual data vectors[53]. The matrix $\mathbf{R}_{\mathsf{LL}}$ represents the noise autocorrelation for detector L and the double index is irrelevant for the time being, but will be useful when dealing with cross-detector correlated noise. We can express $\mathbf{s}_{\mathsf{L}}$ in terms of the two polarizations, treated as independent vector variables: it is convenient to write

$$s_{\mathsf{L}}[i] = \mathsf{h}^t[i] \cdot \mathsf{F}_{\mathsf{L}} = (h_+[i], h_\times[i]) \cdot \begin{pmatrix} F_{\mathsf{L}}^{+} \\ F_{\mathsf{L}}^{\times} \end{pmatrix}; \tag{69}$$

hence h should be regarded as a two column matrix, and $\mathbf{s}_{\mathsf{L}}$ is a vector resulting by contracting one of the indices with those in the vector F_{L}.

The likelihood results to be

$$\Lambda(\boldsymbol{x}|\,\mathsf{h}) = e^{-\frac{1}{2}\mathsf{h}^t\cdot\left(\sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \mathbf{R}_{\mathsf{LL}}^{-1} \otimes \mathsf{F}_{\mathsf{L}}^t\right)\cdot\mathsf{h} + \mathsf{h}^t\cdot\left[\sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \mathbf{y}_{\mathsf{L}}\right]} \tag{70}$$

where we have introduced the δ-filtered data

$$y_{\mathsf{L}}[i] = \left(\mathbf{R}_{\mathsf{LL}}^{-1} \cdot \mathbf{x}_{\mathsf{L}}\right)_{i+d_{\mathsf{L}}} \tag{71}$$

including the d_{L} time shift.

The Gaussian integration over the $h_+, h_\times$ vectors can be performed, and we obtain the network log-likelihood

$$2\ln\Lambda(\boldsymbol{x}|\theta,\phi) = \left[\sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \mathbf{y}_{\mathsf{L}}\right]_{\parallel}^{t} \cdot \left[\sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \left(\mathbf{R}_{\mathsf{LL}}^{-1}\right)_{\parallel} \otimes \mathsf{F}_{\mathsf{L}}^{t}\right]^{-1} \cdot \left[\sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \mathbf{y}_{\mathsf{L}}\right]_{\parallel}; \tag{72}$$

as expected, the fact that the signal is coherent across detectors makes it impossible to factor out the integrated likelihood in a product of terms. The expression obtained is similar to the one proposed by Anderson *et al.*[13, Eq. (5.29)] apart the fact that, as in the case of a single detector, they have chosen the signal prior flat in the metric induced by the matrix

$\left[\sum_{\mathsf{L}} \mathsf{F}_{\mathsf{K}} \otimes \left(\mathbf{R}_{\mathsf{LL}}^{-1}\right)_{\parallel} \otimes \mathsf{F}_{\mathsf{LL}}^{t}\right]^{-1}$.

The formal notation adopted in Eq. (72) deserves some clarifications:

$\mathsf{y} \equiv \begin{pmatrix} \mathsf{y}_{+} \\ \mathsf{y}_{\times} \end{pmatrix} = \sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \mathbf{y}_{\mathsf{L}}$ is a $2 \times N$ vector, where each row contains $\sum_{\mathsf{L}} F_{\mathsf{L}}^{+(\times)} \mathbf{y}_{\mathsf{L}}$, a vector combining data from all the detectors, weighted with the inverse of the correlation matrix pertinent to each of them, and summed with weights depending on the sky direction considered.

$\mathbf{\Theta} \equiv \sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \left(\mathbf{R}_{\mathsf{LL}}^{-1}\right)_{\parallel} \otimes \mathsf{F}_{\mathsf{L}}^{t}$ is a $2 \times N_{\parallel} \times N_{\parallel} \times 2$ matrix, where as in the case of the single detector $N_{\parallel}$ is the dimension of the $\mathcal{V}_{\parallel}$ subspace we are testing for the presence of a burst. It is built summing the matrices relative to each detector, then it must be inverted and contracted with the $2 \times N_{\parallel}$ matrices obtained restricting $\sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \mathbf{y}_{\mathsf{L}}$ to $\mathcal{V}_{\parallel}$ in order to construct the scalar statistic.

The matrix y can be easily computed, while for $\mathbf{\Theta}$, as in Eq. (21), the following identity holds, in absence of signal:

$$\begin{aligned} E\left[\mathsf{y} \otimes \mathsf{y}^{t}\right] &= \sum_{\mathsf{K},\mathsf{L}} \mathsf{F}_{\mathsf{K}} \otimes E\left[\mathbf{y}_{\mathsf{K}} \otimes \mathbf{y}_{\mathsf{L}}^{t}\right] \otimes \mathsf{F}_{\mathsf{L}}^{t} \\ &= \sum_{\mathsf{K},\mathsf{L}} \mathsf{F}_{\mathsf{K}} \otimes \mathbf{R}_{(\mathsf{KL})y} \otimes \mathsf{F}_{\mathsf{L}}^{t}\,; \end{aligned} \tag{73}$$

if the detector noises are statistically independent, it reduces to

$$E\left[\mathsf{y} \otimes \mathsf{y}^{t}\right] = \sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \left(\mathbf{R}_{\mathsf{LL}}^{-1}\right)_{\parallel} \otimes \mathsf{F}_{\mathsf{L}}^{t} = \mathbf{\Theta} \tag{74}$$

where we have used $\mathbf{R}_{(\mathsf{LL})y} = \mathbf{R}_{\mathsf{LL}}^{-1}$ for the correlation matrix of the $\mathbf{y}_{\mathsf{L}}$ time series.

Now we would like to factor the correlation matrices $\left(\mathbf{R}_{\mathsf{LL}}^{-1}\right)_{\parallel}$ relative to each detector: however, each of them admits a different KL expansion over the $\mathcal{V}_{\parallel}$ subspace

$$\left[\mathbf{R}_{\mathsf{LL}}^{-1}\right]_{\parallel} = \sum_{k} \sigma_{k}^{\mathsf{L}} \boldsymbol{\psi}_{\mathsf{L}}^{k} \otimes \boldsymbol{\psi}_{\mathsf{L}}^{k} \tag{75}$$

and the bases $\left\{\boldsymbol{\psi}_{\mathsf{L}}^{k}, k \in \left[1 \ldots N_{\parallel}\right]\right\}$ are generally different for each detector[54], hence the sum of tensor products in Eq. (73) does not factor in a patent way into a product of terms: however, we are going to show that it can be factored introducing two DKL bases.

1. *Exact form for the network statistic*

Let us fully exploit our understanding in Eq. (73) that the matrix

$$\mathbf{\Theta} = \sum_{\mathsf{L}} \begin{bmatrix} \left(F_{\mathsf{L}}^{+}\right)^{2} & F_{\mathsf{L}}^{+} F_{\mathsf{L}}^{\times} \\ F_{\mathsf{L}}^{+} F_{\mathsf{L}}^{\times} & \left(F_{\mathsf{L}}^{\times}\right)^{2} \end{bmatrix} \otimes \left(\mathbf{R}_{\mathsf{LL}}^{-1}\right)_{\parallel} \tag{76}$$

is the correlation matrix of the (vector) signal y. We note that the two time series y_{+} and $\mathsf{y}_{\times}$ are *jointly stationary*, that is also their cross correlation depends just on the relative lag[55].

We can introduce two DKL bases, $\boldsymbol{\psi}_{+,\times}^{k}, k \in \left[1, N_{\parallel}\right]$, for the two time series $\mathsf{y}_{+,\times}$: in terms of these two bases

$$\mathbf{\Theta} = \sum_{k,l} \begin{bmatrix} \boldsymbol{\sigma}_{kl}^{++} \boldsymbol{\psi}_{+}^{k} \otimes \left[\boldsymbol{\psi}_{+}^{l}\right]^{t} & \boldsymbol{\sigma}_{kl}^{+\times} \boldsymbol{\psi}_{+}^{k} \otimes \left[\boldsymbol{\psi}_{\times}^{l}\right]^{t} \\ \boldsymbol{\sigma}_{kl}^{\times+} \boldsymbol{\psi}_{\times}^{k} \otimes \left[\boldsymbol{\psi}_{+}^{l}\right]^{t} & \boldsymbol{\sigma}_{kl}^{\times\times} \boldsymbol{\psi}_{\times}^{k} \otimes \left[\boldsymbol{\psi}_{\times}^{l}\right]^{t} \end{bmatrix}; \tag{77}$$

where the diagonal terms $(\boldsymbol{\sigma})^{++}$ and $(\boldsymbol{\sigma})^{\times\times}$ are simple:

$$(\boldsymbol{\sigma})_{kl}^{++(\times\times)} = \delta_{kl} \sigma_{l}^{+(\times)},$$

where $\sigma_{k}^{+}, \sigma_{k}^{\times}$ are the eigenvalues of the two DKL bases, while the off diagonal terms are

$$\boldsymbol{\sigma}_{kl}^{+\times(\times+)} \equiv \sum_{\mathsf{K}} F_{\mathsf{K}}^{+} F_{\mathsf{K}}^{\times} \left[\boldsymbol{\psi}_{+(\times)}^{k}\right]^{t} \cdot \left(\mathbf{R}_{\mathsf{KK}}^{-1}\right)_{\parallel} \cdot \boldsymbol{\psi}_{\times(+)}^{l} \tag{78}$$

or simply in terms of the estimated cross correlations

$$\begin{aligned} \boldsymbol{\sigma}_{kl}^{+\times(\times+)} &\equiv \left[\boldsymbol{\psi}_{+(\times)}^{k}\right]^{t} \cdot E\left[y_{+(\times)} \otimes y_{\times(+)}\right] \cdot \boldsymbol{\psi}_{\times(+)}^{l} \\ &\equiv E\left[c_{+(\times)}^{k} c_{\times(+)}^{l}\right] \end{aligned} \tag{79}$$

where $c_{+(\times)}^{k}$ are the coefficients of the DKL expansion in the two bases. Therefore the estimation of the matrix $\boldsymbol{\sigma}$ is simple: once the $\boldsymbol{\psi}_{+(\times)}$ eigenvectors are defined, the eigenvalues give immediately the diagonal terms $\boldsymbol{\sigma}^{++(\times\times)}$, while the cross terms are most easily estimated from the data, performing the DKL decomposition and cross correlating the coefficients.

