Improvement of the performance of a classical matched filter by an independent component analysis preprocessing

R. De Rosa, L. A. Forte, F. Garufi, and L. Milano*

Dipartimento di Scienze Fisiche, Università di Napoli Federico II and INFN, Sezione di Napoli

(Received 13 October 2011; published 7 February 2012)

Current gravitational wave searches for compact binaries coalescence are done using a bank of templates (matched filters) on each running detector. Given a network of interferometers, we propose to use a denoising strategy based on an independent component analysis which considers two interferometers at a time and then to use a standard matched filter on the processed data. We show that this method allows to lower the level of noise and increases the signal-to-nose ratio at the output of the matched filter.

DOI: 10.1103/PhysRevD.85.042001 PACS numbers: 04.80.Nn, 07.05.Kf

I. INTRODUCTION

It is commonly believed that gravitational waves (GWs) first detections will come in the next few years. This will be possible thanks to technological upgrades of current laser interferometers (in particular Advanced VIRGO [1] and Advanced LIGO [2]), which will lead a better sensitivity and a bigger volume of the detectable Universe. A final answer to the very first detection (and the following ones) will be reliable only thanks to a coincident signal in all the running antennas. Therefore, a network of GWs interferometers is and will be mandatory. Since the events rate is still rather uncertain, also for the advanced detectors [3,4], and lacking any direct observation (i.e. any true signal) up to now, it is of fundamental importance to use data analysis techniques which maximize the detection probability and at the same time allow to estimate the physical parameters of the astrophysical event in the most precise possible way. With this aim, we propose a method which uses the outputs of two interferometers. We show that applying an independent component analysis (ICA) preprocessing to the data coming from two interferometers, the performance of a classical matched filter is enhanced. Roughly, since any ICA algorithm separates independent components from a mixture (we are interested in separating the ‘‘noise’’ from the ‘‘signal’’ and they are statistically independent), we find that after a preprocessing based on ICA only in one of the resulting components the GW signal will be present, whereas the noise will contaminate the other. If a classical matched filter is applied to the processed time series, it results in a bigger correlation with respect to the matched filter applied directly to the original, unprocessed data. This clearly indicates that the detection probability is higher and since one of the obtained independent component will be less noisy than the other, also the parameters estimation will improve. We focus here only on coalescing binaries: it is well known that the expected GW signal is a so-called chirp signal (see [5] for an excellent review). Nevertheless, the analysis can be run also for other GW signals, for example, gravitational bursts.

In Sec. II, we review basic concepts about matched filters and ICA. In Sec. III, we describe our analysis and show that the preprocessing we propose is useful. Finally, in Sec. IV we summarize the results of our work and outline possible future and conclusive refinements of our investigations on ICA for GW data analysis.

II. GENERALITIES ON MATCHED FILTER AND INDEPENDENT COMPONENT ANALYSIS

In this section, we briefly review the concepts of matched filter (MF) [6] and ICA [7], focusing only on the aspects useful to our purposes.

A classical result states that if one knows exactly the shape of the signal, the matched filter is the optimum (linear) filter which maximizes the output signal-to-nose ratio (SNR) among all the filters. This is true in presence of additive, white, Gaussian noise [6]. If the noise is not Gaussian or correlated (colored), the performance of the matched filter is degraded. This is also the reason to have an instrument with an output noise that is as much Gaussian and stationary as possible. Given an observed signal $x(t) = s(t) + n(t)$ (signal plus noise), the matched filter output can be written as the scalar product (in the frequency domain) between the measured signal and the template, weighted by the power spectral density of the instrument

$$m(t) = 4 \int_0^{+\infty} \frac{\tilde{x}(f)\tilde{s}^*(f)}{S_n(f)} e^{2\pi i f t} df \tag{1}$$

where $S_n(f)$ is the one-sided power spectral density of the noise and s is the template used for the detection (we use definitions similar to the ones accepted by the GW community, as reported, for example, in [9]). If one evaluates the quantities

$$\rho(t) = \frac{|m(t)|}{\sigma}, \tag{2}$$

*forte@na.infn.it

$$\sigma^2 = 4 \int_0^{+\infty} \frac{|\tilde{s}(f)|^2}{S_n(f)} df \tag{3}$$

where σ is a kind of template energy weighted by the power spectral density of the noise, then $\rho(t)$ is the amplitude signal-to-noise ratio and it will have a maximum in correspondence of the time at which the signal has been injected. The value of this maximum, with this normalization, is equal to the input SNR.

