Egc: A Time-Frequency Augmented Template-Based Method For Gravitational Wave Burst Search In Ground-Based Interferometers

A.-C. Clapson, M.-A. Bizouard, V. Brisson, F. Cavalier, M. Davier, P. Hello, N. Leroy, M. Varvella

EGC: a time-frequency augmented template-based method for gravitational-wave burst search in ground-based interferometers

André-Claude Clapson [‡] [§], Matteo Barsuglia, Marie-Anne Bizouard, Violette Brisson,

Fabien Cavalier, Michel Davier, Patrice Hello, Nicolas Leroy, Monica Varvella

LAL, Université Paris-Sud 11, BP 34, 91898 Orsay, France

(Dated: December 2, 2007)

Abstract

The detection of burst-type events in the output of ground gravitational wave detectors is particularly challenging. The potential variety of astrophysical waveforms, as proposed by simulations and analytic studies in General Relativity and the discrimination of actual signals from instrumental noise both are critical issues. Robust methods that achieve reasonable detection performances over a wide range of signals are required.

We present here a hybrid burst-detection pipeline related to time-frequency transforms while based on matched filtering to provide robustness against noise characteristics.

Studies on simulated noise show that the algorithm has a detection efficiency similar to other methods over very different waveforms and particularly good timing even for low amplitude signals : no bias for most tested waveforms and an average accuracy of 1.1 ms (down to 0.1 ms in the best case). Time-frequency-type parameters, useful for event classification, are also derived for noise spectral densities unfavorable to standard time-frequency algorithms.

[‡] To whom correspondence should be addressed (clapson@lal.in2p3.fr)

[§] currently at MPI-K, Heidelberg, Germany.

I.   INTRODUCTION

Production of high frequency gravitational waves only depends on the dynamics of high mass density systems. Expected sources are relativistic astrophysical systems, that is, compact and massive objects in excited states. The most studied gravitational wave (GW) source candidates, though not observed as such yet, are neutron stars and black holes, notably at the formation stage, or the inspiral of a binary system of compact objects. Other mechanisms have been proposed to produce GW such as from neutron-star quakes and cosmic strings. In any case, a large independence from electromagnetic emission can be expected. For this reason, GW are a potentially rich source of information on objects and matter states that have been inaccessible so far.

To observe these waves, km-scale interferometers, TAMA (300 m), GEO (600 m), LIGO [1] (4 km) and Virgo [2] (3 km), are now in operation. This work is motivated by the search for short duration 'bursts' associated, for instance, with gravitational supernovae (SN), in the data of these instruments.

Available information on the signals come from the equations of General Relativity solved either in analytical form, as pursued for binary systems or numerically, in combination with hydrodynamics, nuclear and particle physics models for SN.
Gravitational supernovae (essentially type II supernovae) would be typical burst-like GW emitters, with signal duration below 0.1 second and a large variety of waveforms, according to simulations (see [4] and references therein). A catalog of waveforms [5] referred to hereafter as DFM, from variations of the parameters of the progenitor and neutron star equation of state, presents three waveform families, two of which were subsequently found elsewhere. Recent studies [6] considered 3-dimensional stellar models with realistic progenitors, equations of state and neutrino transport treatment, not found simultaneously before. They indicate the reliability of the core bounce as a source of GWs, to be followed by damped oscillations more dependent on model details. The most frequent waveform family would correspond to the 'type I' of DFM : a short (< 5 ms) peak or double peak of large amplitude followed by much smaller fluctuations.
Nevertheless, alternative possible mechanisms, for instance in non-axisymmetric super-

2

novae [7], through novel gravitational wave production paths [8] or more exotic ones, cannot be ruled out. For existing detectors at the target sensitivity, conservative estimates in observation occurrence rates and amplitudes remain low : a few events per century for supernovae [9] and a signal to noise ratio expected below 10.