The matrix $\mathbf{\Theta}$ can be easily inverted, obtaining

$$\mathbf{\Theta}^{-1} = \sum_{k,l} \begin{bmatrix} \boldsymbol{\rho}_{kl}^{++} \boldsymbol{\psi}_{+}^{k} \otimes \left[\boldsymbol{\psi}_{+}^{l}\right]^{t} & \boldsymbol{\rho}_{kl}^{+\times} \boldsymbol{\psi}_{+}^{k} \otimes \left[\boldsymbol{\psi}_{\times}^{l}\right]^{t} \\ \boldsymbol{\rho}_{kl}^{\times+} \boldsymbol{\psi}_{\times}^{k} \otimes \left[\boldsymbol{\psi}_{+}^{l}\right]^{t} & \boldsymbol{\rho}_{kl}^{\times\times} \boldsymbol{\psi}_{\times}^{k} \otimes \left[\boldsymbol{\psi}_{\times}^{l}\right]^{t} \end{bmatrix} \tag{80}$$

where $\boldsymbol{\rho} \equiv \boldsymbol{\sigma}^{-1}$ is such that

$$\sum_{l=1}^{N_{\parallel}} \sum_{q=+,\times} \boldsymbol{\sigma}_{kl}^{pq} \boldsymbol{\rho}_{lm}^{qr} = \delta_{km} \delta^{pq}\,. \tag{81}$$

It is easy to show that the solution for $\boldsymbol{\rho}$ is (note that $\boldsymbol{\sigma}^{++(\times\times)}$ are diagonal matrices)

$$\boldsymbol{\rho}^{\times\times} = \left(\boldsymbol{\sigma}^{\times\times} - \boldsymbol{\sigma}^{\times+} \cdot \left(\boldsymbol{\sigma}^{++}\right)^{-1} \cdot \boldsymbol{\sigma}^{+\times}\right)^{-1} \tag{82a}$$

$$\boldsymbol{\rho}^{+\times} = -\left(\boldsymbol{\sigma}^{++}\right)^{-1} \cdot \boldsymbol{\sigma}^{+\times} \cdot \boldsymbol{\rho}^{\times\times} \tag{82b}$$

with similar expression exchanging $\times, +$, in terms of products of $N_{\parallel} \times N_{\parallel}$ matrices. We can finally write the statistic $L =$

$\mathsf{y}\cdot(\boldsymbol{\Theta})^{-1}\cdot\mathsf{y}$ as a quadratic form over the coefficients of the two KL expansions:

$$L = \sum_{k,l=1}^{N_\parallel}\sum_{p,q}\boldsymbol{\rho}_{kl}^{pq}\,c_p^k\,c_q^l \text{ , where} \tag{83a}$$

$$c_p^k \equiv \left[\boldsymbol{\psi}_p^k\right]^t\cdot(\mathsf{y}_p)_\parallel = \sum_{\mathsf{K}=1}^{M}F_{\mathsf{K}}^p\left[\boldsymbol{\psi}_p^k\right]^t\cdot\left(\mathbf{R}_{\mathsf{KK}}^{-1}\cdot\mathbf{x}_{\mathsf{K}}\right)_\parallel; \tag{83b}$$

where we should keep in mind that the F_p^{I} functions depend the chosen direction in the sky, and with them both the coefficients σ_{kl}^{pq} and the DKL basis vectors $\boldsymbol{\psi}_p^k$.

As argued in Section II F, the distribution(s) of the DKL eigenvalues may allow to approximate the sum neglecting terms which would result in noisy contributions: a in-depth discussion would require however to consider a realistic noise spectrum.

We will come back to the algorithm defined by Eqs. (83a,83b) after having exposed the simplifications possible in the case of large $N_\parallel$.

2. *Simplified case: large $N_\parallel$*

As in Section II F 2, if $N_\parallel$ is large enough to justify the approximation of the DKL with a Fourier transform, the matrix $\left(\mathbf{R}_{(\mathsf{KK})y}\right)_\parallel = \left(\mathbf{R}_{\mathsf{KK}}^{-1}\right)_\parallel$ can be written as follows:

$$\left(\mathbf{R}_{(\mathsf{KK})y}\right)_\parallel = \frac{f_s}{2}\sum_{k=0}^{N_\parallel-1}S_{(\mathsf{KK})y}[k]\,\mathbf{w}_k\otimes\mathbf{w}_k^H \tag{84}$$

where $S_{(\mathsf{KK})y}[k]$ is the one-sided noise spectrum of the $(\mathbf{y}_{\mathsf{K}})_\parallel$ data; its frequency resolution is $f_s/N_\parallel$, and it does not depend on the direction in the sky. Now the matrix $\boldsymbol{\Theta}$ (see Eq. (73)) can be factored out and we obtain in analogy with the results of Anderson *et al*.[13, Sec. V C]

$$\sum_{\mathsf{K}}\mathsf{F}_{\mathsf{K}}\otimes\left(\mathbf{R}_{\mathsf{KK}}^{-1}\right)_\parallel\otimes\mathsf{F}_{\mathsf{K}}^t = \frac{f_s}{2}\sum_{k=0}^{N_\parallel-1}\mathbf{w}_k\otimes\mathsf{S}_y[k]\otimes\mathbf{w}_k^H \tag{85}$$

where we have introduced a network spectral density S_{y} for the δ filtered data $\mathbf{y}$: each element of S_{y} is a 2×2 matrix, depending on the Euler angles $\phi\theta$ through the F_{K} terms

$$\mathsf{S}_{\mathsf{y}}[k]\equiv\sum_{\mathsf{K}}S_{(\mathsf{KK})y}[k]\;\mathsf{F}_{\mathsf{K}}\otimes\mathsf{F}_{\mathsf{K}}^t\,. \tag{86}$$

We can therefore rewrite the statistic in Eq. (72) as

$$L(\mathsf{x}) = \frac{2\,f_s}{N_\parallel}\sum_{k=1}^{N_\parallel-2}\left[\tilde{\mathsf{y}}[k]\right]_\parallel^H\cdot\left[\mathsf{S}_{\mathsf{y}}[k]\right]^{-1}\cdot\left[\tilde{\mathsf{y}}[k]\right]_\parallel \tag{87a}$$

$$\tilde{\mathsf{y}}[\mathsf{k}] \equiv \sum_{\mathsf{L}=1}^{M}\mathsf{F}_{\mathsf{L}}\tilde{y}_{\mathsf{L}}[k] \tag{87b}$$

in complete analogy with the single detector case, Eq. (36); in this case $\tilde{\mathsf{y}}[\mathsf{k}]$ is a 2 elements vector, whose components are just the Fourier transform of the $\mathsf{y}_{+,\times}$ time series. In analogy with the single detector case it is simple to prove that this statistic converges to the excess power statistic defined in [13, Eq. (5.29)] in the limit $N_\parallel\to N$, or equivalently when the cross correlations between the subspaces $\mathcal{V}_\parallel,\mathcal{V}_\perp$ can be neglected.

3. *Description of the algorithm*

We find useful to briefly outline the detection algorithm in the case of uncorrelated noise across the detectors: we will see that the changes induced by the correlations can be accommodated easily. We can be sketchy because most steps are similar to those described in Section II H.

Our M-detectors network produces $M\times N$ data, represented by the vectors $\mathbf{x}_{\mathsf{K}}$; we are looking for burst of length $N_\parallel\ll N$. Then we should

1. filter the data of the individual detectors for the occurrence of δ events, obtaining new vectors $\mathbf{y}_{\mathsf{K}}$; this operation can be done with the Fourier transform and costs $O(M\times N\ln N)$.
2. For each direction in the sky, shift the $\mathbf{y}_{\mathsf{K}}$ vectors to compensate for the delays τ_I, and sum the M vectors weighed with the $F_{\mathsf{K}}^{+(\times)}$ polarizations, to obtain $y_{+,\times}$; cost $N_{\rm sky}\times O(M\times N)$ where we have introduced $N_{\rm sky}$ as the number of effectively independent directions.
3. Estimate the $\boldsymbol{\psi}_{+,\times}$ DKL bases appropriate for expanding $(y_{+,\times})_\parallel$. This step may not be needed for large $N_\parallel$; if required it costs $O\left(N_{\rm sky}\times N_\parallel^3\right)$.
4. If using the DKL bases, build $\boldsymbol{\Theta}^{-1}$, that is compute the $2\times N_\parallel\times N_\parallel\times2$ matrix $\left(\boldsymbol{\sigma}^{-1}\right)_{lm}^{pq}$; this step requires $O(N_{\rm sky}\times N_\parallel^3)$ operations; otherwise, estimate the "network" spectral density $\mathsf{S}_y[k]$ of the y data, with frequency resolution $f_s/N_\parallel$, for instance averaging over the $N/N_\parallel$ possible $\mathcal{V}_\parallel$ subspaces: a step costing $O\left(\frac{N}{N_\parallel}\times N_\parallel\ln N_\parallel\right)+O\left(M\times N_\parallel\right)$
5. For each possible interval of length $N_\parallel$ in the data vector, perform the decomposition over the bases $\boldsymbol{\psi}_{+,\times}$ or alternatively over the Fourier basis: the cost is $O\left(N_{\rm sky}\times N\times N_\parallel\right)$ in the first case, $O\left(N_{\rm sky}\times N\times\ln N_\parallel\right)$ in the second case.
6. Evaluate the statistic L either using Eq. (83a) or using Eq. (87a), depending on the path followed.

