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: 16th July 2002)

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 $N_{\|}$ 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.

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 discuss exact and approximate forms of the statistic; the choice depends on the 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 magnitude of the 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 method that we propose consists mainly 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 presence 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 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

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

turns out that possible correlations among the detector noises can be taken into account by modifying the δ-filtering step, as a direct consequence of the likelihood maximization.

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 derive the exact expression for the likelihood ratio, for a burst affecting $N_\parallel$ samples in the data stream. In Section II D 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 E we exploit the DKLT expansion and we show that it is a convenient way to implement step **B**. When the supposed burst length $N_\parallel$ is large, an approximate formula using the DFT can be used, as shown in Section II E 1.

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

The distribution of our statistic in absence and presence of a signal is considered in Section II G: 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 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: 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 the paper we adopt a discrete-time, discrete-frequency notation: the conventions for the DFT 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 is zero mean and Gaussian, 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 symmetric Toeplitz [17]. 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. We shall see later that when considering a shorter analysis window (of size $N_\parallel$) contained in a longer data train we should actually exploit also the information contained outside the $N_\parallel$ 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 to estimate the relevance of this boundary effect is the noise spectrum: one knows (see Section A) that

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

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

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

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 should be kept into account when computing false alarm and false dismissal probabilities [24]. In this 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 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 should not forget that the measure is changed by the transformation.

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

B. Detection framework

We suppose that if present the burst affects a interval $T_\parallel = N_\parallel / f_s$ starting at absolute time t_{burst}, that is a number $N_\parallel$ of samples. We are unable to prescribe priors for the signal amplitudes, so we treat then as *nuisance* parameters, to be integrated out. Anderson *et al.* [11, 13] made similar assumptions, with different hypotheses on the integration measure.

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

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

where $P(1), P(0)$ 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 $\mathbf{x}$ given that *some* signal $\mathbf{s}$ is present; $P(\mathbf{x}|0)$ is the probability defined in Eq. (2) of observing the same dataset $\mathbf{x}$ in absence of signals. In turn

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

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). In this paper we assume no a priori information on $P(\mathbf{s})$.

Given our complete ignorance of $P(1), P(0)$, we resort to the integrated likelihood ratio

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

using the Bayes rule $P(1|\mathbf{x})P(\mathbf{x}) = P(\mathbf{x}|1)P(1)$ we can write

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

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 ignorance on the waveform we assume that $P(\mathbf{s})$ is uniform 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] assumed instead that the signals $\mathbf{s}$ are uniformly distributed with respect to the metric induced by the scalar product $\langle x, y\rangle \equiv \mathbf{x}\cdot\mathbf{R}_n^{-1}\cdot\mathbf{y}$.

C. The burst statistic

A vector space notation is useful: let $\mathcal{V}_N$ be the space of all possible data vectors, having length N; we want to test for the 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

$$\Lambda(\mathbf{x}) = \int e^{-\frac{1}{2}s_i(\mathbf{R}_n^{-1})_{ij}s_j + s_i(\mathbf{R}_n^{-1}\cdot\mathbf{x})_i} \prod_i^{N_\parallel} ds_i \tag{10}$$

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

$$(\mathbf{R}_n^{-1})_\parallel \equiv (\mathbf{R}_n^{-1})_{[l:l+N_{burst}, l:l+N_{burst}]} \tag{11}$$

obtained projecting $\mathbf{R}_n^{-1}$ onto $\mathcal{V}_\parallel$. Performing the integrals over the amplitudes s_i

$$\Lambda(\mathbf{x}) \propto \exp\left[\frac{1}{2}x_\alpha(\mathbf{R}_n^{-1})_{\alpha i}\left(\left((\mathbf{R}_n^{-1})_\parallel\right)^{-1}\right)_{ij}(\mathbf{R}_n^{-1})_{j\beta}x_\beta\right] \tag{12}$$

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 split the indexes because the operations of projecting over the "burst" subspace $\mathcal{V}_\parallel$ and of inverting a matrix do not commute[43].

The log-likelyhood statistic is therefore[44]

$$L(\mathbf{x}) \equiv \sum_{i,j\in\mathcal{V}_\parallel} (\mathbf{R}_n^{-1}\mathbf{x})_i \left(\left(\left(\mathbf{R}_n^{-1}\right)_\parallel \right)^{-1} \right)_{ij} (\mathbf{R}_n^{-1}\mathbf{x})_j \,. \tag{13}$$

It may seem awkward to compute: notice however that

$$E\left[\left(\mathbf{R}_n^{-1}\mathbf{n}\right)_i \left(\mathbf{R}_n^{-1}\mathbf{n}\right)_j \right] = \left(\mathbf{R}_n^{-1}\right)_{ij} \,, \tag{14}$$

in other words the time series $\mathbf{R}_n^{-1}\cdot\mathbf{n}$ restricted to the $\mathcal{V}_\parallel$ subspace has autocorrelation matrix $\left(\mathbf{R}_n^{-1}\right)_\parallel$, which is what we need. The last step is to invert the matrix and apply it to the δ-filtered data $\mathbf{R}_n^{-1}\cdot\mathbf{x}$: a efficient tool to accomplish this task is the Karhunen-Loève expansion[17].

D. Karhunen-Loève expansion

An expansion of $\mathbf{R}$ in terms of its eigenvalues and eigenvectors is possible[16, 17, 21]:

$$R_{\alpha\beta} = \sum_{k=1}^{K} \sigma_k \psi_\alpha^k \psi_\beta^k \tag{15}$$

where K is the dimension of the matrix and

$$R_{\alpha\beta}\psi_\beta^k = \sigma_k \psi_\beta^k \tag{16}$$

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

$$\sum_\alpha \psi_\alpha^k \psi_\alpha^l = \delta^{kl} \tag{17}$$

and define a basis in the space $\mathbb{R}^K$:

$$\mathbf{x} = \sum_{k=0}^{K-1} c_k \boldsymbol{\psi}^k, \forall \mathbf{x}\in\mathbb{R}^K, \text{with}\, c_k \equiv \mathbf{x}\cdot\boldsymbol{\psi}^k \,; \tag{18}$$

this decomposition is called discrete Karhunen-Loève transform (DKLT). The Parseval's theorem

$$\mathbf{x}\cdot\mathbf{x} = \sum_{k=1}^{K} c_k^2 \tag{19}$$

holds and it is immediate to show that

$$E[c_k c_l] = \sigma_k \delta_{kl} \tag{20a}$$

$$E[\mathbf{x}\cdot\mathbf{x}] = \sum_{\alpha=1}^{K} R_{\alpha\alpha} = \sum_{k=1}^{K} \sigma_k \,. \tag{20b}$$

The similarity with the Fourier transform goes further[17, sec. 4.7.2]: for large K the $\boldsymbol{\psi}$ converge to sines and cosines, and the eigenvalues converge to the bins of the spectral density.

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.

E. Exact expression for the burst statistic

In Section II C we have shown that the exact statistic, defined in Eq. (13), can be expressed as

$$L = \mathbf{y}_\parallel \cdot \left[(\mathbf{R}_y)_\parallel \right]^{-1} \cdot \mathbf{y}_\parallel \tag{21}$$

with $\mathbf{y}_\parallel \equiv \left(\mathbf{R}_n^{-1}\cdot\mathbf{x}\right)_\parallel$ the vector of δ-filtered data

$$y[l] = \frac{1}{N}\sum_{k=0}^{N-1} \frac{2}{S_n[k]} e^{-i2\pi k l/N} \tilde{x}[k] \,; \tag{22}$$

S_n is the one-sided spectrum corresponding to the correlation matrix $\mathbf{R}_n$: the expression is valid for large N and stationary Gaussian noise. Thanks to the KL transform we have

$$\left[(\mathbf{R}_y)_\parallel \right]^{-1} = \sum_{k=1}^{N_\parallel} \frac{1}{\sigma_k} \boldsymbol{\psi}_\parallel^k \otimes \boldsymbol{\psi}_\parallel^k \,, \tag{23}$$

hence we finally obtain

$$L = \sum_{k=1}^{N_\parallel} \frac{1}{\sigma_k} \left(\boldsymbol{\psi}_\parallel^k \cdot \mathbf{y}_\parallel \right)^2 . \tag{24}$$

The reader might wonder why the DKLT is at all necessary: our statistic in Eq. (21) 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. (24)? 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 $\left\{\boldsymbol{\psi}_\parallel^k\cdot\mathbf{y}_\parallel, k\in\left[1, N_\parallel\right]\right\}$ are statistically uncorrelated renders this procedure sound and simplifies the statistical analysis.