For data analysis aiming at burst detection, a local (in time and frequency) energy or amplitude representation of time series, as provided in the time-frequency (TF) space (see [10] for an overall introduction of the topic), could be valuable to achieve signal isolation from noise with little prior knowledge on waveforms.

Many time-frequency detection algorithms have been developed recently for this purpose, including the Excess Power approaches of [11] and [13], a wavelet-based algorithms, like Waveburst [12], and the Q transform [14].

We turned for this work to the classical signal detection matched filter, developed by Wiener and discussed in [16]. Our choice of templates and subsequent event selection process borrow from regular time-frequency methods.

The Exponential Gaussian Correlator algorithm proposed here offers similar performance levels to what is achieved by other methods, with particularly good timing accuracy for low amplitude signals : no bias for most tested waveforms and an average accuracy of 1.1 ms (down to 0.1 ms in the best case) over a broad range of waveforms. Furthermore, it is not prevented by noise spectral features from providing time-frequency-type parameters.

Missing in the description of noise by its spectral density are non-stationarities and bursts of noise, or glitches. These are likely to generate event candidates, unless some discriminating parameter can be applied. Given the uncertainty on burst signal waveforms, in the context of single instrument analysis it can be argued that glitch identification falls under instrument characterization studies. Therefore we do not discuss their effect on final detection performances, and do not try to include them in the noise series used for Monte Carlo studies.

After a description of the analysis method, implementation issues and parameter choices are evoked, along with the definition of output events and the quantities that can be extracted from them. Monte Carlo simulations are then presented and detection performances,

notably the accuracy of event descriptors, are discussed.

II.   EXPONENTIAL GAUSSIAN CORRELATOR

A.   The EGC transform

EGC is a template-bank matched filter inspired by the S Transform [18] time-frequency algorithm. Matched filtering is a classical approach, introduced for instance in [16]. It consists of a correlation between the data and a fixed test waveform (the template). Normalization by the estimated noise spectral density (given by the Power Spectral Density) allows its use on any time series with a stable PSD.

For a time series h, match with a template s is defined as :

$$c(t) = \frac{< s(t), h(t) >}{||s||} \tag{1}$$

$$with \qquad < s(t), h(t) > = \int_{-\infty}^{\infty} \frac{\hat{s}(f)\hat{h}^*(f)}{S_h(f)}\, df$$

$$and \qquad ||s|| = max_t < s(t), s(t) >^{1/2}$$

with $\hat{}$ indicating the Fourier transform and $S_h(f)$ the spectral power density of the noise. The norm is the maximum value of the match, over time.

The signal to noise ratio (SNR) is the value of c obtained when the test waveform and the signal present in the noise time series are identical. The SNR then depends on the characteristics of the noise.

Using a list, or bank, of templates translates the uncertainties on the potential signal. This approach is a standard in gravitational wave detection for methods dealing with inspiral events, whose waveforms are approximated analytically. See [17] for a discussion on such an analysis.

In this work, the templates are Exponential Gaussian (EG) (fig. 1) :

$$\phi_{f_0,Q}(t) = e^{-t^2 2\pi^2 f_0^2/Q^2} e^{-i2\pi f_0 t}$$

$$\hat{\phi}_{f_0,Q}(f) = \frac{1}{\sqrt{2\pi}} \frac{Q}{f_0} e^{-(\frac{f-f_0}{f_0})^2 Q^2/2} \tag{2}$$

with two free parameters, the oscillation frequency f_0 and a quality factor [19] $Q = \sigma \times 2\pi f_0$, which gives a duration of the order of $2\sigma = \frac{Q_0}{\pi f_0}$.