In the large $N_\parallel$ approximation several further optimizations are possible: for example the Fourier transform of data $\mathbf{y}_{\mathsf{L}}$ can be done once, and the shifts needed to evaluate the sum $\sum_{\mathsf{L}}\mathsf{F}_{\mathsf{L}}\otimes\tilde{\mathbf{y}}_{\mathsf{L}}$ for different sky directions become phase factors to be attached to the F_{L} tensors. The big unknown in this estimation is the number $N_{\rm sky}$ of independent sky direction which should be probed. Given the "spin-2" dependence of the antenna patterns on the sky location, one expects a relatively

slow variation, hence it should be possible to sample the solid angle on a reduced number of ϕ, θ angles, and possibly exploit hierarchical methods to further reduce $N_{\rm sky}$: this may be the subject of future work.

C. The case of correlated noise among the detectors

The problem of correctly writing down the likelihood in the case of correlated noise among M detectors has already been studied in depth by Finn [20], who has also proposed to apply a transformation to the M data channels, which would decorrelate the noise. We face here a technical difficulty, such a transformation would also "rotate" the signal and render awkward the bookkeeping in our derivation: we want here to explore a different approach, motived by the hope that the cross-correlation terms will result to be significantly smaller than the diagonal terms, so that a perturbation expansion is possible.

If there is correlated noise, the likelihood for observing $\boldsymbol{x}$ in presence of a signal h can be written in full generality as [20, Sec. III B]

$$\Lambda\left(\boldsymbol{x}|\,\mathsf{h}\right) = \exp\left[-\frac{1}{2}\left\langle \boldsymbol{s},\boldsymbol{s}\right\rangle_{\boldsymbol{R}} + \left\langle \boldsymbol{s},\boldsymbol{x}\right\rangle_{\boldsymbol{R}}\right] \tag{88}$$

where we have already defined the symbol $\boldsymbol{x}$ for the $M\times N$ matrix representing the M time series, each of length N, produced by the detectors, while

$$\boldsymbol{s} = \mathbf{s}_1 \oplus \mathbf{s}_2 \oplus \cdots \oplus \mathbf{s}_M \tag{89}$$

is the signal at each detector, dependent on $h_+, h_\times$ that we have defined in the wave frame. We borrow from[20] the notation (with a different normalization)

$$\left\langle \boldsymbol{a},\boldsymbol{b}\right\rangle_{\boldsymbol{R}} \equiv \boldsymbol{a}\left(\boldsymbol{R}\right)^{-1}\boldsymbol{b} = \sum_{\mathsf{KL}\,kl} a_K\left[k\right]\left(\boldsymbol{R}^{-1}\right)_{\mathsf{KL}\,kl} b_L\left[l\right]; \tag{90}$$

$\boldsymbol{R}$ is a $MN\times MN$ matrix which we regard as a tensor

$$\boldsymbol{R} \equiv E\left[\boldsymbol{n}\otimes\boldsymbol{n}\right], \tag{91}$$

that is

$$\left(\boldsymbol{R}\right)_{\mathsf{KL}\,kl} = \left(\mathbf{R}_{\mathsf{KL}}\right)_{kl} = E\left[n_{\mathsf{K}}[k+d_{\mathsf{K}}]\,n_{\mathsf{L}}[l+d_{\mathsf{L}}]\right]. \tag{92}$$

Each $N\times N$ matrix $\mathbf{R}_{\mathsf{KL}}$ is a Toeplitz matrix depending only on $k-l+(d_{\mathsf{K}}-d_{\mathsf{L}})$ and not necessarily symmetric, unless $\mathsf{K}=\mathsf{L}$; only the symmetry $\left(\boldsymbol{R}\right)_{\mathsf{KL}\,kl} = \left(\boldsymbol{R}\right)_{\mathsf{LK}\,lk}$ holds, which ensures that $\left\langle \boldsymbol{a},\boldsymbol{b}\right\rangle = \left\langle \boldsymbol{b},\boldsymbol{a}\right\rangle$. Notice also that, similarly to the previous sections, we have shifted the labeling of data on each detector so that the burst is simultaneous in the time series: we must be careful, because the time shifts do not cancel out in the cross terms of the correlation matrix.

The inverse matrix $\boldsymbol{R}^{-1}$is defined, as in [20], such that

$$\delta_{\mathsf{IJ}}\delta_{ij} = \left(\boldsymbol{R}^{-1}\boldsymbol{R}\right)_{\mathsf{IJ}\,ij} = \sum_{\mathsf{K}=1}^{M}\sum_{k=1}^{N}\left(\boldsymbol{R}^{-1}\right)_{\mathsf{IK}\,ik}\left(\boldsymbol{R}\right)_{\mathsf{KJ}\,kj} \tag{93}$$

and the network likelihood can be written explicitly as

$$\Lambda\left(\boldsymbol{x}|\,\mathsf{h}\right) \equiv e^{-\frac{1}{2}s_{\mathsf{K}}[k]\left(\boldsymbol{R}^{-1}\right)_{\mathsf{KL}\,kl}s_{\mathsf{L}}[l]+s_{\mathsf{K}}[k]\left(\boldsymbol{R}^{-1}\right)_{\mathsf{KL}\,kl}x_{\mathsf{L}}[l+d_{\mathsf{L}}]}; \tag{94}$$

summation is implied over the indices k,l labeling the samples, and the indices K,L labeling the detectors.

We have already written in Eq. (69) the explicit form of $s_{\mathsf{L}}\left[l\right]$, and we can proceed as before to integrate over the nuisance parameters $h_{+,\times}$: we obtain formally a similar expression

$$2\ln\Lambda\left(\boldsymbol{x}|\,\theta,\phi\right) = \left[\sum_{\mathsf{I}}\mathsf{F}_{\mathsf{I}}\otimes\mathbf{y}_{\mathsf{I}}\right]^{t}_{\parallel}\cdot\left[\sum_{\mathsf{KL}}\mathsf{F}_{\mathsf{K}}\otimes\left(\boldsymbol{R}^{-1}\right)_{\mathsf{KL}}\otimes\mathsf{F}^{t}_{\mathsf{L}}\right]^{-1}_{\parallel}\cdot\left[\sum_{\mathsf{J}}\mathsf{F}_{\mathsf{J}}\otimes\mathbf{y}_{\mathsf{J}}\right]_{\parallel} = \mathsf{y}_{\parallel}\cdot\left[\boldsymbol{\Theta}\right]^{-1}_{\parallel}\cdot\mathsf{y}_{\parallel} \tag{95}$$

where again the suffix $\parallel$ means restriction of time indices to the $\boldsymbol{V}_{\parallel}$ subspace; $\boldsymbol{\Theta}$ is a $2\times N_{\parallel}\times N_{\parallel}\times 2$ matrix, constructed contracting the detector indices in $\boldsymbol{R}$ with the corresponding indices in the $2\times M$ matrix F, while $\mathsf{y}_{\parallel}$ is a $2\times N_{\parallel}$ matrix obtained contracting the detector index of F with the corresponding index in the matrix $\boldsymbol{y}$.

Notice that $\mathbf{y}_{\mathsf{I}}$ combines data from the different interferometers: one has by definition

$$y_{\mathsf{I}}\left[i\right] \equiv \sum_{\mathsf{J}\,j}\left(\boldsymbol{R}^{-1}\right)_{\mathsf{IJ}\,ij}x_{\mathsf{J}}\left[j+d_{\mathsf{J}}\right], \tag{96}$$

a sort of δ function filtering for multiple interferometers.

Given the formal exact solution in Eq. (95), we proceed with the assumption that the cross correlations among detectors are much smaller than the internal correlations: more precisely, that

$$\frac{||\mathbf{R}_{\mathsf{KL}}||}{\sqrt{||\mathbf{R}_{\mathsf{KK}}||\,||\mathbf{R}_{\mathsf{LL}}||}} \ll 1 \tag{97}$$

where $||\mathbf{A}|| \equiv \max_{||\mathbf{x}||=1}||\mathbf{A}\cdot\mathbf{x}||$ is the matrix norm induced by the standard vector norm $||\mathbf{x}|| \equiv \sqrt{\mathbf{x}\cdot\mathbf{x}}$. We split $\boldsymbol{R}$ into a block diagonal $\boldsymbol{D}$, and a off-diagonal $\boldsymbol{O}$:

$$\boldsymbol{R} = \boldsymbol{D}+\boldsymbol{O} \tag{98a}$$

$$\left(\boldsymbol{D}\right)_{\mathsf{KL}\,kl} = \delta_{\mathsf{KL}}\left(\mathbf{R}_{\mathsf{KK}}\right)_{kl} \tag{98b}$$

and expand $\boldsymbol{R}^{-1}$ in powers of $\boldsymbol{O}$ using the identity

$$(\boldsymbol{D}+\boldsymbol{O})^{-1} = \boldsymbol{D}^{-1} - \boldsymbol{D}^{-1}\boldsymbol{O}\,(\boldsymbol{D}+\boldsymbol{O})^{-1} \tag{99}$$

which allows to approximate

$$\boldsymbol{R}^{-1} \simeq \boldsymbol{D}^{-1} - \boldsymbol{D}^{-1}\boldsymbol{O}\left[\boldsymbol{D}^{-1} - \boldsymbol{D}^{-1}\boldsymbol{O}\left(\boldsymbol{D}^{-1} - \ldots\right)\right] \tag{100}$$

to any desired precision. This expression is useful because the inverse of the block diagonal matrix $\boldsymbol{D}$ is simple

$$\left(\boldsymbol{D}^{-1}\right)_{\mathsf{KL}\,kl} = \delta_{\mathsf{KL}}\left(\mathbf{R}^{-1}_{\mathsf{KK}}\right)_{kl}\,, \tag{101}$$

in terms of inverse correlation matrices on each detector.