ICA is a way to solve, for example, the classical cocktail party problem. This refers to the situation in which one has several people speaking with their voices recorded by a certain number of microphones. If the number of people and microphones is the same (the so-called determined ICA), it is possible, under some quite general assumptions, to separate (or estimate) the original voices (the waveforms) using only the observed data, i.e. the recorded time series [10]. Note that since noise always enters in the recording procedure (through echo effects, environmental noise, or an original noise source rather than a clean voice), the separation process is also a denoising technique. For more details, see [8,11]. Thus, let $\mathbf{x}$ be the vector of the observations and suppose that it is linearly related to the original sources by a square matrix $\mathbf{A}$ (the so-called mixing matrix)

$$\mathbf{x} = \mathbf{A}\mathbf{s}. \tag{4}$$

The purpose of any ICA algorithm is to estimate the sources $\mathbf{s}$ using only information contained in the data $\mathbf{x}$. Of course, the mixing matrix, representing the full dynamics of the physical system (with or without noise components), is not known, otherwise the estimation problem would be trivial. Indeed, one has to estimate at the same time the demixing matrix $\mathbf{W}$, i.e. the inverse of $\mathbf{A}$, and the original components. In order to do that, one can make different assumptions on the nature of the sources, depending on the problem one has to face. These assumptions select the separation criteria. The most interesting ones are based on the non-Gaussianity of the sources (all of them except at most one), on their different spectral contents and on their nonstationarity. In particular, ICA was formulated initially to separate non-Gaussian components and a corresponding algorithm was developed, named FastICA [12]. FastICA is a classical ICA algorithm which uses, for separation, higher-order statistical methods, for example, concepts like mutual information and neg-entropy. The computational complexity of FastICA is $O(dN)$, where N is the number of samples and d is the number of sources. The main iterations of FastICA are the following:

(1) choose an initial random vector $\mathbf{w}$,
(2) $\mathbf{w}^\dagger = E\{\mathbf{x} g(\mathbf{w}^{tr}\mathbf{x})\} - E\{g'(\mathbf{w}^{tr}\mathbf{x})\}\mathbf{w}$,
(3) $\mathbf{w} = \frac{\mathbf{w}^\dagger}{||\mathbf{w}^\dagger||}$,
(4) if not converged, go back to 2

where g is certain nonlinearity function (and g' is its first derivative). The FastICA algorithm tries to maximize the non-Gaussianity of $\mathbf{w}^{tr}x$ (which is an estimated independent component) measuring the neg-entropy of $\mathbf{w}^{tr}\mathbf{x}$. For more details, see [11]. Successively, more algorithms have been developed to cope also with different situations. A famous second-order algorithm has been developed in [13], and named SOBI (second-order blind identification). This takes into account the spectral properties of the sources, and it is possible to separate also Gaussian components. SOBI uses second-order statistical methods based on covariance matrices. In fact, the algorithm evaluates a certain number of covariance matrices at different time lags and by a joint diagonalization of these latter it is possible to estimate the original components. The computational complexity of SOBI is $O(d^4M)$ where d is the number of sources and M the number of covariance matrices involved. For our simulations, we have used some refinements of FastICA and SOBI known as EFICA (Efficient FastICA) [14,15] and WASOBI (Weights Adjusted SOBI) [16], respectively. These latter are implemented in a simple and efficient way in MATLAB code (see [17] for a MATLAB toolbox which contains also more BSS algorithms and the webpages [18] for the original version of the codes). EFICA has a complexity which is only slightly higher than that of the original FastICA algorithm. In particular, the test of saddle points and the adaptive choice of the nonlinearity function (which are the main refinements of the algorithm) have complexity $O(d^2N)$ and $O(dN)$, respectively. WASOBI has a computational complexity $O(d^6 + d^3M^3)$, but the number M of covariance matrices involved can be significantly lower than that of SOBI.