The complete algorithm adds an *ad hoc* treatment of the resulting coefficients and TF representation capability, introduced below. The normalization can also be interpreted as a data 'whitening' procedure [15] meant to remove the effects of the colored PSD.
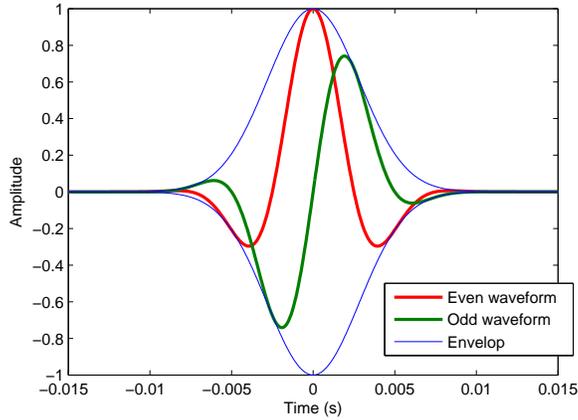

FIG. 1: Time domain template: the tested waveform is the complex sum of the two quadratures, here for $f_0 = 100$ Hz and $Q = 1$.

The match between time series $h(t)$ and a given template is, expressed in the Fourier domain :

$$c_{(f_0,Q)}(t) = \frac{1}{\sqrt{2\pi}} \frac{Q}{f_0} \int_{-\infty}^{\infty} \frac{e^{-(\frac{f-f_0}{f_0})^2 Q^2/2} \times \hat{h}^*(f)}{S_h(f)} e^{i2\pi ft} \, df(t) \tag{3}$$

In this case (complex-valued templates) c is a complex. Its squared modulus $y = |c|^2$ provides an energy-analog quantity, tat will serve here, while $\rho = \sqrt{2 \times y}$ is the usual SNR.

While no signal could match a complex domain template, the orthogonality (defined as a null correlation) between the real and imaginary parts

$$e^{-t^2 2\pi^2 f_0^2/Q^2} \cos(2\pi f_0 t) \qquad \text{and} \qquad e^{-t^2 2\pi^2 f_0^2/Q^2} \sin(2\pi f_0 t) \tag{4}$$

makes this implementation equivalent to the use of both Gaussian Cosine and Gaussian Sine real templates.

5

B. Implementation

1. Discrete form

In discrete form, the EGC transform reads :

$$|c|^2_{(f_0,Q)}(t_k) = \frac{1}{2\pi} \frac{Q^2}{f_0{}^2} \left| \sum_{m=0}^{N-1} \frac{e^{-(\frac{f_m-f_0}{f_0})^2 Q^2/2} \times H^*(f_m)}{PSD(f_m)} e^{i2\pi f_m t_k} \right|^2 \qquad (5)$$

Computing gains are achieved by doing the correlation in the Fourier domain. To reduce edge effects for data segments of limited size, the Welch overlapping scheme [21] is applied : data is treated in chunks of size N, with an overlap of N/2 (typically $N = 2^{18}$ in this work). Each chunk is windowed by the Hann function [22], and the two outputs of the transform for the same data bin from successive chunks are weighted and summed up to provide the final coefficient. PSD is estimated over several chunks (20 in this work) from the same windowed data.

2. Template bank selection

Any actual search must rely on limited resources, hence the importance of selecting a set of complementary templates that maximizes coupling with signal. This is quantified by the mismatch between two waveforms ϕ and ψ, given by the ambiguity function Γ :

$$\Gamma(\phi, \psi) = max_t < \phi(t), \psi(t) > \qquad (6)$$

assuming $||\phi|| = ||\psi|| = 1$, using the notation of eqn. (1).