We can use this expansion to write down an approximate likelihood: the δ filtered data $\mathbf{y}_{\mathsf{K}}$ are

$$\mathbf{y}_{\mathsf{K}} \simeq \mathbf{R}^{-1}_{\mathsf{KK}}\cdot\mathbf{x}_{\mathsf{K}} - \sum_{\mathsf{L}\neq\mathsf{K}}\mathbf{R}^{-1}_{\mathsf{KK}}\cdot\mathbf{R}_{\mathsf{KL}}\cdot\mathbf{R}^{-1}_{\mathsf{LL}}\cdot\mathbf{x}_{\mathsf{L}} \tag{102}$$

where we understand the shift of the data $\mathbf{x}$, to simplify the notation, and we keep only the first order in $\mathbf{R}_{\mathsf{KL}}$. This expression can be evaluated in the Fourier space

$$y_{\mathsf{K}}[l] \simeq \frac{2}{N}\sum_{k=1}^{N-2}\frac{e^{-i2\pi kl/N}}{S_{\mathsf{KK}}[k]}\left[\tilde{x}_{\mathsf{K}}[k] - \sum_{\mathsf{L}\neq\mathsf{K}}\frac{S_{\mathsf{KL}}[k]}{S_{\mathsf{LL}}[k]}\tilde{x}_{\mathsf{L}}[k]\right]\,; \tag{103}$$

and can be easily extended to higher orders, leading to

$$\boldsymbol{y} = \frac{2}{N}\sum_{k=1}^{N-2}e^{-i2\pi kl/N}\left[\boldsymbol{S}[k]\right]^{-1}\cdot\boldsymbol{x}[k] \tag{104}$$

where $\boldsymbol{x}$ is the $M\times N$ data "vector" of the network, and $\boldsymbol{S}[k]$ is the $M\times M$ matrix with elements

$$S_{\mathsf{KL}}[k] \equiv \frac{2}{f_s}\mathbf{w}_k^H\cdot\mathbf{R}_{\mathsf{KL}}\cdot\mathbf{w}_k\,. \tag{105}$$

To prevent misunderstandings we underline that this procedure is not a whitening, and it does not correspond to defining uncorrelated data channels: it is instead the multi-detector analogous of filtering for the occurrence of δ-function events.

Having obtained in this way an approximation to some desired order of the $\mathsf{y} = \begin{pmatrix}\mathsf{y}_+\\ \mathsf{y}_\times\end{pmatrix} = \sum_{\mathsf{J}}\mathsf{F}_{\mathsf{J}}\otimes\mathbf{y}_{\mathsf{J}}$ data matrix (of size $2\times N$), we are actually able to proceed with the same methods exploited in the uncorrelated noise case. In order to test for the occurrence of a burst in the subspace $\mathcal{V}_\parallel$, we need to restrict the y data to the burst subspace $\mathcal{V}_\parallel$ and define there the DKL bases $\boldsymbol{\psi}_{+\times}$, as in III B 1, or Fourier bases if $N_\parallel$ is large enough.

We are then able to *estimate* the matrix $\boldsymbol{\Theta} = E\left[\mathsf{y}\otimes\mathsf{y}\right]$ and compute the elements of the $\boldsymbol{\sigma}$ matrix (see Eq. (77)) exactly as in that case: the diagonal elements $\boldsymbol{\sigma}^{++(\times\times)}$ just as diagonal matrices built from the eigenvalues $\sigma_k^{+(\times)}$, and the off diagonal matrices $\boldsymbol{\sigma}^{+(\times)}$ using Eq. (79); all the remaining derivation goes unchanged.

1. *Simplified case: large $N_\parallel$*

As in Section III B 2, a considerable simplification is possible if the DKL $\boldsymbol{\psi}_{+(\times)}$ bases of the $\mathcal{V}_\parallel$ space converge to the Fourier bases: in this case we have, by the very definition $\boldsymbol{\Theta} = E\left[\mathsf{y}\otimes\mathsf{y}\right]$, and in analogy with Eq. (85)

$$\sum_{\mathsf{IJ}}\mathsf{F}_{\mathsf{I}}\otimes\left(\boldsymbol{R}^{-1}\right)_{\mathsf{IJ}}\otimes\mathsf{F}^t_{\mathsf{J}} \equiv \frac{1}{2}\sum_k\mathbf{w}_k\otimes\mathsf{S}_{\mathsf{y}}[k]\otimes\mathbf{w}_k^H \tag{106a}$$

$$\mathsf{S}_{\mathsf{y}}[k] \equiv \begin{pmatrix} S_{++}[k] & S_{+\times}[k]\\ S_{\times+}[k] & S_{\times\times}[k]\end{pmatrix} \tag{106b}$$

where in turn $S_{pq}[k]$ are the 4 possible cross-spectra, at frequency resolution $f_s/N_\parallel$, defined from the data $y_+, y_\times$. Hence the log-likelihood has exactly the same expression as in the uncorrelated noise case

$$L(\mathsf{x}) = \frac{2f_s}{N_\parallel}\sum_{k=1}^{N_\parallel-2}\left[\tilde{\mathsf{y}}[k]\right]_\parallel^H\cdot\left[\mathsf{S}_{\mathsf{y}}[k]\right]^{-1}\cdot\left[\tilde{\mathsf{y}}[k]\right]_\parallel \tag{107a}$$

$$\tilde{\mathsf{y}}[\mathsf{k}] \equiv \sum_{\mathsf{L}=1}^{M}\mathsf{F}_{\mathsf{L}}\tilde{y}_{\mathsf{L}}[k] \tag{107b}$$

with the difference that the y_{K} have been obtained combining data from the different detectors, in a manner dependent on the direction in the sky; approximately as in Eq. (103), or exactly as in Eq. (104).

The algorithm described in Section III B 3 goes almost unchanged: the only real change is in the computation of the data $\mathbf{y}_{\mathsf{K}}$. In particular, we have to rearrange steps (1) and (2) because now the δ-filtering must be preceded by the time shift of the data $\mathbf{x}_{\mathsf{L}}$: the two operations no longer commute. All the other steps remain unaltered.

D. Distribution of the network statistic

Having given in Section II I a demonstration that the statistic L is a χ^2 distributed variable, we can generalize those results: we know that the network statistic

$$L = \mathsf{y}_\parallel\cdot\boldsymbol{\Theta}^{-1}\cdot\mathsf{y}_\parallel \tag{108}$$

is written in terms of the inverse correlation matrix of the variables $\mathsf{y}_\parallel$ themselves, viz.

$$\boldsymbol{\Theta} = E\left[\mathsf{y}_\parallel\otimes\mathsf{y}_\parallel\right] \tag{109}$$

where the average is taken in absence of signal. Hence again L is a χ^2 variable, as can be directly checked considering its first two moments

$$\begin{aligned} E\left[L|H_0\right] &= E\left[\mathsf{y}_\parallel\boldsymbol{\Theta}^{-1}\mathsf{y}_\parallel\right] = \mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{\Theta}^{-1}\right)\\ &= 2N_\parallel \end{aligned} \tag{110a}$$

$$\begin{aligned} E\left[L^2|H_0\right] &= 2\,\mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{\Theta}^{-1}\boldsymbol{\Theta}\boldsymbol{\Theta}^{-1}\right) + \left[\mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{\Theta}^{-1}\right)\right]^2\\ &= 2N_\parallel\left(2N_\parallel+2\right) \end{aligned} \tag{110b}$$

and generalizing with a bit of combinatorics to

$$E\left[L^m|H_0\right]=\prod_{k=1}^{m}\left[2N_{\|}+2\left(k-1\right)\right] \tag{111}$$

which is the moment expansion of a $\chi^2\left(2N_{\|}\right)$ variable. In the same way, in presence of a signal the distribution $d_1\left(L\right)$ will be given by Eq. (56) with the substitution $N_{\|}\to 2N_{\|}$ and the SNR defined as

$$\mathsf{SNR}=\frac{E\left[L|H_1\right]-E\left[L|H_0\right]}{\sqrt{2\left(2N_{\|}\right)}}; \tag{112}$$

the extra factor of $\sqrt{2}$ in the denominator results from the definition of the signal in terms of two polarizations.

E. Example: network sensitivity to δ events

The "network spectral density" introduced in Eqs. (86,106b) is a 2×2 matrix of spectra and (complex) cross-spectra, which depends on the sky direction: it is therefore interesting to derive some scalar quantity which can be plotted in a spherical projection and give a visual idea of the sensitivity of the network.