Before moving on to our simulations, we need to remind that ICA has an ambiguity which cannot be eliminated. In fact, there is not a reasonable or simple (i.e. natural) criterion to order the output components, in other words the output of any ICA algorithm lacks any ordering definition. This is different from what happens, e.g., in the more popular principal component analysis, where one selects the principal components according to an energy criterion, i.e. the ones corresponding to the biggest eigenvalues of the covariance matrix. The problem of ordering is an important point in our analysis and will be discussed in the next section.

Finally, in this paper we use determined ICA algorithms, which assume the same number of sensors and sources. This situation is not exactly what one encounters in real life, since one has a certain number of interferometers (the sensors), each one with its own noise, and one or more GW signals (and the situation is even more intricate due to GW polarizations, different antenna patterns, etc). We are confident that determined ICA methods may work because the noises in the interferometers, although stochastic processes, are similar. Thus, one can hope to achieve at least a good denoising preprocessing if not a perfect separation; and we show that this is the case. To be more sophisticated

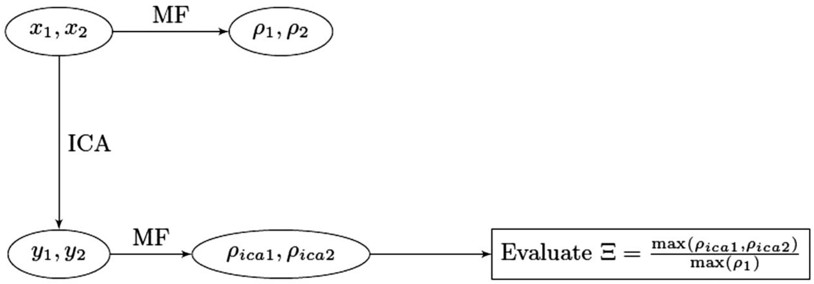

FIG. 1. The flowchart of our analysis.

and correct, one should use under-determined ICA methods, which assume a lesser number of sensors than sources. This is, of course, a more difficult problem: the same well-posedness of the problem (i.e. the existence and the uniqueness of the solution) is still open in the general case. In recent years, there has been a growing literature on this subject, most approaches assume sparseness of the sources, a concept which might not be true for GW signals. Other approaches seem to be more promising, especially the ones where one assumes the knowledge of (a part of) the mixing matrix, which in the GW case corresponds to targeted searches (tracking of a known source in a known direction *without* assuming the shape of the signal). We hope to investigate this idea in forthcoming papers.

An interesting approach about how to cope with the under-determined case is given in [19], which uses classical, determined ICA methods (in the same way as we do in this paper) followed by a so-called empirical mode decomposition in order to retrieve the original sources. The authors show that their method is effective in a low-noise environment, which is the typical case for GW experiments. Reference [19] also contains a quick introduction to the concept of independent component analysis. Another way to look at the under-determined case has been studied in [20] with the tool of the Wigner-Ville transform which has the important property to localize (linear) chirp signals in the time-frequency plane. Reference [20] contains a more detailed introduction to the ICA methods used also in this paper.