For practical purposes, the ambiguity function can be approximated, within the hypothesis of close narrow-band waveforms previously normalized, as :

$$\Gamma(\phi, \psi) = max_t \int_{-\infty}^{\infty} \phi(\tau)\psi(\tau + t) \, d\tau \qquad (7)$$

Within the family of EG templates this gives :

$$\Gamma_{EG}(f_1, Q_1, f_2, Q_2) = \sqrt{2} \sqrt{\frac{f_1/Q_1 \times f_2/Q_2}{(f_1/Q_1)^2 + (f_2/Q_2)^2}} \times e^{-\frac{(f_1-f_2)^2}{2((f_1/Q_1)^2+(f_2/Q_2)^2)}} \qquad (8)$$

As Γ is defined as a scalar product, a tiling of the 2-dimensional parameter space with a near minimum template number can be proposed using, for instance, the method of [23] for a set minimum match (MM) in the form $\Gamma > MM$.
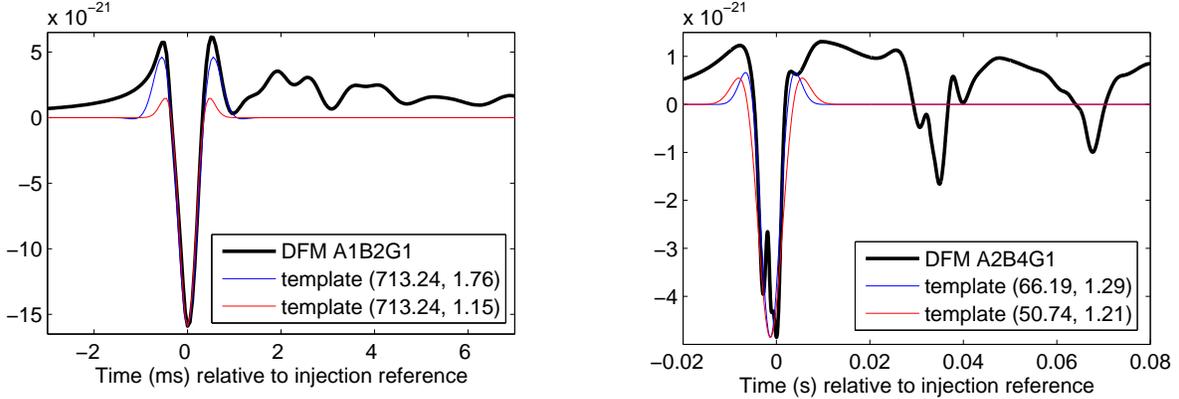
FIG. 2: Adequacy of EGC templates with two test signals from the DFM catalog, A1B2G1 (*Left*) and A2B4G1 (*Right*). The templates are identified by their parameters (f_0, Q) and time-shifted to their extremal amplitude value. Only the Gaussian-Cosine part of the complex templates are shown here. The match between these and signal reaches 9.6 for A1B2G1 and 8 for A2B4G1.

The parameter space is *a priori* bounded by minimum and maximum frequencies and quality factors. From the expected sensitivity of Virgo, the target frequency region would be 10 Hz - 1k Hz. Considering template duration, estimated from the width of the Gaussian envelope, it appears that this area encloses long duration structures at low frequency. To ensure that only burst-like signals would be matched, an additional constraint on maximum template duration was added.

The quality factor range was established as follows. The lower limit $Q > 1$ comes from two observations : the approximate ambiguity function (eqn. (8)) becomes less accurate below this limit and templates do not exhibit additional shapes. The upper bound, $Q \leq 16$, was arbitrarily set. It effectively limits the search to the region of the parameter space where the injected waveforms reside. This was a time-saving choice with no physical motivation.

The tiling minimal match parameter that sets the density of the coverage was fixed at $MM \simeq 0.95$. Of course, in a real search, the MM can be fixed to some higher value, as permitted by computing power.