As a simple example, motivated by the short duration of the impulsive features in some of the model waveforms [14, Fig. 2 (Model A)], we may consider the response of the network to a burst of duration $dt_{\rm burst}=1/f_s$, where f_s is the sampling rate in the detectors, and having amplitudes $A_+,A_\times$ in the two polarizations. We assume that the noise is uncorrelated across the detectors, and that the data have been shifted to capture the event in every data streams at the same time index a: we have

$$\tilde{x}_{\mathsf{L}}\left[k\right]=\left(A_+F_{\mathsf{L}}^+ + A_\times F_L^\times\right)\frac{1}{f_s}e^{-j2\pi ka/N} \tag{113}$$

and the δ-filtered signal is (see Eq. (4))

$$y_{\mathsf{L}}\left[l\right]=\left(A_+F_{\mathsf{L}}^+ + A_\times F_{\mathsf{L}}^\times\right)\frac{2}{f_sN}\sum_{k=1}^{N-2}\frac{e^{i2\pi k(l-a)/N}}{S_{\mathsf{LL}}\left[k\right]}; \tag{114}$$

projecting on the $\mathcal{V}_{\|}$ subspace means setting $l=a$, hence

$$y_{\mathsf{L}}\left[a\right]=\left(A_+F_{\mathsf{L}}^+ + A_\times F_{\mathsf{L}}^\times\right)\mathrm{rms}\left(y_{\mathsf{L}}\right) \tag{115}$$

where we have defined

$$\begin{aligned}\mathrm{rms}\left(y_{\mathsf{L}}\right) &\equiv \frac{1}{f_sN}\sum_{k=1}^{N/2-1}\frac{1}{S_{\mathsf{LL}}\left[k\right]}\\ &\simeq \frac{1}{f_s^2}\int_{f_{\mathsf{seism}}}^{f_{\mathsf{Nyquist}}}\frac{df}{S_{\mathsf{LL}}\left(f\right)};\end{aligned} \tag{116}$$

this is the same quantity resulting from the analysis in [40], where burst signals with uniform spectrum in the detection band had been considered. Next, we have

$$\mathsf{y}_{\|}=\sum_{\mathsf{L}}\mathrm{rms}\left(y_{\mathsf{L}}\right)\left(A_+F_{\mathsf{L}}^+ + A_\times F_{\mathsf{L}}^\times\right)\begin{pmatrix}F_{\mathsf{L}}^+\\ F_{\mathsf{L}}^\times\end{pmatrix} \tag{117}$$

and, noticing that $\left(\mathbf{R}_{\mathsf{LL}}^{-1}\right)\left[a,a\right]=\mathrm{rms}\left(\mathsf{y}_{\mathsf{L}}\right)$

$$\mathbf{\Theta}=\sum_{\mathsf{L}}\mathrm{rms}\left(\mathsf{y}_{\mathsf{L}}\right)\begin{pmatrix}\left(F_{\mathsf{L}}^+\right)^2 & F_{\mathsf{L}}^+F_{\mathsf{L}}^\times\\ F_{\mathsf{L}}^+F_{\mathsf{L}}^\times & \left(F_{\mathsf{L}}^\times\right)^2\end{pmatrix}; \tag{118}$$

the log-likelihood statistic $L=\mathsf{y}_{\|}\cdot\mathbf{\Theta}^{-1}\cdot\mathsf{y}_{\|}$ can now be easily evaluated: we average over $A_+,A_\times$ keeping their geometric mean $A\equiv\sqrt{A_+^2+A_\times^2}$ fixed, and evaluate the resulting SNR as in Eq. (112) for networks of interferometric detectors built out of different partitions of the instruments currently under commissioning. Because of the f_s^{-2} factor in $\mathrm{rms}\left(\mathsf{y}_{\mathsf{L}}\right)$, the scale is set by the effective amplitude $A\,dt=A\,f_s^{-1}$.

We report in Fig. (1) two polar plots of the SNR, obtained setting $A\,dt=10^{-23}s$; with $f_s^{-1}=O\left(1ms\right)$ this would correspond to a strain $A=O\left(10^{-20}\right)$, possible for a core collapse event at a distance of 10 kpc [9, 14]. We have considered either the network of three LIGOs interferometers, or a network including also GEO600, TAMA and Virgo: the details on the detectors are reported in Section B 2, where the nominal noise spectrum is modeled in Eq. (B16) and in Tab. II, while locations and orientations are reported in Tab. I.

The plots have been obtained using a *Mathematica*$^{\rm tm}$ notebook which is available from the author upon request.

The global interferometric network appears significantly more sensitive, and much of the effect is due to the contribution of Virgo: however the result should be considered merely illustrative, because the chosen shape of the burst (a δ-function) corresponds to a flat spectrum in the frequency domain; this choice favors the Virgo detector substantially, as already noticed in [40], because of the wide bandwidth of the model sensitivity of Virgo.

IV. CONCLUSIONS AND OUTLOOK

In this paper we have defined a statistic for the detection of burst signals, which is well suited to be applied to data affected by colored noise, thus properly generalizing the *excess power* statistic [11, 12, 13] to the case in which the spectral noise density varies significantly over the frequency band of interest, and the signal prior is assumed to be flat in $\mathbb{R}^N$. It is optimal in the Bayes sense, under the two hypotheses that the signal is distributed uniformly in amplitude, and is contaminated by additive Gaussian noise. The extension to the network case was straightforward and it was also possible to take into account in a natural way the possible presence of Gaussian noise correlations among the detectors, either perturbatively or exactly.

The lack of assumptions on the GW signal distribution is both an advantage and a disadvantage. We believe that the proposed statistic is the correct one for a detection strategy free of *a priori* assumptions, apart the duration of the burst; yet we are aware that it does not lend itself to easily include assumptions on the amplitude distribution of the signal, as it was possible in [13], thus making difficult to set Bayesian thresholds.

Our generalization is not much more expensive, from the computational point of view, than the statistic discussed by

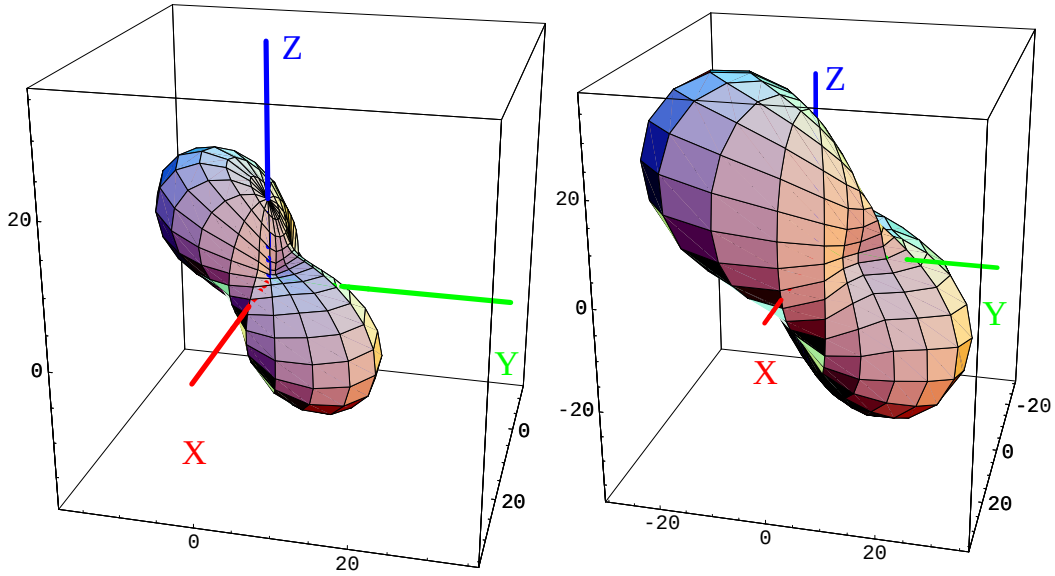

Figure 1: Polar plot of the SNR for bursts having $A\,dt = 10^{-23}s$, as a function of the direction of the source. The figure at left refers to the network of LIGO detectors, while the figure at right includes also GEO600, TAMA and Virgo. The axes of the network frame are shown: we recall that **Z** points toward the geographical north, and **X** crosses the Greenwich meridian.

Anderson *et al.* [12, 13]; in its simplest implementation it amounts to perform a matched filtering for δ-functions followed by the calculation of a "energy" over the time window to be tested for the occurrence of a burst. The whole analysis chain can be implemented combining a few standard time-series analysis tools and therefore implemented with efficient algorithms. The limited computational cost of the single detector search method suggests that it may be applied to the full data set, before any triggering is performed: in this sense it is a method proposed for the *on-line* search.

An evaluation of the actual detection performance, when considering theoretical waveforms and simulated noise, remains to be done; this will be the subject of future work, much along the lines of [7, 8, 10].

The cost of network detection was not fully estimated: it depends on the number of directions in the sky effectively independent, which was not studied in this paper.

Besides the detection of gravitational events, we propose this statistic as a tool for excess noise characterization: real interferometric detectors are definitely affected by non Gaussian noise [33], and as long as the excess noise is dominated by burst signals, like Poisson distributed creep events, we can think of using this statistic as a tool for detecting and characterizing them. Once a event results in a instance of the L statistic above the threshold, and therefore a burst of excess noise or a gravitational wave is detected, our algorithm provides a way to encode this information in a manner optimal with respect to the distribution of the noise [18]. In fact, the discrete Karhunen-Loève transform that we have chosen as a tool to compute the "energy" over the burst time window is equivalent to a principal component analysis: the coefficients of the DKLT are ordered by the amount of RMS noise contributed by the corresponding basis vectors. A candidate event can therefore be encoded selecting just the largest DKLT coefficients, while retaining most of its relevant "energy", that is, the energy that is distributed in less noisy components.

Moreover, we recall that in absence of signals the DKLT coefficients are statistically uncorrelated: one may instead anticipate peculiar, spurious correlations when a signal of any nature is present. These correlations will emerge as regularities if the events occur repeatedly in time, and it should be possible to catalog them in an automatic way, for instance by using clustering methods in the vector space of the DKLT coefficients.

Acknowledgments

The author gratefully thanks Alessandra Buonanno and Albert Lazzarini for enlightening discussions, and Éanna É. Flanagan and Jolien D.E. Creighton for reading an earlier version of the manuscript and providing me with valuable comments and suggestions. Warm thanks are also due to the LSC reviewer, who spotted and corrected several misprints.

The LIGO Project and LIGO Laboratory are supported by the National Science Foundation under cooperative agreement PHY-9210038.

This document has been assigned LIGO Document Control Center number LIGO-P010019-01-E.