Note that the δ-filtering in Eq. (22) copes with long correlations in the data, assuming no *deterministic* line to be present: in fact, the Wold theorem[17, sec. 7.6.2] states that a random process can be decomposed in the sum

$$x(t) = x_r(t) + x_p(t) \tag{25}$$

of a *regular* process[45] $x_r(t)$ and a *predictable* process[46] $x_p(t)$. The latter could correspond to a harmonic of the power line: it would contribute to the spectrum a term

$$\sigma_p^2 \delta(\nu - \nu_p) \tag{26}$$

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

$$\frac{N}{f_s}\delta_{k,k_p} = T\,\delta_{k,k_p} \,, \tag{27}$$

a trend in T which is the symptom of a infinite correlation length. Such spectral features cannot be properly handled by spectral factorization methods, but can be subtracted from the data [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].

1. *Approximate expression: $N_\parallel$ large*

In general we search for a burst of length $N_\parallel$ in a data N sufficiently long to resolve the narrow spectral features which give rise to long correlation times. The sampling rate $O(20\,\mathrm{kHz})$ needed to exploit the spectral range available in interferometer data may lead to consider $N_\parallel$ of the order of several hundreds, even for signals of a few tens of ms.

For large $N_\parallel$ an approximate expression based on the Fourier transform exists:

$$\left[(\mathbf{R}_y)_\parallel\right]^{-1} \simeq f_s^{-1} \sum_{k=1}^{N_\parallel-2} \frac{2}{S_y[k]} \mathbf{w}_k \otimes \mathbf{w}_k^H \tag{28}$$

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 approximate statistic

$$L(\mathbf{x}) \approx \frac{f_s}{N_\parallel} \sum_{k=1}^{N_\parallel/2-1} \frac{1}{S_y[k]} \left|\tilde{y}_\parallel[k]\right|^2 \tag{29}$$

as 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], applied to data filtered for the occurrence of δ.

F. In which sense the statistic is "optimal"

It is 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

$$\Lambda(\mathbf{x}) = \int e^{-\frac{1}{2}\mathbf{s}\cdot\mathbf{R}_n^{-1}\cdot\mathbf{s}+\mathbf{s}\cdot\mathbf{R}_n^{-1}\cdot\mathbf{x}} P(\mathbf{s})\, d\mathbf{s}. \tag{30}$$

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

$$d\mathbf{s} = \rho^{N_\parallel-1} d\rho\, d\Omega_{N_\parallel}(\hat{\mathbf{s}}) \tag{31}$$

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 on the distribution $p(\rho)$, we would like to follow the same approach as in [13, Sec. III] and write

$$\Lambda(\mathbf{x}) = \int p(\rho)\, \Lambda(\mathbf{x}|\rho)\, d\rho \tag{32}$$

where

$$\Lambda(\mathbf{x}|\rho) \equiv \int \delta\left(\rho - \sqrt{\mathbf{s}\cdot\mathbf{s}}\right) e^{-\frac{1}{2}\mathbf{s}\cdot\left(\mathbf{R}_n^{-1}\right)_\parallel\cdot\mathbf{s}+\mathbf{s}\cdot\mathbf{y}_\parallel} d\mathbf{s} \tag{33}$$

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

$$\begin{aligned}\Lambda(\mathbf{x}|\rho) &= \int \delta\left(\rho - \sqrt{\boldsymbol{\varsigma}\cdot\boldsymbol{\varsigma}}\right) e^{-\frac{1}{2}\sum_k \sigma_k \varsigma_k^2 + \boldsymbol{\varsigma}\cdot\mathbf{c}} d\boldsymbol{\varsigma} \\ &= \int e^{-\frac{\rho^2}{2}\sum_k \sigma_k^2 \hat{\varsigma}_k^2 + \rho\sum_k \hat{\varsigma}_k c_k} d\Omega_{N_\parallel}(\hat{\varsigma})\,;\end{aligned} \tag{34}$$

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 our case the expression in Eq. (34) is not 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 signal space.

This discussion shows that the statistic L we have proposed is strictly speaking *optimal* only for a signal prior $P(\mathbf{s})$ constant[47]: 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

$$L = \sum_{k,l} \mathsf{L}_{kl} \frac{c_k c_l}{\sqrt{\sigma_k \sigma_l}} \tag{35}$$

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.

G. Statistical analysis

Given the statistic L we need the distributions $d_0(L)$ and $d_1(L|\mathbf{s})$ under the 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].

Thanks to the fact that the coefficients of the DKL expansion are Gaussian, uncorrelated random variables, the distribution $d_0(L)$ is a $\chi^2\left(N_\parallel\right)$[16, Sec. 4-3]:

$$d_0(L) = \frac{L^{N_\parallel/2-1} e^{-L/2}}{2^{N_\parallel/2}\Gamma(N_\parallel/2)}. \tag{36}$$

When a signal is present, the distribution is a non-central $\chi^2\left(N_\parallel\right)$:

$$d_1(L) = \frac{L^{N_\parallel/2-1} e^{-\frac{1}{2}\left(L+\mathsf{SNR}\sqrt{2N_\parallel}\right)} {}_0\mathrm{F}_1\left(;\frac{N_\parallel}{2};\frac{\mathsf{SNR}L\sqrt{2N_\parallel}}{4}\right)}{2^{N_\parallel/2}\Gamma(N_\parallel/2)} \tag{37}$$

where ${}_0F_1$ is a hyper-geometric function[32, sec. 9.14]. The SNR explicit form is

$$\mathsf{SNR} = \frac{\left(\mathbf{R}_n^{-1}\cdot\mathbf{s}\right)_{\parallel}\cdot\left(\left(\mathbf{R}_y\right)_{\parallel}\right)^{-1}\cdot\left(\mathbf{R}_n^{-1}\cdot\mathbf{s}\right)_{\parallel}}{\sqrt{2N_{\parallel}}} \tag{38}$$

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

$$\mathsf{SNR} \equiv \frac{\left|E\left[L|H_1\right]-E\left[L|H_0\right]\right|}{\sqrt{E\left[\left(L-E\left[L|H_0\right]\right)^2|H_0\right]}}, \tag{39}$$

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. The SNR in Eq. (39) is not the *intrinsic* signal-to-noise ratio which would result from matched filtering, $\mathsf{SNR}_{\mathsf{intrinsic}} \propto \sqrt{\int\left|\tilde{s}\left(f\right)\right|^2/S_n\left(f\right)df}$; in particular, it is quadratic in the signal amplitude.

Given the distributions $d_0\left(L\right)$ and $d_1\left(L\right)$ it is immediate to compute false alarm and detection probabilities[48]. How should however d_0 and d_1 be modified, had we chosen instead to compute an approximate statistic?