A final specification comes from the observation that templates with low quality factors, that are not oscillatory, are potentially important for some expected signals. Figure 2 shows how low quality factor templates (here the even quadrature) recover peak structures, much like Gaussian templates would do. So we chose to distinguish two areas in the parameter space illustrated in fig. 3 :
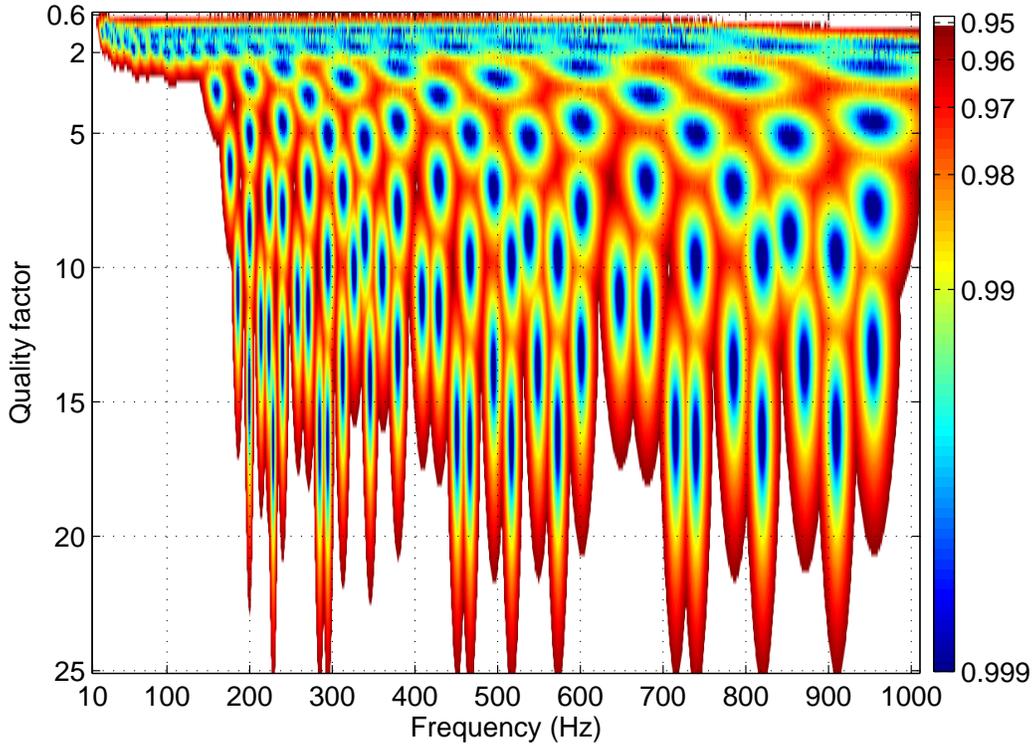
7

FIG. 3: Best match between EG waveforms and the selected 122 templates, with a logarithmic color scale. Values at and below 0.95 were set to the same color (white). Figure produced by smoothing over values at 500000 randomly chosen (f, Q) points.

$Q \geq 2.2$, $MM = 0.95$: Templates are oscillatory. Discarding low frequencies, the low frequency limit was shifted to 150 Hz.

$1 < Q \leq 2.2$: Peak-like shapes, as illustrated in fig. 1, dominate. An increased minimal match ($MM \simeq 0.98$) was introduced to have a regular coverage over the whole frequency band.