III. SIMULATIONS AND RESULTS

We consider the ideal case of data coming from two interferometric antennas: we assume that the (calibrated) data have been (temporally) whitened and neglect time delays, GW polarizations, etc. [21]. Basically, we inject the same chirp signal into two independent time series of WGN (White Gaussian Noise) at a certain SNR. Given the noise assumptions, the input SNR can be defined as

$$\mathrm{SNR}^2 = E(\mathrm{signal})^2/\mathrm{var}(\mathrm{noise}) \tag{5}$$

where E is the signal energy. In practice, given a certain chirp signal $s(t)$ (that we have generated with its own physical units using LAL [22], a collection of libraries written in C) we generate two independent realizations n_1, n_2 of WGN (with unit power) and then we add the signal to the noise at the chosen SNR [23], i.e.

$$x_i(t) = n_i(t) + \mathrm{SNR} * \frac{s(t)}{|s(t)|^2} \quad (i = 1, 2). \tag{6}$$

The two time series x_1, x_2 model the output of two different GW interferometers. Since we assume to know the GW signal, the best strategy is to apply the matched filter. This is what is generally used in the search for GWs: in the case of coalescing binaries a bank of matched filters is used in order to cope with the different physical parameters of the astrophysical system. Finally, the analysis takes advantage of the fact that there exists a network of antennas and a coincident search is performed. Given a couple of interferometers, we propose to preprocess the data by an independent component analysis before applying a matched filter search. In fact, as we mentioned in the previous section, ICA separates a certain number of input time series (2 in our case) on the basis of a certain criterion (we focus on non-Gaussianity or spectral diversity) and the idea is to separate the interferometers noise as much as possible from the GW signal. As we show, this will lead in an enhancement of the output SNR after applying the matched filter, with respect to the case in which one only applies the matched filter [24]. This is a desirable property considering that the events rate is not so high [25] and that the true gravitational waveforms could (substantially or marginally) differ from the (approximate) analytical waveforms that we have now or the ones given by numerical relativity methods.

The flowchart of our analysis is simple and is shown in Fig. 1 (all simulations have been made in MATLAB). In one case, we apply the matched filter to $x_1(t)$ and $x_2(t)$ separately, obtain two outputs (i.e. two correlations with the clean signal we have injected) and record the maximum of these. In the other case, we first apply an ICA method and obtain two time series, say $y_1(t), y_2(t)$. On these latter, we apply the matched filter. The result is that the MF on a processed time series will give a better correlation with the template (i.e. a higher value for the maximum) and the MF filter on the other processed time series will give a negative

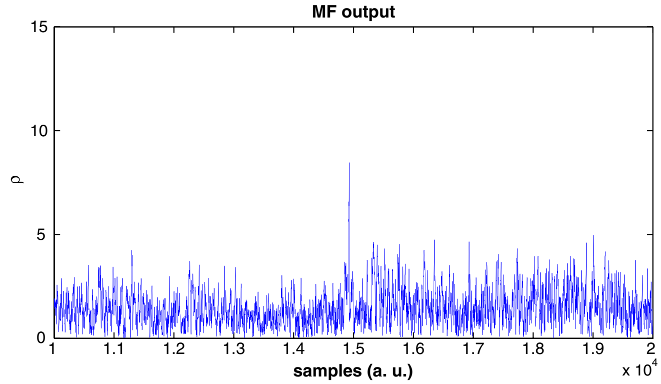

FIG. 2 (color online). Typical matched filter output (SNR = 8).

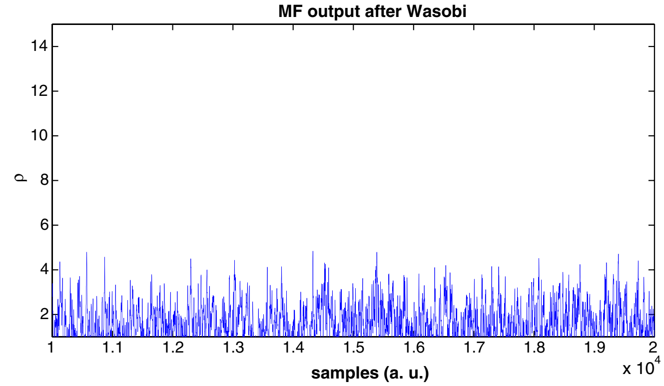

FIG. 4 (color online). Typical output for the matched filter between the template and the other time series after an ICA algorithm (SNR = 8). Note the absence of any clear maximum.

result (no signal detected). This confirms our previous results [20]: classical (determined) ICA techniques work as a powerful denoising strategy and can be very useful if used in a network of GW interferometers.