1. Approximate statistic distribution

We discussed in Section II E the possibility of defining the statistic L using just a subset $N_{KL} < N_{\parallel}$ of the DKL basis vectors:

$$L_{\text{approx}} = \sum_{k=1}^{N_{KL}}\frac{1}{\sigma_k}\left(\boldsymbol{\psi}^k\cdot\left(\mathbf{R}_n^{-1}\cdot\mathbf{n}\right)_{\parallel}\right)^2; \tag{40}$$

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. (20a)); they are also Gaussian variables, because they are linear combinations of Gaussian variables. Hence L_{approx} is distributed as a $\chi^2\left(N_{KL}\right)$, and analogously the formula in Eq. (37) for $d_1\left(L|\mathsf{SNR}\right)$ holds, with $N_{\parallel}\to N_{KL}$ and

$$\mathsf{SNR} \equiv \frac{1}{\sqrt{2N_{KL}}}\sum_{k=1}^{N_{KL}}\frac{1}{\sigma_k}\left[\boldsymbol{\psi}^k\cdot\left(\mathbf{R}_n\cdot\mathbf{s}\right)_{\parallel}\right]^2; \tag{41}$$

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].

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 useful definitions and formulas in Section B 1, and we refer to [37], whose notation we follow closely, for a complete treatment.

The signal at the L-th detector is

$$s_{\mathsf{L}}\left(t\right) = h_+\left(t-\tau_{\mathsf{L}}\right)F_{\mathsf{L}}^+ + h_\times\left(t-\tau_{\mathsf{L}}\right)F_{\mathsf{L}}^\times \tag{42}$$

where τ_{L} is the delay of the signal with respect to a detector at the center of the Earth: it depends on the source direction. The antenna patterns $F_{\mathsf{L}}^+, F_{\mathsf{L}}^\times$ are given in Section B 1 as functions of the source direction and the position and orientation of the detector L; as we deal with burst signals we omit their dependence on time.

B. Network likelihood with uncorrelated noise

We consider first a simpler case, assuming that the Gaussian noise of the individual detectors is uncorrelated: then the likelihood ratio is just the product of the ratios for the M individual detectors

$$\Lambda\left(\boldsymbol{x}|\,\mathsf{h}\right) \equiv \prod_{\mathsf{L}=1}^{M}\Lambda_{\mathsf{L}}\left(\mathbf{x}_{\mathsf{L}}|\,\mathbf{s}\right) \tag{43}$$

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. (42) in terms of the same two polarizations $h_{+,\times}$: we have

$$\Lambda_{\mathsf{L}}\left(\mathbf{x}_{\mathsf{L}}|\,\mathbf{s}\right) \equiv e^{-\frac{1}{2}(\mathbf{s}_{\mathsf{L}})_i(\mathbf{R}_{\mathsf{LL}}^{-1})_{ij}(\mathbf{s}_{\mathsf{L}})_j+(\mathbf{s}_{\mathsf{L}})_i(\mathbf{R}_{\mathsf{LL}}^{-1}\cdot\mathbf{x}_{\mathsf{L}})_{i+d_{\mathsf{L}}}}. \tag{44}$$

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 τ_{L}. The matrix $\mathbf{R}_{\mathsf{LL}}$ represents the noise autocorrelation for detector L and the double index 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

$$s_{\mathsf{L}}\left[i\right] = \mathsf{h}^t\left[i\right]\cdot\mathsf{F}_{\mathsf{L}} = \left(h_+\left[i\right],h_\times\left[i\right]\right)\cdot\begin{pmatrix}F_{\mathsf{L}}^+\\F_{\mathsf{L}}^\times\end{pmatrix}; \tag{45}$$

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

The likelihood can be rewritten as

$$\Lambda\left(\boldsymbol{x}|\,\mathsf{h}\right) = e^{-\frac{1}{2}\mathsf{h}^t\cdot\left(\sum_{\mathsf{L}}\mathsf{F}_{\mathsf{L}}\otimes\mathbf{R}_{\mathsf{LL}}^{-1}\otimes\mathsf{F}_{\mathsf{L}}^t\right)\cdot\mathsf{h}+\mathsf{h}^t\cdot\left[\sum_{\mathsf{L}}\mathsf{F}_{\mathsf{L}}\otimes\mathbf{y}_{\mathsf{L}}\right]} \tag{46}$$

where we have introduced the δ-filtered data

$$y_{\mathsf{L}}\left[i\right] = \left(\mathbf{R}_{\mathsf{LL}}^{-1}\cdot\mathbf{x}_{\mathsf{L}}\right)_{i+d_{\mathsf{L}}} \tag{47}$$

which include the d_{L} time shift.

The Gaussian integration over $h_+, h_\times$ can be performed, resulting in the log-likelihood

$$2\ln\Lambda(\boldsymbol{x}) = \left[\sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \mathbf{y}_{\mathsf{L}}\right]^t_{\parallel} \cdot \left[\sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \left(\mathbf{R}^{-1}_{\mathsf{LL}}\right)_{\parallel} \otimes \mathsf{F}^t_{\mathsf{L}}\right]^{-1} \cdot \left[\sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \mathbf{y}_{\mathsf{L}}\right]_{\parallel} ; \tag{48}$$

the presence of a signal coherent across detectors makes it impossible to factor out the integrated likelihood in a product of terms[49].

The notation in Eq. (48) deserves some clarification:

$\mathsf{y} \equiv \begin{pmatrix} \mathsf{y}_+ \\ \mathsf{y}_\times \end{pmatrix} = \sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \mathbf{y}_{\mathsf{L}}$ is a $2 \times N$ matrix; each row contains $\sum_{\mathsf{L}} F_{\mathsf{L}}^{+(\times)} \mathbf{y}_{\mathsf{L}}$, in turn a vector which combines data from all the detectors, filtered by the appropriate δ-filter and scaled with weights depending on the sky direction.

$\boldsymbol{\Theta} \equiv \sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \left(\mathbf{R}^{-1}_{\mathsf{LL}}\right)_{\parallel} \otimes \mathsf{F}^t_{\mathsf{L}}$ is a $2 \times N_\parallel \times N_\parallel \times 2$ matrix; as in the single detector case $N_\parallel$ is the dimension of the $\mathcal{V}_\parallel$ subspace we are testing for the presence of a burst.

The matrix y can be easily computed: for $\boldsymbol{\Theta}$ the following identity holds, in absence of signal:

$$\begin{aligned} E\left[\mathsf{y} \otimes \mathsf{y}^t\right] &= \sum_{\mathsf{K,L}} \mathsf{F}_{\mathsf{K}} \otimes E\left[\mathbf{y}_{\mathsf{K}} \otimes \mathbf{y}^t_{\mathsf{L}}\right] \otimes \mathsf{F}^t_{\mathsf{L}} \\ &= \sum_{\mathsf{K,L}} \mathsf{F}_{\mathsf{K}} \otimes \mathbf{R}_{(\mathsf{KL})y} \otimes \mathsf{F}^t_{\mathsf{L}} ; \end{aligned} \tag{49}$$

if the detector noises are independent, it reduces to

$$E\left[\mathsf{y} \otimes \mathsf{y}^t\right] = \sum_{\mathsf{L}} \mathsf{F}_{\mathsf{L}} \otimes \left(\mathbf{R}^{-1}_{\mathsf{LL}}\right)_{\parallel} \otimes \mathsf{F}^t_{\mathsf{L}} = \boldsymbol{\Theta} \tag{50}$$

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

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

$$\left[\mathbf{R}^{-1}_{\mathsf{LL}}\right]_{\parallel} = \sum_k \sigma^{\mathsf{L}}_k \boldsymbol{\psi}^k_{\mathsf{L}} \otimes \boldsymbol{\psi}^k_{\mathsf{L}} \tag{51}$$

and the bases $\left\{\boldsymbol{\psi}^k_{\mathsf{L}}, k \in \left[1 \ldots N_\parallel\right]\right\}$ are generally different for each detector, hence the sum of tensor products in Eq. (49) does not factor into a product of terms.