This selection was used in the analysis of Virgo commissioning runs, see [20], and will be the reference set hereafter.
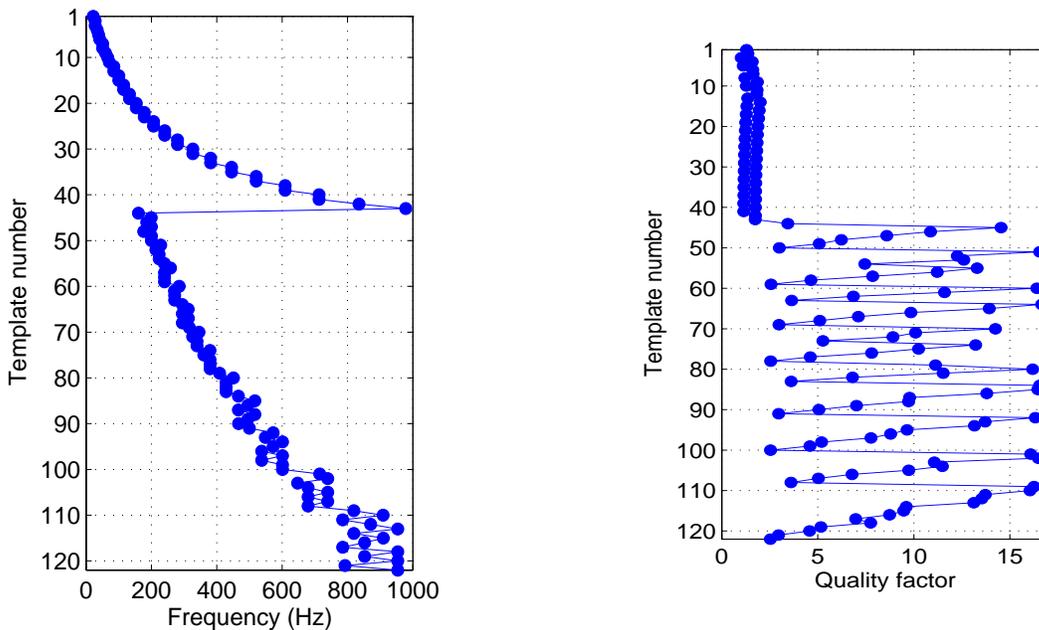
FIG. 4: Stacking order of the 122 selected templates in the output map, for the regular EGC implementation.

C. Event definition and parameters

The elementary output of the transform is a coefficient indexed in time and template. Analysis relies on maps of coefficients, or 'pixels' and requires some basic image processing, namely suppression of individual coefficients below a set threshold and clustering of adjacent coefficients, to extract meaningful events. This assumes that adjacent coefficients refer to related signal components.

A TF-like representation was therefore sought, by adjusting the stacking order of the correlation output, identified by three parameters (frequency, quality factor, time) for the selected templates. A compromise was retained to try and regroup selected templates with similar characteristics : a first group with $1 < Q \leq 2.2$ is ordered by increasing frequency; at higher quality factor, templates are regrouped by 50 Hz wide frequency bands, wherein they are ordered by decreasing Q (see figure 4).

In principle this allows the localization of energy in frequency while separating oscillating events from peak-dominated ones.

Figure 5 illustrates the transform when the ordered template set is used. The peak-like
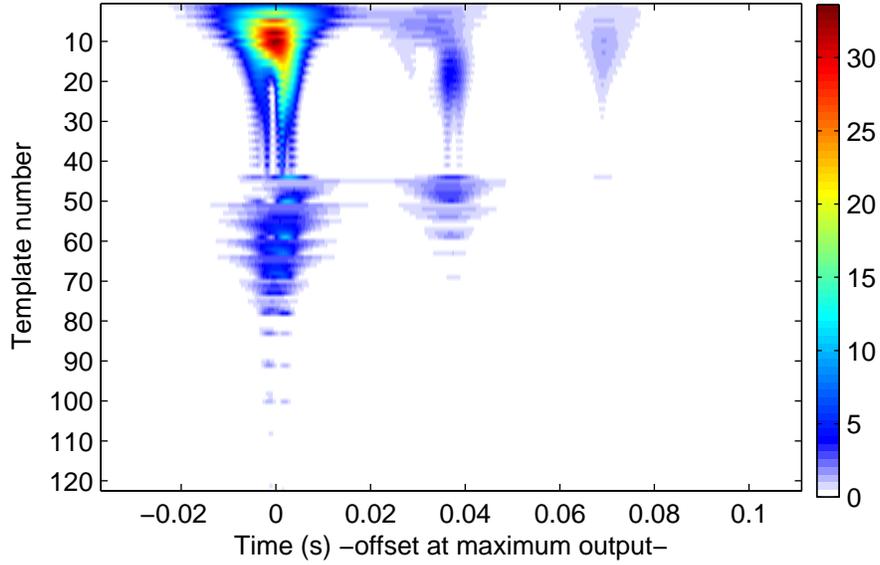
FIG. 5: EGC transform of DFM A2B4G1 waveform, considering a Virgo-type noise background and SNR 10. Templates were selected using a tiling algorithm and stacked as proposed in figure 4. The colorscale has the dimension of a squared SNR.

nature of the tested waveform can be seen clearly : most of the energy is concentrated at low quality factors, according to the numbering of fig. 4. Conversely, an oscillatory signal would be mostly seen on the templates with high ordering numbers.