Appendix A: DFT CONVENTIONS

We list here our conventions for the discrete time, discrete frequency representation of stochastic processes. Let f_s be the sampling frequency, and N the length of the data sample for

the discrete time process $x[l]$; then the Fourier transform pair $\mathbf{x} \leftrightarrow \tilde{\mathbf{x}}$ is

$$\tilde{x}[k] \equiv \frac{1}{f_s}\sum_{l=0}^{N-1} e^{-i2\pi kl/N} x[l], \tag{A1a}$$

$$x[l] = \frac{f_s}{N}\sum_{k=0}^{N-1} e^{i2\pi kl/N}\tilde{x}[k]; \tag{A1b}$$

the one-sided sample spectrum $S_x[k]$ is *defined* by

$$\frac{1}{2}S_x[k] = \frac{f_s}{N}E\left[|\tilde{x}_{\mathcal{H}}[k]|^2\right]; \tag{A2}$$

where $\tilde{\mathbf{x}}_{\mathcal{H}} \leftrightarrow (\mathcal{H} * \mathbf{x})$ is the Fourier transform of a suitably windowed realization of $\mathbf{x}$. For large N one has that the correlation function $R_x\left(\frac{a-b}{f_s}\right) = (\mathbf{R}_x)[a,b]$ and the sample spectrum are *approximately* Fourier pairs

$$R_x(\tau) \simeq \frac{f_s}{N}\sum_{k=0}^{N-1}\frac{1}{2}S_x[k]\, e^{i2\pi k\tau f_s/N}; \tag{A3}$$

or equivalently

$$\mathbf{R}_x \simeq f_s\sum_{k=0}^{N-1}\frac{1}{2}S_x[k]\,\mathbf{w}_k \otimes \mathbf{w}_k^H, \tag{A4}$$

where $\mathbf{w}_k$ are the Fourier orthonormal basis vectors

$$\mathbf{w}_k = \frac{1}{\sqrt{N}}\left[1, \omega^k, \omega^{2k}, \ldots \omega^{(N-1)k}\right] \tag{A5}$$

with $\omega \equiv e^{i2\pi/N}$. For a zero-mean process one has

$$\left(\mathbf{R}_x^{-1}\right)[a,b] \simeq \frac{1}{f_s N}\sum_{k=1}^{N-2}\frac{2}{S_x[k]}e^{i2\pi(a-b)k/N} \tag{A6}$$

and the useful relations

$$(\mathbf{R}_x \cdot \mathbf{y})[l] \simeq \frac{f_s^2}{N}\sum_{k=0}^{N-1}\frac{1}{2}S_x[k]\,\tilde{y}[k]\, e^{i2\pi kl/N} \tag{A7a}$$

$$\left(\mathbf{R}_x^{-1} \cdot \mathbf{y}\right)[l] \simeq \frac{1}{N}\sum_{k=1}^{N-2}\frac{2}{S_x[k]}\tilde{y}[k]\, e^{i2\pi kl/N} \tag{A7b}$$

where in the second one also y is a a zero mean process. These are approximate relations because of the finite N.

Appendix B: NETWORK MODEL

In this appendix we summarize the mathematics needed to describe a network of detectors, the essential geometric characteristic of the interferometers under construction, and their anticipated model spectral densities.

1. Geometry

We adopt the following reference frames[37]:

network frame centered on Earth and chosen with the **Z** axis aligned along the geographical north, the **X** axis crossing the Greenwich meridian;

detector frames centered on the beam splitter of each detector, the **Z** axis pointing toward the local zenith and the **X** axis bisecting the detector arms;

wave frame having the **Z** axis aligned along the direction of propagation of the wave, and **X** axis lying in the **(X, Y)** plane of the network frame.

Rotations of coordinates from one frame to the other are expressed in terms of Euler angles, that is

$$\mathsf{x}_{\mathsf{wave}} = O(\phi,\theta,\psi)\cdot \mathsf{x}_{\mathsf{network}} \tag{B1a}$$

$$\mathsf{x}_{\mathsf{detector_L}} = O(\alpha_{\mathsf{L}},\beta_{\mathsf{L}},\gamma_{\mathsf{L}})\cdot \mathsf{x}_{\mathsf{network}}; \tag{B1b}$$

if θ_s, ϕ_s are elevation and azimuth of the source in the network frame, the relation with the Euler angles ϕ, θ is

$$\phi = \phi_s - \frac{\pi}{2};\ \theta = \pi - \theta_s\,. \tag{B2}$$

The ψ angle is zero, according to the definition of the *wave frame*. One can introduce the wave tensor

$$\mathbf{w}(t) = \frac{1}{2}\left[(h_+(t) + i\,h_\times(t))\,\mathbf{e}_R + (h_+(t) - i\,h_\times(t))\,\mathbf{e}_L\right] \tag{B3}$$

where the helicity states $\mathbf{e}_R, \mathbf{e}_L$ can be written as

$$\mathbf{e}_{L,R} = \frac{1}{2}(\mathbf{e}_X \pm i\,\mathbf{e}_Y)\otimes(\mathbf{e}_X \pm i\,\mathbf{e}_Y) \tag{B4}$$

in terms of unit vectors $\mathbf{e}_X, \mathbf{e}_Y$ specifying the **X**,**Y** axes of the wave frame as. In the network frame they can also be written as Symmetric Trace Free tensors of second rank (STF-2) $\mathcal{Y}_{mn}$[35, 36]:

$$\begin{aligned}\mathbf{e}_{L,R} &= \sqrt{\frac{8\pi}{15}}\,(\mathcal{Y}_{2\pm2})_{\mathsf{wave}} \\ &= \sqrt{\frac{8\pi}{15}}T_{\pm2n}(\phi,\theta,0)\,(\mathcal{Y}_{2n})_{\mathsf{network}}\end{aligned} \tag{B5}$$

where T_{mn}, with $m,n = 0,\pm1,\pm2$ are Gel'fand's functions of rank 2 and depend on the Euler angles $\phi,\theta,\psi(=0)$ needed to rotate the network frame to the wave frame.

For completeness, the explicit form of the T^l_{mn} functions for arbitrary rank is[34] as follows:

$$T^l_{mn}(\phi,\theta,\psi) = e^{-i(n\phi+m\psi)}P^l_{mn}(\cos\theta), \tag{B6}$$

$$P^l_{mn}(\mu) = \frac{(-1)^{l-m}\,i^{n-m}}{2^l\,(l-m)!}\sqrt{\frac{(l-m)!\,(l+n)!}{(l+m)!\,(l-n)!}}\frac{(1-\mu)^{\frac{m-n}{2}}}{(1+\mu)^{\frac{m+n}{2}}}\frac{d^{l-n}}{d\mu^{l-n}}\left[\frac{(1+\mu)^{m+l}}{(1-\mu)^{m-l}}\right], \tag{B7}$$

where $m,n\in[-l,l]$.

The detector response is encoded by the tensor $\mathbf{d}_{\mathsf{L}}$

$$\mathbf{d}_{\mathsf{L}} = \sin(2\Omega_{\mathsf{L}})\left(\mathbf{n}_{(\mathsf{L})1}\otimes\mathbf{n}_{(\mathsf{L})1}-\mathbf{n}_{(\mathsf{L})2}\otimes\mathbf{n}_{(\mathsf{L})2}\right) \tag{B8}$$

in terms of the unit vectors aligned along the L-th interferometer arms, and the aperture angle $2\Omega_{\mathsf{L}}$ of the arms. The factor $\sin(2\Omega_{\mathsf{L}})$ is 1 for all the detectors apart GEO600, where it is 0.997; consequently we will understand the factor in the following, to simplify the notation.

The tensor $\mathbf{d}$ has a simple expression

$$\mathrm{tr}\left[\mathbf{d}\cdot(\mathcal{Y}_{2m})_{\mathsf{detector}}\right] = -i\sqrt{\frac{15}{8\pi}}\left(\delta_{m2}-\delta_{m-2}\right) \tag{B9}$$

in terms of STF-2 tensors in the detector frame. With two successive rotations we obtain the coefficients of its expansion in terms of STF-2 tensors in the wave frame

$$\mathrm{tr}\left[\mathbf{d}\cdot(\mathcal{Y}_{2m})_{\mathsf{network}}\right] = -i\sqrt{\frac{15}{8\pi}}\Big[T^*_{2m}(\alpha_{\mathsf{L}},\beta_{\mathsf{L}},\gamma_{\mathsf{L}}) - T^*_{-2m}(\alpha_{\mathsf{L}},\beta_{\mathsf{L}},\gamma_{\mathsf{L}})\Big] \tag{B10a}$$

$$\mathrm{tr}\left[\mathbf{d}\cdot(\mathcal{Y}_{2m})_{\mathsf{wave}}\right] = \sum_n T_{mn}(\phi,\theta,0)\,\mathrm{tr}\left[\mathbf{d}\cdot(\mathcal{Y}_{2n})_{\mathsf{network}}\right] \equiv \sqrt{\frac{15}{8\pi}}D_m(\phi,\theta,0,\alpha_{\mathsf{L}},\beta_{\mathsf{L}},\gamma_{\mathsf{L}}) \tag{B10b}$$

where we have introduced in the last formula a short-hand notation.