1. *Exact form for the network statistic*

Recall (see Eq. (49)) that the matrix

$$\boldsymbol{\Theta} = \sum_{\mathsf{L}} \begin{bmatrix} \left(F^+_{\mathsf{L}}\right)^2 & F^+_{\mathsf{L}} F^\times_{\mathsf{L}} \\ F^+_{\mathsf{L}} F^\times_{\mathsf{L}} & \left(F^\times_{\mathsf{L}}\right)^2 \end{bmatrix} \otimes \left(\mathbf{R}^{-1}_{\mathsf{LL}}\right)_{\parallel} \tag{52}$$

is the correlation matrix of the (vector) signal y. The two time series y_+ and $\mathsf{y}_\times$ are *jointly stationary*, that is also their cross correlation depends just on the relative lag, and we can introduce two DKL bases, $\boldsymbol{\psi}^k_{+,\times}, k \in \left[1, N_\parallel\right]$. In terms of them

$$\boldsymbol{\Theta} = \sum_{k,l} \begin{bmatrix} \boldsymbol{\sigma}^{++}_{kl} \boldsymbol{\psi}^k_+ \otimes \left[\boldsymbol{\psi}^l_+\right]^t & \boldsymbol{\sigma}^{+\times}_{kl} \boldsymbol{\psi}^k_+ \otimes \left[\boldsymbol{\psi}^l_\times\right]^t \\ \boldsymbol{\sigma}^{\times+}_{kl} \boldsymbol{\psi}^k_\times \otimes \left[\boldsymbol{\psi}^l_+\right]^t & \boldsymbol{\sigma}^{\times\times}_{kl} \boldsymbol{\psi}^k_\times \otimes \left[\boldsymbol{\psi}^l_\times\right]^t \end{bmatrix} ; \tag{53}$$

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

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

with $\sigma^+_k, \sigma^\times_k$ are the eigenvalues of the two DKL bases. The off diagonal terms are

$$\boldsymbol{\sigma}^{+\times(\times+)}_{kl} \equiv \sum_{\mathsf{K}} F^+_{\mathsf{K}} F^\times_{\mathsf{K}} \left[\boldsymbol{\psi}^k_{+(\times)}\right]^t \cdot \left(\mathbf{R}^{-1}_{\mathsf{KK}}\right)_{\parallel} \cdot \boldsymbol{\psi}^l_{\times(+)} \tag{54}$$

or simply in terms of the estimated cross correlations

$$\begin{aligned} \boldsymbol{\sigma}^{+\times(\times+)}_{kl} &\equiv \left[\boldsymbol{\psi}^k_{+(\times)}\right]^t \cdot E\left[y_{+(\times)} \otimes y_{\times(+)}\right] \cdot \boldsymbol{\psi}^l_{\times(+)} \\ &\equiv E\left[c^k_{+(\times)} c^l_{\times(+)}\right] \end{aligned} \tag{55}$$

with $c^k_{+(\times)}$ the coefficients of the DKLT in the two bases[50]. The matrix $\boldsymbol{\Theta}$ can be inverted

$$\boldsymbol{\Theta}^{-1} = \sum_{k,l} \begin{bmatrix} \boldsymbol{\rho}^{++}_{kl} \boldsymbol{\psi}^k_+ \otimes \left[\boldsymbol{\psi}^l_+\right]^t & \boldsymbol{\rho}^{+\times}_{kl} \boldsymbol{\psi}^k_+ \otimes \left[\boldsymbol{\psi}^l_\times\right]^t \\ \boldsymbol{\rho}^{\times+}_{kl} \boldsymbol{\psi}^k_\times \otimes \left[\boldsymbol{\psi}^l_+\right]^t & \boldsymbol{\rho}^{\times\times}_{kl} \boldsymbol{\psi}^k_\times \otimes \left[\boldsymbol{\psi}^l_\times\right]^t \end{bmatrix} \tag{56}$$

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

$$\sum_{l=1}^{N_\parallel} \sum_{q=+,\times} \boldsymbol{\sigma}^{pq}_{kl} \boldsymbol{\rho}^{qr}_{lm} = \delta_{km} \delta^{pq} . \tag{57}$$

The solution for $\boldsymbol{\rho}$ is (the $\boldsymbol{\sigma}^{++(\times\times)}$ are diagonal matrices)

$$\boldsymbol{\rho}^{\times\times} = \left(\boldsymbol{\sigma}^{\times\times} - \boldsymbol{\sigma}^{\times+} \cdot \left(\boldsymbol{\sigma}^{++}\right)^{-1} \cdot \boldsymbol{\sigma}^{+\times}\right)^{-1} \tag{58a}$$

$$\boldsymbol{\rho}^{+\times} = -\left(\boldsymbol{\sigma}^{++}\right)^{-1} \cdot \boldsymbol{\sigma}^{+\times} \cdot \boldsymbol{\rho}^{\times\times} \tag{58b}$$

with similar expression exchanging $\times, +$. We finally write the statistic $L = \mathsf{y} \cdot (\boldsymbol{\Theta})^{-1} \cdot \mathsf{y}$ as:

$$L = \sum_{k,l=1}^{N_\parallel} \sum_{p,q} \boldsymbol{\rho}^{pq}_{kl} c^k_p c^l_q \text{ , where} \tag{59a}$$

$$c^k_p \equiv \left[\boldsymbol{\psi}^k_p\right]^t \cdot (\mathsf{y}_p)_\parallel = \sum_{\mathsf{K}=1}^{M} F^p_{\mathsf{K}} \left[\boldsymbol{\psi}^k_p\right]^t \cdot \left(\mathbf{R}^{-1}_{\mathsf{KK}} \cdot \mathbf{x}_{\mathsf{K}}\right)_\parallel ; \tag{59b}$$

where we should keep in mind that the F^{I}_p functions, the coefficients σ^{pq}_{kl} and the DKL basis vectors $\boldsymbol{\psi}^k_p$ depend on the direction in the sky.

2. Simplified case: large $N_\parallel$

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

$$\left(\mathbf{R}_{(\mathsf{KK})y}\right)_\parallel = \frac{f_s}{2}\sum_{k=0}^{N_\parallel-1} S_{(\mathsf{KK})y}[k]\,\mathbf{w}_k\otimes\mathbf{w}_k^H \tag{60}$$

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 $\mathbf{\Theta}$ (see Eq. (49)) can be factored out and we obtain in analogy with the results of Anderson *et al.*[13, Sec. V C]

$$\sum_\mathsf{K} \mathsf{F}_\mathsf{K}\otimes\left(\mathbf{R}^{-1}_{\mathsf{KK}}\right)_\parallel\otimes\mathsf{F}^t_\mathsf{K} = \frac{f_s}{2}\sum_{k=0}^{N_\parallel-1}\mathbf{w}_k\otimes\mathsf{S}_\mathsf{y}[k]\otimes\mathbf{w}_k^H \tag{61}$$

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 source direction through the F_K terms

$$\mathsf{S}_\mathsf{y}[k] \equiv \sum_\mathsf{K} S_{(\mathsf{KK})y}[k]\,\mathsf{F}_\mathsf{K}\otimes\mathsf{F}^t_\mathsf{K}\,. \tag{62}$$

We can therefore rewrite the statistic in Eq. (48) as

$$L(\mathsf{x}) \;=\; \frac{2\,f_s}{N_\parallel}\sum_{k=1}^{N_\parallel-2}\left[\tilde{\mathsf{y}}[k]\right]^H_\parallel\cdot\left[\mathsf{S}_\mathsf{y}[k]\right]^{-1}\cdot\left[\tilde{\mathsf{y}}[k]\right]_\parallel \tag{63a}$$

$$\tilde{\mathsf{y}}[\mathsf{k}] \;\equiv\; \sum_{\mathsf{L}=1}^{M}\mathsf{F}_\mathsf{L}\tilde{y}_\mathsf{L}[k] \tag{63b}$$

in complete analogy with the single detector case, Eq. (29); now $\tilde{\mathsf{y}}[\mathsf{k}]$ is a 2 elements vector, whose components are just the Fourier transform of the $\mathsf{y}_{+,\times}$ time series. 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.