The clustering of the time-frequency images considers that all neighbouring (the 8 surrounding elements) selected pixels belong to the same event. This reduction is achieved by using the Hoshel-Kopelman algorithm [24] [25], which requires a single sweep to group pixels, by comparing pixels to neighbouring values and numbering clusters.

Several quantities can be extracted from these events, to constrain event coincidences or to describe event populations :

- Total energy : the sum of all coefficients in an event $E_{cluster} = \sum_{i \in cluster} |c_i|^2$.

- Peak energy : the maximum coefficient value over the cluster.

- Event Time : the location of the peak energy.

- Central Frequency : averaged over all cluster elements, weighted by their energy.

10

- Quality factor : averaged, like the frequency.

- Duration : the width of a cluster on the time axis, from all cluster members (total duration) or represented by the σ parameter from the template with maximum energy in the cluster.

The above *total energy* generally does not represent the true energy that can be attributed to a signal, owing both to the incompleteness in frequency coverage and the non-orthogonality between the templates that comes from the envelop term in equation (2) as well as from the possibility for several templates of identical center frequency but different quality factors.

D.   Event selection algorithm

Given the complex TF structure of astrophysical signals (see for example fig. 5), some clusters that belong to the same signal could be split by the threshold on coefficient value. Associating these 'orphan' clusters increases the visibility of the signal in the final event list while decreasing the false alarm rate.
To remain robust with respect to the imprint left by a signal on the TF map, the following clustering method was selected, where three thresholds are applied :

a) Background restriction : A low threshold is applied to map coefficients to discard most of the noise-consistent elements.

b) Clustering : Events are created from all adjacent remaining pixels.

c) Background suppression : A threshold higher than in step a) is applied to coefficients in the clusters. Low value coefficients, which may be noise fluctuations, are removed. Higher amplitude coefficients in the cluster remain grouped as before. This step, in principle, permit a lower threshold in a) to initially keep a larger fraction of the signal.

d) Simultaneity test : Events overlapping in time or separated by less than a reference delay are merged.

e) Energy selection : The candidate list contains only events with a total energy above a final threshold.

Remaining events are ranked according to peak energy (equivalent to maximum template correlation). A threshold on this last value can be freely tuned to build a ROC curve or fix the final false alarm rate, while the preceding steps are set, as discussed below.
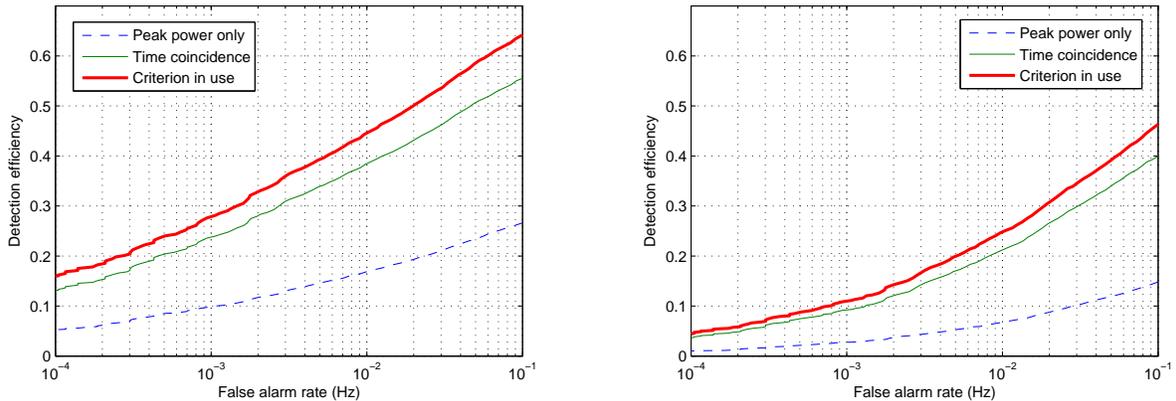