Finally the signal at the L-th detector, after proper h-reconstruction in order to deconvolve the interferometer response function, will be the scalar

$$s_{\mathsf{L}}(t) = \mathrm{tr}\left[\mathbf{w}(t-\tau_{\mathsf{L}}(\phi\theta))\cdot\mathbf{d}_{\mathsf{L}}\right] \tag{B11}$$

where τ_{L} is the delay at the L-th detector with respect to the network frame; it can be positive or negative depending on the direction of the source. In terms of the (complex) beam pattern functions $F^{L,R}_{\mathsf{L}}\equiv\mathrm{tr}(\mathbf{e}_{L,R}\cdot\mathbf{d}_{\mathsf{L}})$ for the two left and right wave polarizations

$$F^L_{\mathsf{L}} = D_{-2}(\phi,\theta,0,\alpha_{\mathsf{L}},\beta_{\mathsf{L}},\gamma_{\mathsf{L}}) \tag{B12a}$$

$$F^R_{\mathsf{L}} = \left(F^L_{\mathsf{L}}\right)^* = D_2(\phi,\theta,0,\alpha_{\mathsf{L}},\beta_{\mathsf{L}},\gamma_{\mathsf{L}}) \tag{B12b}$$

one can alternatively write

$$s_{\mathsf{L}}(t) = \Re\left[\left(h_+(t-\tau_{\mathsf{L}})+ih_\times(t-\tau_{\mathsf{L}})\right)F^R_{\mathsf{L}}\right] = \frac{1}{2}\left[(h_++ih_\times)D_2(\ldots)+(h_+-ih_\times)D_{-2}(\ldots)\right] \tag{B13}$$

or equivalently and more conveniently for our work

$$s_{\mathsf{L}}(t) = h_+F^+_{\mathsf{L}}+h_\times F^\times_{\mathsf{L}} \tag{B14}$$

where, reintroducing the aperture angle

$$F^+_{\mathsf{L}} \equiv \sin(2\Omega_{\mathsf{L}})\,\Re\left[D_{-2}(\phi,\theta,0,\alpha_{\mathsf{L}},\beta_{\mathsf{L}},\gamma_{\mathsf{L}})\right] \tag{B15a}$$

$$F^\times_{\mathsf{L}} \equiv \sin(2\Omega_{\mathsf{L}})\,\Im\left[D_{-2}(\phi,\theta,0,\alpha_{\mathsf{L}},\beta_{\mathsf{L}},\gamma_{\mathsf{L}})\right]. \tag{B15b}$$

The given expression for the signal and for the antenna patterns, as stressed in[37], is convenient because it keeps in factor form the rotations among the various frames.

2. Interferometer network characteristics

The geometrical characteristics of the detectors considered in this study are listed in Table I where we quote the latitude north of the equator, the longitude east of Greenwich, the azimuths of the **X** and **Y** arms of the detectors, measured counter-clockwise from the local east, and the α,β,γ Euler angles needed to rotate coordinates in the network frame to coordinates in the detector frame: all the angles are measured in radians. The orientation data are taken from[41] and updated with informations from the web sites of the different collaborations[42]; the naming of the axes has been changed in one case so that all the detectors have azimuth(**X** arm) < azimuth(**Y** arm). The resulting Euler angles should be taken with care because are computed in the approximation of spherical Earth and neglecting the elevation of the detector sites and the fact that the arms are cords and not tangents of the surface. For a more accurate model, please see [44].

We show in Fig. (2) the locations of the detectors and the reference frames attached to them, from two viewpoints above Europe and the United States of America.

The other important characteristic of the detector is their planned sensitivity. We have chosen to include only the baseline thermal and shot-noise sources, omitting resonances in the observation band: the model for the noise spectrum (filtered to deconvolve the response function to gravitational waves) is therefore

$$S_n(f) = \frac{S_{\mathrm{pend}}}{f^5}+\frac{S_{\mathrm{mirror}}}{f}+S_{\mathrm{shot}}\left[1+\left(\frac{f}{f_{\mathrm{knee}}}\right)^2\right] \tag{B16}$$

where S_{pend} quantifies the thermal noise of the mirror pendular mode, above the pendulum resonance; S_{mirror} quantifies the $1/f$ tail of the internal modes of the mirror, excited by thermal noise; S_{shot} and f_{knee} parameterize the optical read-out noise.

Detector	Lat.	Long.	X azim.	Y azim.	α, β, γ Euler angles
GEO600	0.911935	0.171217	0.377166	2.02353	-1.20035, -0.658862, -1.74201
LIGO Liv.	0.533373	-1.58424	3.45575	5.02655	2.04204, -1.03742, 0.013439
LIGO Han.	0.810705	-2.0841	2.21308	3.78387	-2.99848, -0.760091, 0.513301
TAMA	0.622733	2.43543	3.14159	4.71239	2.35619, -0.948063, 2.27696
Virgo	0.761487	0.18326	1.24791	2.81871	-2.03331, -0.809309, -1.75406

Table I: detector locations and orientations, and Euler angles (approximated) needed to express coordinates in the network frame in terms of coordinates in the detector frames.

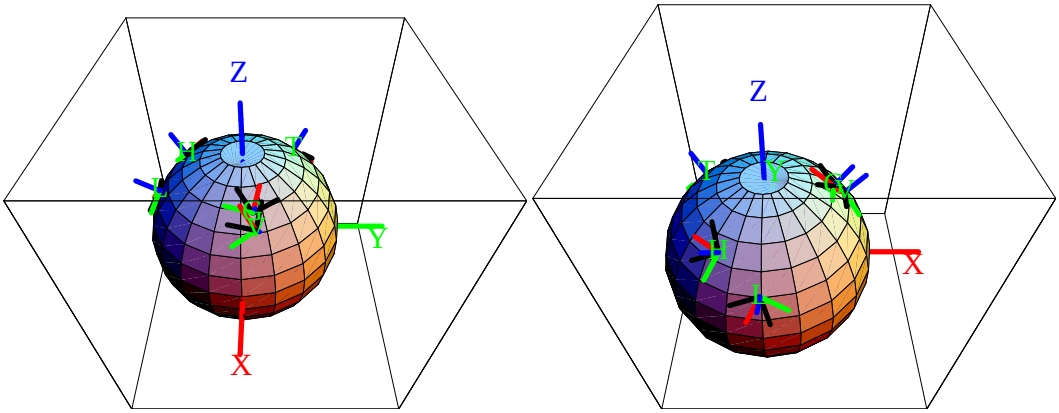

Figure 2: The locations of the detectors on Earth, labeled by their initials (H and L for LIGO Hanford and LIGO Livingston); the axes of the network frame, labeled **X**,**Y**,**Z** and of the various detector frames are also shown. To the left: view from above Europe. To the right: view from above the United States of America.

In addition to these parameters, we call $f_{\rm seism}$ the cutoff below which the seismic noise is supposed to dominate over the thermal noise. This simplified model does not include at least two important effects, the thermal violin mode resonances and the internal mirror resonance peaks, and should be considered merely illustrative.

We report in Table II the numerical values of these parameters, deduced from [43], and in Fig. (3) the comparison of the different noise spectral densities.

For the model at hand, the inverse correlation function

$$R_n^{-1}(\tau) \propto \Re \int_0^\infty e^{i2\pi f\tau} \frac{1}{S_n(f)}\, df \qquad \text{(B17)}$$

is well approximated by an expression of the form

$$\begin{aligned} R_n^{-1}(\tau) \;\simeq\; & A_0 e^{-|\tau|/\tau_0} \cos(2\pi f_0 \tau + \phi_0) \\ & + A_1 e^{-|\tau|/\tau_1} \cos(2\pi f_1 \tau + \phi_1)\,; \end{aligned} \qquad \text{(B18)}$$

we list in Table III the values of the decay times $\tau_{0,1}$ and of the "ringing" frequencies $f_{0,1}$ deduced from the values in Table II: the small values of the $\tau_{0,1}$ simply reflect the absence of resonances in the model noise curves.

[1] A.Abramovici *et al.* Science **256**, 325 (1992).
[2] B.Caron *et al.* Nucl. Phys. B (Proc.Suppl.) **54**, 167 (1997).
[3] K.Danzmann *et al.*, in *First Edoardo Amaldi Conference on Gravitational wave experiments*, ed. E.Coccia *et al.* (World Scientific, Singapore, 1995).
[4] K.Tsubono, in *First Edoardo Amaldi Conference on Gravitational Wave Experiments*, ed. E.Coccia *et al.* (World Scientific, Singapore, 1995).
[5] B.Owen and B.S.Sathyaprakash, Phys. Rev. D **60**, 022002 (1999).
[6] T.Pradier, N.Arnaud, M.A.Bizouard, F.Cavalier, M.Davier and P.Hello, Phys. Rev. D **63,** 042002 (2001).
[7] T.Pradier, N.Arnaud, M.A.Bizouard, F.Cavalier, M.Davier and P.Hello, Int. J. Mod. Phys. D **9**, 309-314 (2000).
[8] N.Arnaud, M.Barsuglia, M.-A.Bizouard, P.Canitrot, F.Cavalier, M.Davier, P.Hello and T.Pradier, *Detection in coincidence of gravitational wave bursts with a network of interferometric detectors (I): Geometric acceptance and timing*, gr-qc/0107081.
[9] T.Zwerger and E.Müller, Astron. Astrophys. **267**, 623 (1993).
[10] N. Arnaud, M. Davier, F. Cavalier and P.Hello, Phys. Rev. D **59**, 082002 (1999).
[11] É.É.Flanagan, S.A.Hughes, Phys. Rev. D **57**, 4566 (1998).
[12] W.G.Anderson, P.R.Brady, J.D.E.Creighton, É.É.Flanagan, Int. J. Mod. Phys. D **9**, 303 (2000).
[13] W.G.Anderson, P.R.Brady, J.D.E.Creighton, É.É.Flanagan, Phys. Rev. D **63**, 042003 (2001).