C. The case of correlated noise among the detectors

The problem of writing down the likelihood in the case of correlated noise among M detectors has already been studied in depth by Finn [20], who proposed to apply a transformation to the M data channels, which would de-correlate 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 adopt a different approach, motived by the hope that the cross-correlation terms will be significantly smaller than the diagonal terms, so that a perturbation expansion is possible.

If the noise is correlated the likelihood for having $\boldsymbol{x}$ in presence of a signal h is [20, Sec. III B]

$$\Lambda(\boldsymbol{x}|\,\mathsf{h}) = \exp\left[-\frac{1}{2}\langle\boldsymbol{s},\boldsymbol{s}\rangle_{\boldsymbol{R}} + \langle\boldsymbol{s},\boldsymbol{x}\rangle_{\boldsymbol{R}}\right] \tag{64}$$

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

$$\boldsymbol{s} = \mathbf{s}_1\oplus\mathbf{s}_2\oplus\cdots\oplus\mathbf{s}_M \tag{65}$$

is the direct sum of the signals at each detector. We borrow from[20] the notation

$$\langle\boldsymbol{a},\boldsymbol{b}\rangle_{\boldsymbol{R}} \equiv \boldsymbol{a}\,(\boldsymbol{R})^{-1}\,\boldsymbol{b} = \sum_{\mathsf{KL}\,kl} a_K[k]\,\left(\boldsymbol{R}^{-1}\right)_{\mathsf{KL}\,kl} b_L[l]\,; \tag{66}$$

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

$$\boldsymbol{R}\equiv E\left[\boldsymbol{n}\otimes\boldsymbol{n}\right], \tag{67}$$

that is

$$(\boldsymbol{R})_{\mathsf{KL}\,kl} = (\mathbf{R}_\mathsf{KL})_{kl} = E\left[n_\mathsf{K}[k+d_\mathsf{K}]\,n_\mathsf{L}[l+d_\mathsf{L}]\right]. \tag{68}$$

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 $(\boldsymbol{R})_{\mathsf{KL}\,kl} = (\boldsymbol{R})_{\mathsf{LK}\,lk}$ holds, which ensures that $\langle\boldsymbol{a},\boldsymbol{b}\rangle = \langle\boldsymbol{b},\boldsymbol{a}\rangle$. 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

$$\delta_{\mathsf{IJ}}\delta_{ij} = \left(\boldsymbol{R}^{-1}\boldsymbol{R}\right)_{\mathsf{IJ}\,ij} = \sum_{\mathsf{K}=1}^{M}\sum_{k=1}^{N}\left(\boldsymbol{R}^{-1}\right)_{\mathsf{IK}\,ik}(\boldsymbol{R})_{\mathsf{KJ}\,kj} \tag{69}$$

and the network likelihood can be written explicitly as

$$\Lambda(\boldsymbol{x}|\,\mathsf{h}) \equiv e^{-\frac{1}{2}s_\mathsf{K}[k]\left(\boldsymbol{R}^{-1}\right)_{\mathsf{KL}\,kl}s_\mathsf{L}[l] + s_\mathsf{K}[k]\left(\boldsymbol{R}^{-1}\right)_{\mathsf{KL}\,kl}x_\mathsf{L}[l+d_\mathsf{L}]}; \tag{70}$$

where the indices k,l label the samples, and the indices K,L label the detectors.

Given the form of $s_\mathsf{L}[l]$ (Eq. (45)) we can integrate over the nuisance parameters $h_{+,\times}$, obtaining

$$2\ln\Lambda(\boldsymbol{x}|\,\theta,\phi) = \left[\sum_\mathsf{I}\mathsf{F}_\mathsf{I}\otimes\mathbf{y}_\mathsf{I}\right]^t_\parallel\cdot\left[\sum_{\mathsf{KL}}\mathsf{F}_\mathsf{K}\otimes\left(\boldsymbol{R}^{-1}\right)_{\mathsf{KL}}\otimes\mathsf{F}^t_\mathsf{L}\right]^{-1}_\parallel\cdot\left[\sum_\mathsf{J}\mathsf{F}_\mathsf{J}\otimes\mathbf{y}_\mathsf{J}\right]_\parallel = \mathsf{y}_\parallel\cdot[\mathbf{\Theta}]^{-1}_\parallel\cdot\mathsf{y}_\parallel\,; \tag{71}$$

$\mathbf{\Theta}$ is a $2\times N_\parallel\times N_\parallel\times 2$ matrix, constructed contracting the detector indices in $\boldsymbol{R}$ and in the $2\times M$ matrix F, while $\mathsf{y}_\parallel$ is a $2\times N_\parallel$ matrix obtained contracting the detector index in F and in $\boldsymbol{y}$.

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

$$y_\mathsf{I}[i] \equiv \sum_{\mathsf{J}\,j} \left(\boldsymbol{R}^{-1}\right)_{\mathsf{IJ}\,ij} x_\mathsf{J}[j+d_\mathsf{J}] \,, \tag{72}$$

a sort of δ function filtering for multiple interferometers.

We can suppose that the cross correlations among detectors are much smaller than the internal correlations[51]: we split $\boldsymbol{R}$ into a block diagonal $\boldsymbol{D}$, and a off-diagonal $\boldsymbol{O}$:

$$\boldsymbol{R} = \boldsymbol{D} + \boldsymbol{O} \tag{73a}$$

$$(\boldsymbol{D})_{\mathsf{KL}\,kl} = \delta_{\mathsf{KL}} \left(\mathbf{R}_{\mathsf{KK}}\right)_{kl} \tag{73b}$$

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

$$\boldsymbol{R}^{-1} \approx \boldsymbol{D}^{-1} - \boldsymbol{D}^{-1}\boldsymbol{O}\left[\boldsymbol{D}^{-1} - \boldsymbol{D}^{-1}\boldsymbol{O}\left(\boldsymbol{D}^{-1} - \ldots\right)\right] ; \tag{74}$$

the inverse of the block diagonal matrix $\boldsymbol{D}$ is simple

$$\left(\boldsymbol{D}^{-1}\right)_{\mathsf{KL}\,kl} = \delta_{\mathsf{KL}} \left(\mathbf{R}_{\mathsf{KK}}^{-1}\right)_{kl} \,, \tag{75}$$

in terms of the 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

$$\mathbf{y}_\mathsf{K} \approx \mathbf{R}_{\mathsf{KK}}^{-1} \cdot \mathbf{x}_\mathsf{K} - \sum_{\mathsf{L}\neq\mathsf{K}} \mathbf{R}_{\mathsf{KK}}^{-1} \cdot \mathbf{R}_{\mathsf{KL}} \cdot \mathbf{R}_{\mathsf{LL}}^{-1} \cdot \mathbf{x}_\mathsf{L} \tag{76}$$

where we understand the shift of the data $\mathbf{x}$ and we keep the first order in $\mathbf{R}_{\mathsf{KL}}$. In Fourier space

$$y_\mathsf{K}[l] \approx \frac{2}{N} \sum_{k=1}^{N-2} \frac{e^{-i2\pi kl/N}}{S_{\mathsf{KK}}[k]} \left[\tilde{x}_\mathsf{K}[k] - \sum_{\mathsf{L}\neq\mathsf{K}} \frac{S_{\mathsf{KL}}[k]}{S_{\mathsf{LL}}[k]} \tilde{x}_\mathsf{L}[k]\right] ; \tag{77}$$

the extension to all orders is immediate

$$\boldsymbol{y} = \frac{2}{N} \sum_{k=1}^{N-2} e^{-i2\pi kl/N} \left[\boldsymbol{S}[k]\right]^{-1} \cdot \boldsymbol{x}[k] \tag{78}$$

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

$$S_{\mathsf{KL}}[k] \equiv \frac{2}{f_s} \mathbf{w}_k^H \cdot \mathbf{R}_{\mathsf{KL}} \cdot \mathbf{w}_k \,. \tag{79}$$

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 an approximation to some 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 can proceed along the same lines followed in the uncorrelated noise case. To test for the occurrence of a burst in the subspace $\mathcal{V}_\parallel$, we 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 can *estimate* the matrix $\boldsymbol{\Theta} = E[\mathsf{y} \otimes \mathsf{y}]$ and compute the elements $\boldsymbol{\sigma}$ (see Eq. (53)) exactly as before: the diagonal elements $\boldsymbol{\sigma}^{++(\times\times)}$ are diagonal matrices built from the eigenvalues $\sigma_k^{+(\times)}$, and the off diagonal matrices $\boldsymbol{\sigma}^{+(\times)}$ using Eq. (55); all the remaining derivation goes unchanged.