These statements can be easily understood looking at the Figs. 2–4. In Fig. 2, the typical output of the matched filter on series $x_1(t)$ is shown (the output of the matched filter for series $x_2(t)$ is not shown, but it is completely analogous). It corresponds to an SNR = 8. Figures 3 and 4 show the output of the matched filter after applying an ICA separation to the series x_1 and x_2. It is clear that the matched filter gives a higher correlation only with one of the processed time series and no correlation with the other. We call Ξ the ratio between peaks as in Figs. 3 and 2. To quantify the power of our method, we have run 10 000 simulations, each time injecting the same chirp signal into two independent noisy time series. The performance of our results is measured by the probability that $\Xi > 1$, i.e. how many times the separation through ICA gives a better correlation with respect to the plain matched filter. These probabilities are given in Table I. As one expects, the separation performance improves by increasing the SNR, and, as already noted in [20], the WASOBI algorithm is slightly better than EFICA. Probability distributions for these ratios (which measure the enhancement of the output SNR) are given in Fig. 5. One can see that, on average, the (linear) SNR is boosted by a factor around 1.5 in most cases; whereas only in 1%–2% of cases (as reported in the table), a simple matched filter is better; however the minimum value for the ratios in Fig. 5 is always between 0.8 and 0.9.

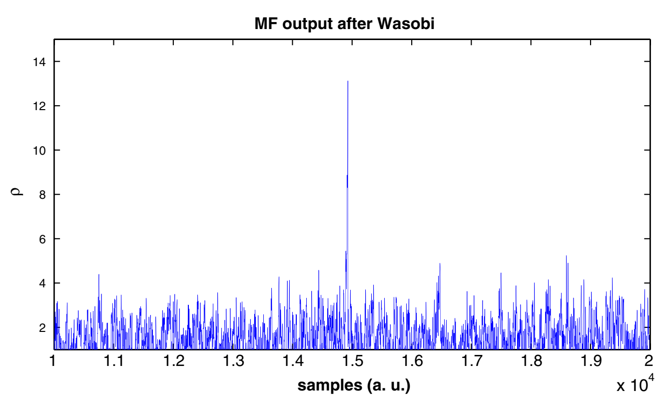

FIG. 3 (color online). Typical output for the matched filter between the template and one of the two time series after an ICA algorithm (SNR = 8). Note the value of the maximum.

Final plots summarizing the results of our analysis are shown in Figs. 6 and 7. Figure 6 shows the CDF (cumulative distribution function) of Ξ for various SNR (in accordance to Fig. 5). Figure 7 shows the CDF of the output SNR, more precisely the CDF of the peaks at the output of the matched filter (MF only, EFICA + MF and WASOBI + MF). The green curve is the usual MF curve and can be reproduced (for WGN) also via theoretical/analytical tools (see [6]). Basically, it is related to the so-called receiving operator characteristic of the matched filter for different SNR. The true receiving operator characteristic curve contains also the information regarding the detection probability and the false alarm probability. One can give different but equivalent interpretations to the curves in Fig. 7. In fact, one can say that the preprocessing based on ICA allows to increase (thanks to a network of interferometers) the detection probability and to lower the false alarm probability with respect to the plain matched filter, keeping fixed the input SNR. This is possible because part of the noise has been taken off through the separation algorithm, thus the optimality properties of the matched filter are not violated, of course.

TABLE I. Probability that $\Xi > 1$ for the various algorithms at different SNR.