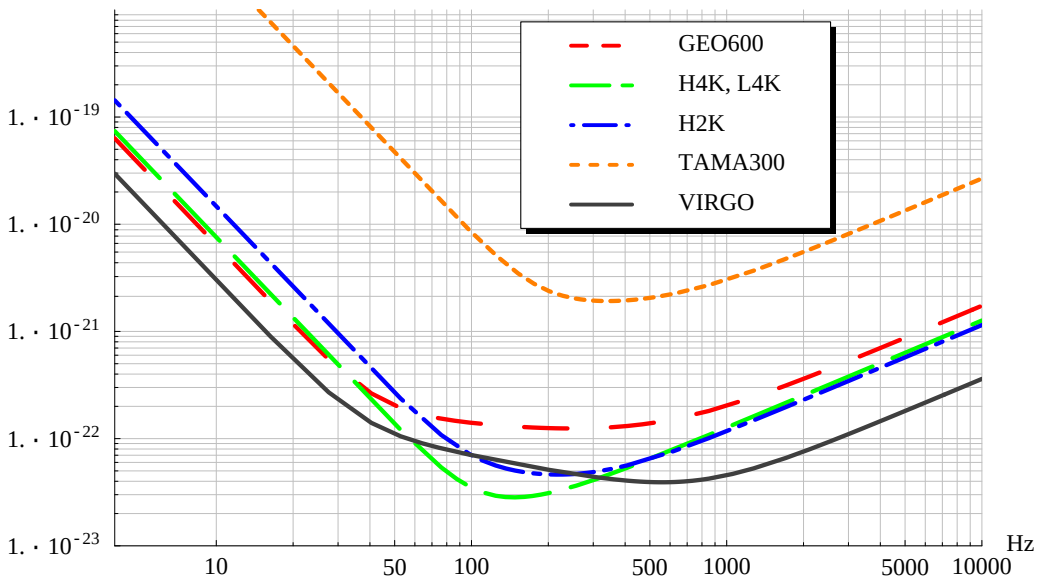

Figure 3: A comparison of the baseline spectral sensitivities, in units of $\mathrm{Hz}^{-1/2}$, for the different interferometers currently under commissioning, according to the simplified model in Eq. (B16). The two LIGO 4km interferometers "H4K" and "L4K", located in Hanford and Livingston respectively, are supposed to have the same sensitivity. For clarity's sake we do not display the so called "seismic wall" at f_{seism}, which should be understood as dictated by Table II.

Detector	f_{seism}	S_{pend}	S_{mirror}	S_{shot}	f_{knee}
GEO600	50	4.1e-36	9e-43	1e-44	577
LIGO 2K	40	2.1e-35	2.25e-43	4.35e-46	182
LIGO 4K	40	5.6e-36	3.9e-44	1.1e-46	83
TAMA	50	6.6e-31	3.2e-40	1.78e-42	500
Virgo	4	9e-37	4.5e-43	3.24e-46	500

Table II: Parameters characterizing the baseline noise of the detectors in the network; cfr. Eq. (B16).

Detector	τ_0	f_0	τ_1	f_1
GEO600	5.6e-3	32	2.7e-4	44
LIGO 2K	2.4e-3	70	6.1e-4	107
LIGO 4K	2.5e-3	83	1.1e-3	51
TAMA	1.4e-3	141	3.1e-4	82
Virgo	6.0e-3	27	2.2e-4	293

Table III: The characteristic ringing frequencies (measured in Hz) and correlation times (in sec.), for each interferometer, deduced from the simplified noise models we have adopted.

[14] H.Dimmelmeier, J.A.Font and E.Müller, *Gravitational waves from relativistic core collapse*, astro-ph/0103088 (2001).

[15] G. Pizzella, *Fisica sperimentale del campo gravitazionale*, (La Nuova Italia, Firenze, 1993).

[16] A.Papoulis, *Probability, random variables, and stochastic processes* (McGraw-Hill, New York, 1991).

[17] C.W.Therrien, *Discrete random signals and statistical signal processing* (Prentice-Hall, Englewood Cliffs, 1992).

[18] A.Ortolan, G.Vedovato, M.Cerdonio and S.Vitale, Phys. Rev. D **50**, 4737 (1994).

[19] G.Cella, V.Chickarmane and A.Viceré, Int. J. Mod. Phys. D **9**, 269 (2000).

[20] L.S.Finn, Phys. Rev. D **63**, 102001 (2001).

[21] R.N.Donough and A.D.Whalen, *Detection of Signals in Noise* (Academic Press, San Diego, 1995).

[22] Abramovici *et al.*, Phys. Lett. A **218**, 157-163 (1996).

[23] L.S.Finn and S.Mukherjee, Phys. Rev. D **63**, 062004 (2001).

[24] B.F.Schutz, *Data processing, analysis and storage for interferometric antennas*, in *The detection of gravitational waves*, ed. D.G.Blair, (Cambridge Univ. Press, Cambridge, 1991).

[25] E.Cuoco, G.Calamai, L.Fabbroni, G.Losurdo, M.Mazzoni, R.Stanga and F.Vetrano, Class. Quant. Grav. **18**, 1727-1752 (2001).

[26] E.Cuoco, G.Losurdo, G.Calamai, L.Fabbroni, M.Mazzoni, R.Stanga, G.Guidi and F.Vetrano, *Noise parametric identification and whitening for LIGO 40-meter interferometer data*, gr-qc/0104071 (2001).

[27] L.S.Finn, Phys. Rev. D **46**, 5236 (1992).

[28] D.B.Percival and A.T.Walden, *Spectral analysis for physical applications*, (Cambridge Univ. Press, Cambridge, 1993).

[29] A.M.Sintes, B.F.Schutz, Phys. Rev. D **58**, 122003 (1998).

[30] D.Buskulic, L.Derome, R.Flaminio, R.Hermel, J.C.Lacotte, F.Marion, A.Masserot, L.Massonnet, B.Mours, R.Morand, P.Puppo, D.Verkindt and M.Yvert, Int. J. of Mod. Phys D **9**, 263-267 (2000).

[31] C.W.Helstrom, *Statistical Theory of Signal Detection* (Pergamon Press, New York, 1960).

[32] I.S.Gradshteyn and I.M.Ryzhik, *Table of Integrals, Series, and*

Products (Academic Press, New York, 1965).

[33] B.Allen *et al.*, Phys. Rev. Lett. **83**, 1498-1501 (1999).

[34] S.Dhurandhar and M.Tinto, Mon. Not. R. astr. Soc. **234**, 663-676 (1988).

[35] M.Gel'fand, R.A.Minlos and Z.Ye.Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (Pergamon Press, New York, 1963).

[36] K.S.Thorne, Rev. Mod. Phys. **52**, 299-339 (1980).

[37] A.Pai, S.Dhurandhar and S.Bose, Phys. Rev. D **64**, 042004 (2001).

[38] A.Lazzarini and A.Viceré, *Analysis of the effects of long-term correlations that may compromise long-integration cross-correlation measurements*, LIGO technical note LIGO-T010101-00-E (2001).

[39] N.Christensen, Ph.D. thesis, Massachussetts Institute of Technology, (1990).

[40] B.Bhawal and S.V.Dhurandhar, *Coincidence detection of broadband signals by networks of the planned interferometric gravitational wave detectors*, (Ganeshkhind, IUCAA), IUCAA-29-95, gr-qc/9509042 (1995).

[41] B.Allen, *Gravitational wave detector sites*, gr-qc/9607075 (1996).

[42] Sources of information about detector locations and orientations:
www.geo600.uni-hannover.de/geo600/project/location.html
www.ligo.caltech.edu/docs/T/T980044-08.pdf
tamago.mtk.nao.ac.jp/tama/facility/faciliti.htm
wwwlapp.in2p3.fr/virgo/virgo.html

[43] Reference sites for the noise curves cited in the text:
www.geo600.uni-hannover.de/geo600/project/wideband.html
www.ligo.caltech.edu/~kent/ASIS_NM/noise_models.html
tamago.mtk.nao.ac.jp/tama/recom/recom3/index.html
www.virgo.infn.it/senscurve

[44] W.G.Anderson, P.R.Brady, D.Chin, J.D.E.Creighton, K.Riles and J.T.Whelan, *Beam Pattern Response Functions and Times of Arrival for Earthbound Interferometers*, http://phys.utb.edu/UTBRG/activities/papers

[45] We are assuming that the noise processes are causal: real noise is expected to be such, unless it is processed with a non-causal filter. For instance a matched filter is in general non-causal, because its impulse response is non-zero for negative times[21, sec. 6.5].

[46] It is well known[17, chap. 2] that a correlation matrix $\mathbf{R}$, being hermitian and symmetric, can be expressed as $\mathbf{R} = \mathbf{L} \cdot \mathbf{D}_L \cdot \mathbf{L}^t$, where $\mathbf{L}$ is a unit lower matrix and $\mathbf{D}_L$ is a diagonal matrix. This implies that $\mathbf{R}^{-1} = (\mathbf{L}^t)^{-1} \cdot \mathbf{D}_L^{-1} \cdot \mathbf{L}^{-1}$, and note that also $\mathbf{L}^{-1}$ is unit lower triangular. For our purposes we can take the square root of the diagonal matrix $\mathbf{D}_L$ so as to define $\mathbf{W} \equiv \sqrt{\mathbf{D}_L^{-1}} \cdot \mathbf{L}^{-1}$ and obtain $\mathbf{R}^{-1} = \mathbf{W}^t \cdot \mathbf{W}$.

[47] It is also worth noting that while $\mathbf{R}_n$ is a Toeplitz matrix, $\mathbf{R}_n^{-1}$ is not necessarily so, though it converges to a Toeplitz matrix in the $N \to \infty$ limit.

[48] This approximation will be better far from the boundaries of the data train, a fact which suggests, in the analysis of a continuous stream of data, to work on (long) overlapping segments; another consequence is that the DKL basis can be estimated once and repeatedly used for subsequent data segments.

[49] The statistic L would of course be optimal also for any prior which is itself a function of the statistic L.

[50] A regular random process $x[l]$ satisfies the Paley-Wiener condition[17, sec. 5.5.2] $\int_{-\pi}^{\pi} |\ln S_x[\omega]| \; d\omega < \infty$ and can be written as white noise filtered through a causal process: this is usually the case for physical noises.

[51] A process is said predictable if the output at time $t + dt$ can be written in terms of past values *without* error: such a process clearly cannot be written as white noise filtered through a causal filter[17, sec. 7.6.1].

[52] It can be expressed in terms of Γ and Bessel functions I as follows: ${}_0F_1(;x,y) = y^{\frac{1-x}{2}} \Gamma(x) \; \mathrm{I}_{x-1}\left(2\sqrt{y}\right)$

[53] We are also assuming that all the detectors share data resampled at the same frequency.

[54] With the possible exception of detectors which are designed equal, such as the two 4km LIGOs.

[55] Note also that the cross correlation is symmetric, because it is the sum of symmetric matrices $\mathbf{R}_{\mathrm{K}}^{-1}$.