1. Simplified case: large $N_\parallel$

As in Section III B 2, a simplification is possible if the DKL $\boldsymbol{\psi}_{+(\times)}$ bases of the $\mathcal{V}_\parallel$ space converge to the Fourier bases: by the very definition $\boldsymbol{\Theta} = E[\mathsf{y} \otimes \mathsf{y}]$, and in analogy with Eq. (61)

$$\sum_{\mathsf{IJ}} \mathsf{F}_\mathsf{I} \otimes \left(\boldsymbol{R}^{-1}\right)_{\mathsf{IJ}} \otimes \mathsf{F}_\mathsf{J}^t \equiv \frac{1}{2} \sum_k \mathbf{w}_k \otimes \mathsf{S}_\mathsf{y}[k] \otimes \mathbf{w}_k^H \tag{80a}$$

$$\mathsf{S}_\mathsf{y}[k] \equiv \begin{pmatrix} S_{++}[k] & S_{+\times}[k] \\ S_{\times+}[k] & S_{\times\times}[k] \end{pmatrix} \tag{80b}$$

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

$$L(\mathsf{x}) = \frac{2 f_s}{N_\parallel} \sum_{k=1}^{N_\parallel - 2} [\tilde{\mathsf{y}}[k]]_\parallel^H \cdot [\mathsf{S}_\mathsf{y}[k]]^{-1} \cdot [\tilde{\mathsf{y}}[k]]_\parallel \tag{81a}$$

$$\tilde{\mathsf{y}}[k] \equiv \sum_{\mathsf{L}=1}^{M} \mathsf{F}_\mathsf{L} \tilde{y}_\mathsf{L}[k] \tag{81b}$$

with the difference that the y_K combine data from different detectors (Eq. (77), or Eq. (78)).

D. Example: network sensitivity to δ events

The "network spectral density" defined in Eqs. (62,80b) is a 2×2 matrix of spectra and (complex) cross-spectra, which depends on the sky direction: it is interesting to derive a scalar quantity to be plotted in a spherical projection to 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 $dt_{\text{burst}} = 1/f_s$, where f_s is the sampling rate in the detectors, and having amplitudes $A_+, A_\times$ in the two polarizations. We assume that the detector noises are uncorrelated, and that the data have been shifted to capture the event in every data streams at the same time index a: we have

$$\tilde{x}_\mathsf{L}[k] = \left(A_+ F_\mathsf{L}^+ + A_\times F_L^\times\right) \frac{1}{f_s} e^{-j2\pi ka/N} \tag{82}$$

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

$$y_\mathsf{L}[l] = \left(A_+ F_\mathsf{L}^+ + A_\times F_\mathsf{L}^\times\right) \frac{2}{f_s N} \sum_{k=1}^{N-2} \frac{e^{i2\pi k(l-a)/N}}{S_{\mathsf{LL}}[k]} ; \tag{83}$$

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

$$y_\mathsf{L}[a] = \left(A_+ F_\mathsf{L}^+ + A_\times F_\mathsf{L}^\times\right) \text{rms}(y_\mathsf{L}) \tag{84}$$

where we have defined

$$\text{rms}(y_\mathsf{L}) \equiv \frac{1}{f_s N} \sum_{k=1}^{N/2-1} \frac{1}{S_{\mathsf{LL}}[k]} \simeq \frac{1}{f_s^2} \int_{f_{\text{seism}}}^{f_{\text{Nyquist}}} \frac{df}{S_{\mathsf{LL}}(f)} ; \tag{85}$$

this is the same quantity resulting from the analysis in [38], where burst signals with uniform spectrum in the detection band had been considered. Next, we have

$$\mathsf{y}_{\|} = \sum_{\mathsf{L}} \mathrm{rms}\left(y_{\mathsf{L}}\right)\left(A_+ F_{\mathsf{L}}^+ + A_\times F_{\mathsf{L}}^\times\right)\begin{pmatrix} F_{\mathsf{L}}^+ \\ F_{\mathsf{L}}^\times \end{pmatrix} \tag{86}$$

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

$$\mathbf{\Theta} = \sum_{\mathsf{L}} \mathrm{rms}\left(\mathsf{y}_{\mathsf{L}}\right)\begin{pmatrix} \left(F_{\mathsf{L}}^+\right)^2 & F_{\mathsf{L}}^+ F_{\mathsf{L}}^\times \\ F_{\mathsf{L}}^+ F_{\mathsf{L}}^\times & \left(F_{\mathsf{L}}^\times\right)^2 \end{pmatrix}; \tag{87}$$

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 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 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 [38], 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 correct 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 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.

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].

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, 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 enlightnening 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-P010019-02-E.

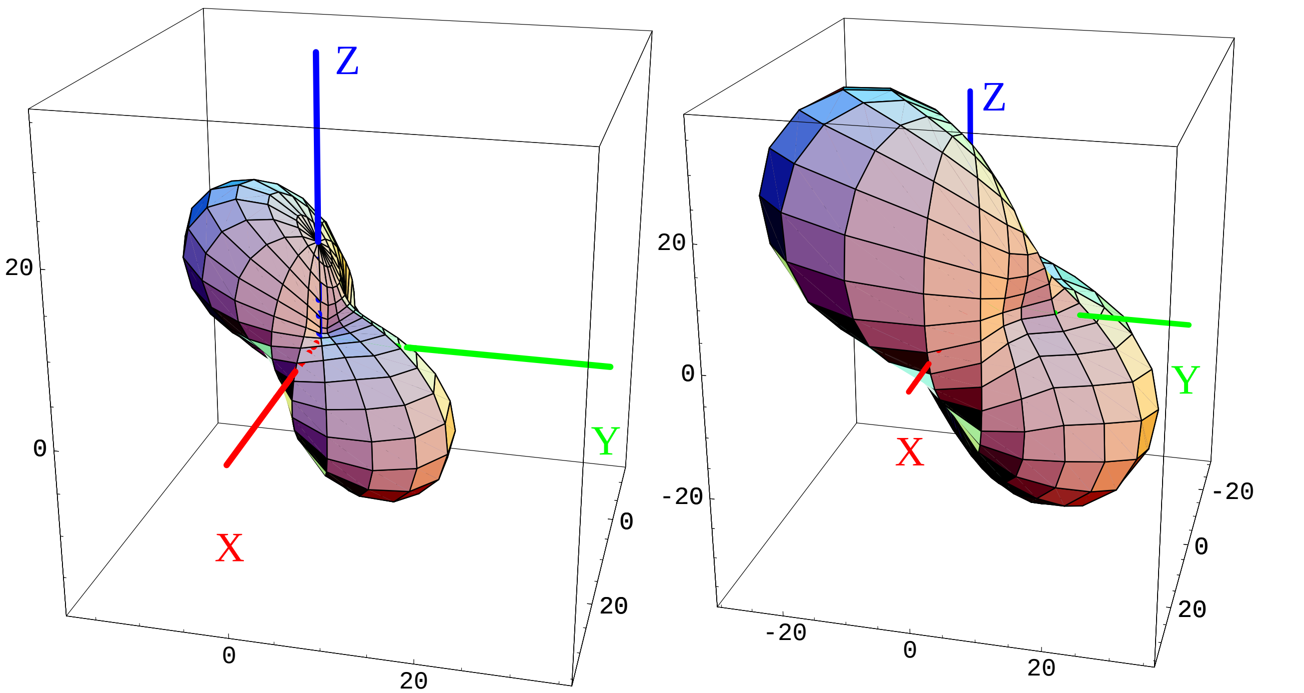

Figure 1: Polar plot of the SNR for bursts having $A\,dt = 10^{-23}s$, as a function of the source direction. The figure at left refers to the LIGO network; at right also GEO600, TAMA and Virgo are included. The network frame axes are shown: **Z** points toward the geographical north, and **X** crosses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

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 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

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

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

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

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

approximate relations because of the finite N.