SNR	EFICA	WASOBII
8	98.05%	98.27%
10	98.33%	98.96%
12	98.60%	99.30%

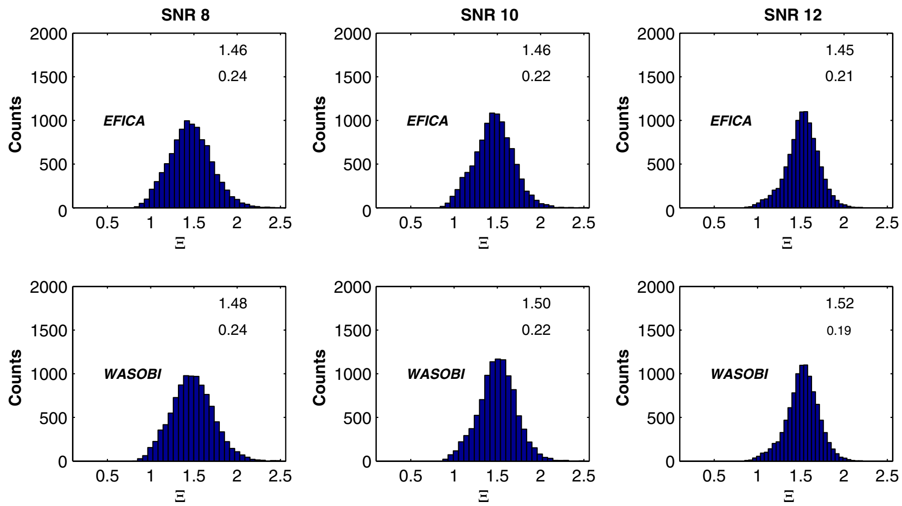

FIG. 5 (color online). Probability density function for the ratio between peaks Ξ for the various SNR. Mean and standard deviation are indicated for each distribution.

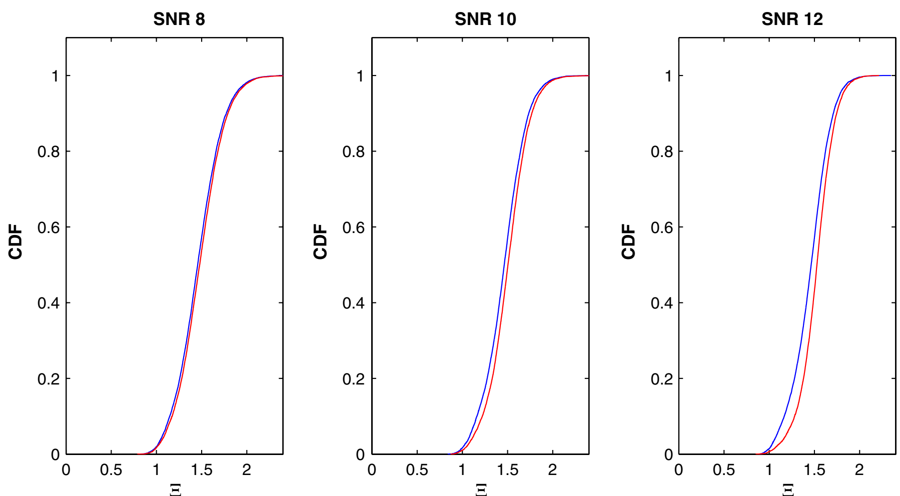

FIG. 6 (color online). CDF for the ratio between peaks Ξ. The two curves (blue) EFICA, (red) WASOBI tend to separate with an increasing SNR.

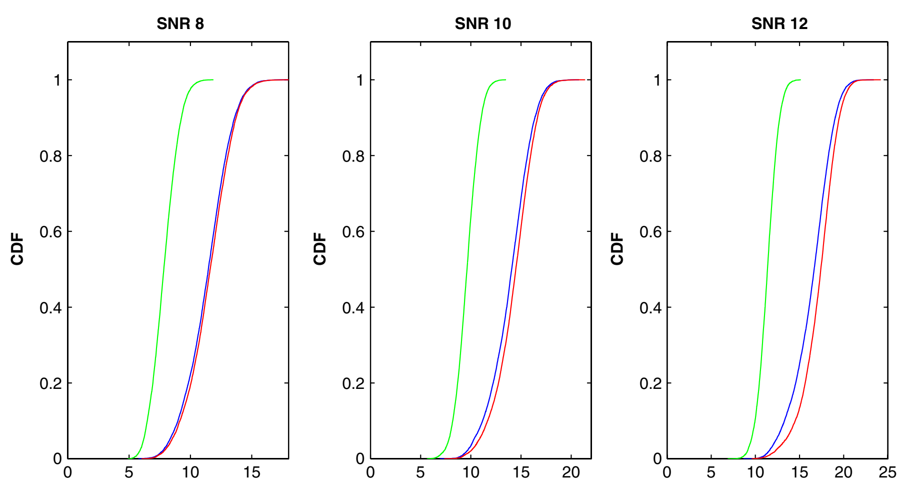

FIG. 7 (color online). CDF for the recovered SNR. Green (MF), Blue (EFICA), Red (WASOBI).

Before moving to the conclusions, we need to clarify the following point. An important issue concerns whether to run ICA (i.e. separation) on a whole time series and run the matched filter on some segments inside the original time series or rather split the original time series into a certain number of segments and run both ICA and the matched filter on each segment. Indeed, since any ICA algorithm has an inherent ambiguity problem for the ordering of the output components, it would be dangerous to run ICA on the same segments on which one runs the matched filter. In fact, it might happen that part of the searched signal, a chirp in our case, could be present in, say, segment 1 and segment 2. Running ICA on segment 1 could move part of the chirp to a certain output component and move part of the chirp in segment 2 to another output component, and one, *a priori*, is not able to concatenate in a sensible way the output components coming from segments 1 and 2. This is the main reason we have chosen to separate on the whole time series and run the matched filters on subsegments as usual. In this way, one can have a look also at the whole processed time series after separation by ICA, for example, investigating time-frequency plans (as in [20]) in a consistent way.

IV. CONCLUSIONS AND FUTURE PERSPECTIVES

In this paper, we have investigated the possibility to use an ICA preprocessing to data coming from a network of interferometers for the purpose of decreasing the level of noise and increase the SNR at the output of the matched filter. This leads to a bigger detection probability.

Note that our considerations are based on the contemporary use of the data from two GW antennas: this is an important point. Indeed, current detection pipelines (real-time or offline) run a matched filter on all the operating interferometers *separately*, and produce coincident triggers only when the temporal delays of the single triggers are compatible with the ones of a passage of a GW. We investigate the possibility to use the network of antennas before applying the matched filter with the goal to reduce the noise and increase the detection probabilities. Note that our analysis requires running the matched filter twice with respect to single-interferometer pipelines, because one has to run the matched filter twice for each pair of interferometers. This is because any ICA algorithm has an inherent ambiguity which is the order of the output components: there is no way to select interesting components. Thus, one needs to run the matched filter on every estimated component. This is the price to pay in order to increase the detection capabilities. The total computational complexity is the one of the matched filter (as described, for example, in [9] in relation to GW searches) augmented by the complexity of the chosen ICA algorithm as reported in Sec. II.

Since our method boosts the output SNR [26], it is clear that also the maximum size of the detectable Universe grows. In fact, if one defines, as usual [9], the horizon distance as the distance of a particular binary system (two neutron stars, $1.4M_\odot - 1.4M_\odot$, optimally oriented) that would produce a signal-to-noise ratio equal to 8, then our method allows to search for GWs a little bit further. In fact, since the horizon distance is proportional to the linear signal-to-noise ratio, it is clear that with our preprocessing one can gain a factor of about 1.5. A similar phenomenon happens for a coherent search rather than a coincident one; in fact, our method is a kind of coherent search, focused on denoising.