Appendix B: NETWORK MODEL

We summarize the mathematics describing 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 another are expressed in terms of Euler angles

$$\begin{aligned} x_{\mathsf{wave}} &= \mathcal{O}(\phi,\theta,\psi)\cdot x_{\mathsf{network}} && \text{(B1a)}\\ x_{\mathsf{detector}_\mathsf{L}} &= \mathcal{O}(\alpha_\mathsf{L},\beta_\mathsf{L},\gamma_\mathsf{L})\cdot x_{\mathsf{network}}\,; && \text{(B1b)}\end{aligned}$$

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

$$\phi=\phi_s-\frac{\pi}{2};\ \theta=\pi-\theta_s\,. \qquad \text{(B2)}$$

The ψ angle is zero, according to the wave frame orientation. One introduces the wave tensor

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

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

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

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 2nd rank Symmetric Trace Free tensors $\mathcal{Y}_{mn}$[35, 36]:

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

where T_{mn}, with $m,n=0,\pm1,\pm2$ are 2nd rank Gel'fand's functions 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:

$$\begin{aligned} T^l_{mn}\left(\phi,\theta,\psi\right) &= e^{-i(n\phi+m\psi)}P^l_{mn}\left(\cos\theta\right), && \text{(B6)}\\ P^l_{mn}\left(\mu\right) &= \frac{(-1)^{l-m}\,i^{n-m}}{2^l\,(l-m)!}\sqrt{\frac{(l-m)!\,(l+n)!}{(l+m)!\,(l-n)!}}\frac{(1-\mu)^{\frac{m-n}{2}}}{(1+\mu)^{\frac{m+n}{2}}}\frac{d^{l-n}}{d\mu^{l-n}}\left[\frac{(1+\mu)^{m+l}}{(1-\mu)^{m-l}}\right], && \text{(B7)}\end{aligned}$$

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

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

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

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 $\sin\left(2\Omega_\mathsf{L}\right)$ 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 **d** has a simple expression

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

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

$$\begin{aligned}\mathrm{tr}\left[\mathbf{d}\cdot\left(\mathcal{Y}_{2m}\right)_{\mathsf{network}}\right] &= -i\sqrt{\frac{15}{8\pi}}\Big[T^*_{2m}\left(\alpha_\mathsf{L},\beta_\mathsf{L},\gamma_\mathsf{L}\right)\\ &\qquad -T^*_{-2m}\left(\alpha_\mathsf{L},\beta_\mathsf{L},\gamma_\mathsf{L}\right)\Big] && \text{(B10a)}\\ \mathrm{tr}\left[\mathbf{d}\cdot\left(\mathcal{Y}_{2m}\right)_{\mathsf{wave}}\right] &= \sum_n T_{mn}\left(\phi,\theta,0\right)\,\mathrm{tr}\left[\mathbf{d}\cdot\left(\mathcal{Y}_{2n}\right)_{\mathsf{network}}\right]\\ &\equiv \sqrt{\frac{15}{8\pi}}D_m\left(\phi,\theta,0,\alpha_\mathsf{L},\beta_\mathsf{L},\gamma_\mathsf{L}\right) && \text{(B10b)}\end{aligned}$$

where following [37] a short hand notation is used. Finally the signal at the L-th detector will be

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

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 $F^{L,R}_\mathsf{L}\equiv\mathrm{tr}\left(\mathbf{e}_{L,R}\cdot\mathbf{d}_\mathsf{L}\right)$ for the two left and right wave polarizations

$$F_{\mathsf{L}}^{L} \;=\; D_{-2}\left(\phi,\theta,0,\alpha_{\mathsf{L}},\beta_{\mathsf{L}},\gamma_{\mathsf{L}}\right) \qquad \text{(B12a)}$$

$$F_{\mathsf{L}}^{R}=\left(F_{\mathsf{L}}^{L}\right)^{*} \;=\; D_{2}\left(\phi,\theta,0,\alpha_{\mathsf{L}},\beta_{\mathsf{L}},\gamma_{\mathsf{L}}\right) \qquad \text{(B12b)}$$

one can alternatively write

$$\begin{aligned} s_{\mathsf{L}}(t) &= \Re\left[\left(h_{+}\left(t-\tau_{\mathsf{L}}\right)+i h_{\times}\left(t-\tau_{\mathsf{L}}\right)\right) F_{\mathsf{L}}^{R}\right] \qquad \text{(B13)} \\ &= \frac{1}{2}\left[\left(h_{+}+i h_{\times}\right) D_{2}(\ldots)+\left(h_{+}-i h_{\times}\right) D_{-2}(\ldots)\right] \end{aligned}$$

or equivalently and more conveniently for our work

$$s_{\mathsf{L}}(t)=h_{+}F_{\mathsf{L}}^{+}+h_{\times}F_{\mathsf{L}}^{\times} \qquad \text{(B14)}$$

where, reintroducing the aperture angle

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

$$F_{\mathsf{L}}^{\times} \;\equiv\; \sin\left(2\Omega_{\mathsf{L}}\right)\Im\left[D_{-2}(\phi,\theta,0,\alpha_{\mathsf{L}},\beta_{\mathsf{L}},\gamma_{\mathsf{L}})\right]. \qquad \text{(B15b)}$$

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. The orientation data are taken from[39] and updated with informations from the web sites of the collaborations[40]; 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 spherical Earth approximation, 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 [42].

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 noise spectrum model is therefore

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

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.

In addition to these parameters, we call f_{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 [41], and in Fig. (3) the comparison of the different noise spectral densities. 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

$$\begin{aligned} R_n^{-1}(\tau) \;&\simeq\; A_0e^{-|\tau|/\tau_0}\cos\left(2\pi f_0\tau+\phi_0\right) \\ &+A_1e^{-|\tau|/\tau_1}\cos\left(2\pi f_1\tau+\phi_1\right); \qquad \text{(B18)} \end{aligned}$$

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: their smallness reflects the absence of resonances in the model noises.

[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, P.Hello, Phys. Rev. D**63,** 042002 (2001).
[7] T.Pradier,N.Arnaud,M.A.Bizouard,F.Cavalier,M.Davier, 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, T.Pradier, Phys.Rev. D**65,** 042004 (2002).
[9] T.Zwerger and E.Müller, Astron. Astrophys. **267**, 623 (1993).
[10] N. Arnaud, M. Davier, F. Cavalier, 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).
[14] H.Dimmelmeier, J.A.Font and E.Müller, Astrophys.J. **560**, L163 (2001).
[15] G. Pizzella, *Fisica sperimentale del campo gravitazionale*, (La Nuova Italia, Firenze, 1993).

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.