One more issue regards the fact that compact binaries coalescence search is blind, in the sense that one does not know *a priori* the parameters of the binary, and a bank of filters is used. This means that a more interesting comparison needs to be done between a matched filter on a bank of templates with and without the preprocessing we have proposed. This is a much more fair and interesting comparison, since the optimality property of the matched filter is lost using a bank of filters (whatever be the chosen value of the fitting factor), thus any preprocessing which separates part of the noise could be very useful, especially in the cases of low SNR events. We will discuss this point in a forthcoming paper.

ACKNOWLEDGMENTS

We thank the referee for useful suggestions and careful reading of the manuscript.

[1] https://wwwcascina.virgo.infn.it/advirgo/.
[2] G. M. Harry (the LIGO Scientific Collaboration), Classical Quantum Gravity **27**, 084006 (2010).
[3] V. K. *et al.*, Astrophys. J. **601**, L179 (2004).
[4] J. A. *et al.*, Classical Quantum Gravity **27**, 173001 (2010).
[5] M. Maggiore, *Gravitational Waves: Theory and Experiments* (Oxford University, New York, 2007), Vol. 1.
[6] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection* (Prentice-Hall, Englewood Cliffs, NJ, 1998), Vol. 2.
[7] Note that for some of our algorithms the term blind source separation (BSS) would be more appropriate, see [8]. We use the term ICA (which is a way to solve the BSS problem) only for simplicity.
[8] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing* (Wiley, New York, 1998).
[9] B. Allen, W. G. Anderson, Patrick R. Brady, Duncan A. Brown, and J. D. E. Creighton, arXiv:gr-qc/0509116v2 [Phys. Rev. D (to be published)].

[10] The separation procedure does not require any *a priori* knowledge of the signals, i.e. no templates are needed by any ICA algorithm.

[11] J. K. A. Hyvarinen and E. Oja, *Independent Component Analysis* (Wiley, New York, 1998).

[12] http://www.cis.hut.fi/projects/ica/fastica/.

[13] J.-F. C. A. Belouchrani, K. Abed-Meraim, and E. Moulines, IEEE Trans. Signal Process. **45**, 434 (1997).

[14] Z. Koldovsky, P. Tichavsky, and E. Oja, IEEE Transactions on Neural Networks **17**, 1265 (2006).

[15] Z. Koldovsky, P. Tichavsky, and E. Oja, IEEE Trans. Signal Process. **54** 1189 (2006).

[16] A. Y. P. Tichavsky, E. Doron, and J. Nielsen, Proceedings of EUSIPCO, Florence, Italy, 2006 (unpublished).

[17] http://bssgui.wz.cz/.

[18] http://si.utia.cas.cz/Tichavsky.html.

[19] Shih-Lin Lin, Pi-Cheng Tung, and Norden E. Huang, Phys. Rev. E **79**, 066705 (2009).

[20] L. A. Forte, F. Garufi, L. Milano, R. P. Croce, V. Pierro, and I. Pinto, Phys. Rev. D **83**, 122006 (2011).

[21] A possible use of our analysis is for the purpose of a better estimation once standard pipelines have found a coincident trigger, in other words an interesting event. In fact, we can imagine to use our methods by whitening original series, time-shifting, etc.

[22] https://www.lsc-group.phys.uwm.edu/daswg/.

[23] We have chosen values for SNR equal to 8, 10, and 12.

[24] Note that this fact is not in contradiction with the optimal properties of the matched filter, since the data are preprocessed.

[25] Indeed, according to [3,4], tens of compact binaries coalescence events are expected for the advanced generation of ground-based detectors. Yet, the uncertainties on the events rate are quite big.

[26] One could argue that in this way also the SNR of detector glitches is enhanced. This is true as long as one encounters coincident glitches. Usually, coincident (say triple) glitches are very rare, and they are killed by instruments vetoes (data quality cuts) if they are not genuine GW signals.