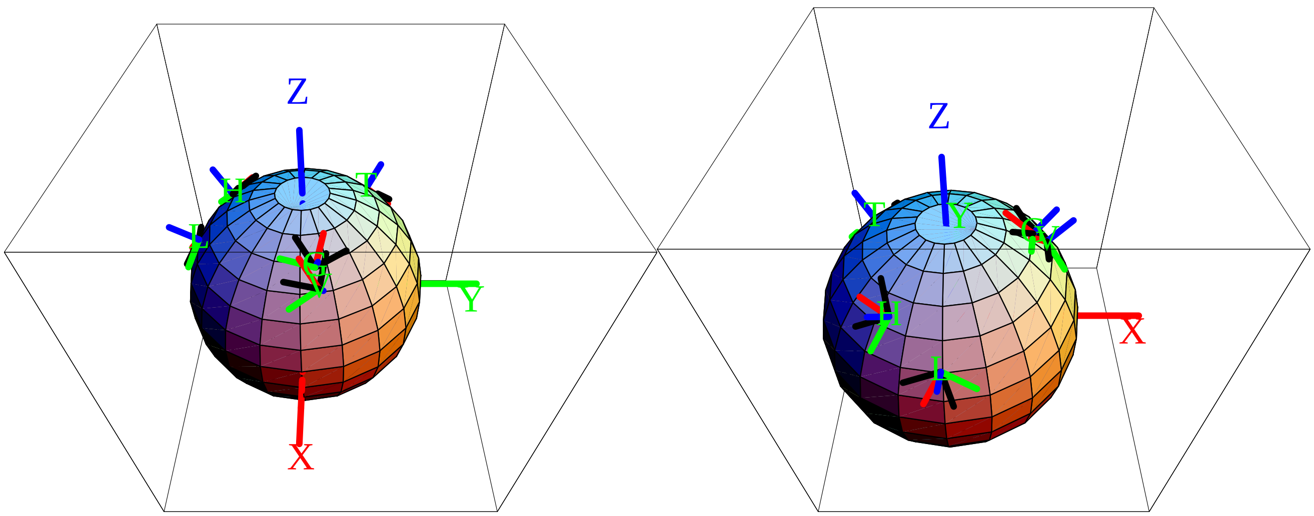

Figure 2: The locations of the detectors on Earth, labeled by their initials (H, L for LIGO Hanford and Livingston); the axes **X**,**Y**,**Z** of the network frame and of the detector frames are shown. Left: view from above Europe. Right: view from above the United States of America.

Detector	$f_{\rm seism}$	$S_{\rm pend}$	$S_{\rm mirror}$	$S_{\rm shot}$	$f_{\rm knee}$
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.

[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, F.Vetrano, Phys.Rev. D**64,** 122002 (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,

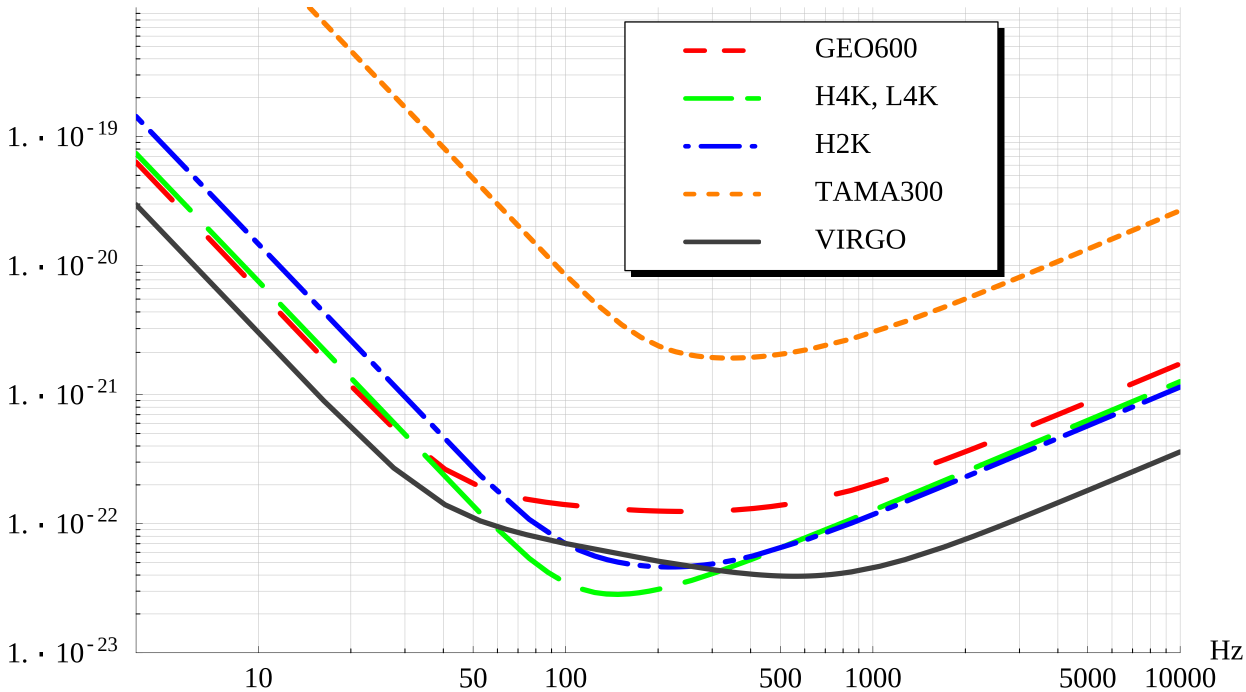

Figure 3: A comparison of the baseline spectral sensitivities, in units of $\mathrm{Hz}^{-1/2}$, for the different interferometers currently under commissioning, according to the simplified model in Eq. (B16). The two LIGO 4km interferometers "H4K" and "L4K", located in Hanford and Livingston respectively,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

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

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

[40] 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

[41] 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

[42] 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

[43] They do only when $N = N_{\|}$: in that case one would have

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

where we used the spectral factorization of $\mathbf{R}_n^{-1}$, hence $\mathbf{W}_n \cdot \mathbf{x}$ is distributed as white gaussian noise. The likelihood would be just a monotonic function of the *energy* $(\mathbf{W}_n \cdot \mathbf{x})^2$ of the whitened sample.

[44] The burst statistic reduces to a standard δ-filtering when dealing with bursts of unit duration: one has

$$\Lambda(\mathbf{x}) = \exp\left[\frac{1}{2}\frac{\left[(\mathbf{R}_n^{-1}\cdot\mathbf{x})_a\right]^2}{(\mathbf{R}_n^{-1})_{aa}}\right]$$

and the interesting statistic for a certain arrival index a is

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

which is the Wiener filter for a δ function[15].

[45] A regular 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 by a causal process: this is usually the case for physical noises.

[46] A process is 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].

[47] The statistic L would of course be optimal also for any prior being a function of the statistic L.

[48] The false alarm probability is $Q_f(L_0) \equiv \int_{L_0}^{\infty} d_0(L)\, dL = \Gamma\left(\frac{N_\|}{2}, \frac{L_0}{2}\right)/\Gamma\left(\frac{N_\|}{2}\right)$, while the detection probability$Q_d(L_0|\mathsf{SNR})$ cannot be written in closed form. One can approximate $d_1(L|\mathsf{SNR})$ with a Gaussian distribution, obtaining

$$Q_d \approx \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{N_\| + \mathsf{SNR}\sqrt{2N_\|} - L_0}{2\sqrt{N_\| + 2\,\mathsf{SNR}\sqrt{2N_\|}}}\right)\right]$$

which is valid for SNR large.

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

[50] Notice that 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.

[51] More precisely, that $\frac{||\mathbf{R}_{\mathsf{KL}}||}{\sqrt{||\mathbf{R}_{\mathsf{KK}}||\,||\mathbf{R}_{\mathsf{LL}}||}} \ll 1$, where $||\mathbf{A}|| \equiv \max_{||\mathbf{x}||=1} ||\mathbf{A}\cdot\mathbf{x}||$ is the matrix norm induced by the standard vector norm $||\mathbf{x}|| \equiv \sqrt{\mathbf{x}\cdot\mathbf{x}}$.