FIG. 6: Comparison of EGC performances for a threshold on peak energy, according to the event selection rule, for two waveforms, DFM A1B2G1 to the left, Gaussian Cosine 820 Hz Q 15 to the right. These results come from some 6000 injections spread over 24 hours of simulated Virgo-type data, with an injected SNR of 5.

After applying the minimum pixel energy and neighbour clustering steps, the first gain comes from the introduction of time coincidence between clusters. This allows to combine simultaneous time-frequency structures, thus reducing the false alarm rate without affecting detection efficiency, assuming that the structures within the coincidence time window belong to the same event. This evolution is shown in fig. 6. Further improvement is obtained after the total energy threshold. Adjustments cannot be made independently of the other steps in the algorithm. For instance, raising the pixel energy threshold from the selected value lowers the number of events clearing the total energy threshold; while decreasing it does not alter the performances much. This is explained by the correction brought by the time coincidence, here with a delay allowance of 0.05 seconds. The adjustment of thresholds was done mostly independently for each stage and may therefore not constitute an optimum. However, given the variety of waveforms that are targeted, this was considered sufficient.

E. Other information

1. Computational cost

The full pipeline, including the event selection steps of section II D, was run without optimization (notably no parallelization) on Opteron computers. In single processor mode, the average ratio of computing time to data time ($\simeq$30 minutes per job) was about 7.

2. A simple comparison to the Q transform

The development of this algorithm took place in parallel to that on the Q transform without influence in either direction. This precision is required given the similarity of the two filters. Let us mention briefly the steps where the methods differ. We will not compare parameters values, as these could be adjusted for a given instrument or target signal.
In the Q transform, rhe templates are placed at fixed Q values, then frequencies are selected to satisfy a minimal match threshold. There is also a choice of times where the correlation should be estimated, also against a minimal match threshold. Finally, the maps of coefficients in ($time$, $frequency$) are stacked and the events built by comparing values at different Q. Besides, the Q transform relies on an external whitening procedure instead of PSD normalization.

There has been no full comparison of the two methods. The template selection of the Q transform may be less sparse in the (f, Q) plane, while some loss in timing accuracy could be expected compared to EGC. Computational cost reduction was an argument for these choices.

III. MONTE CARLO PERFORMANCE STUDIES

Performance tests were carried out using typical testbed waveforms and Gaussian colored noise. The efficiency of the method is characterized by Receiver Operating Characteristics (ROC) curves, in which detection efficiency is plotted against false-alarm (FA) rate.

A. Data

Data from operating instruments do not follow a Gaussian distribution but, as the minimum hypothesis model, it is a necessary first step in the evaluation of analysis methods. The Gaussian noise series used here is a 24 hours stretch, whose power spectral density is the target Virgo sensitivity, including for large spectral lines due to the suspension system or mirror internal resonance modes that are distinctive features of Virgo.

Eight waveforms were used, covering different categories. There are six regular waveforms: two Gaussian signals, of characteristic duration (σ_t) of 1 and 4 milliseconds, and four Gaussian Cosine, of central frequency (f_0) 235 and 820 Hz, and quality factor ($Q = 2\pi f_0\sigma_t$) 5 and 15. The last two signals are DFM A1B2G1 and DFM A2B4G1 from the catalog [5] selected to present very different characteristics.

Amplitudes for the injections are normalized in units of SNR, 5 and 10 here [26]. Each waveform was injected 6000 times [27].

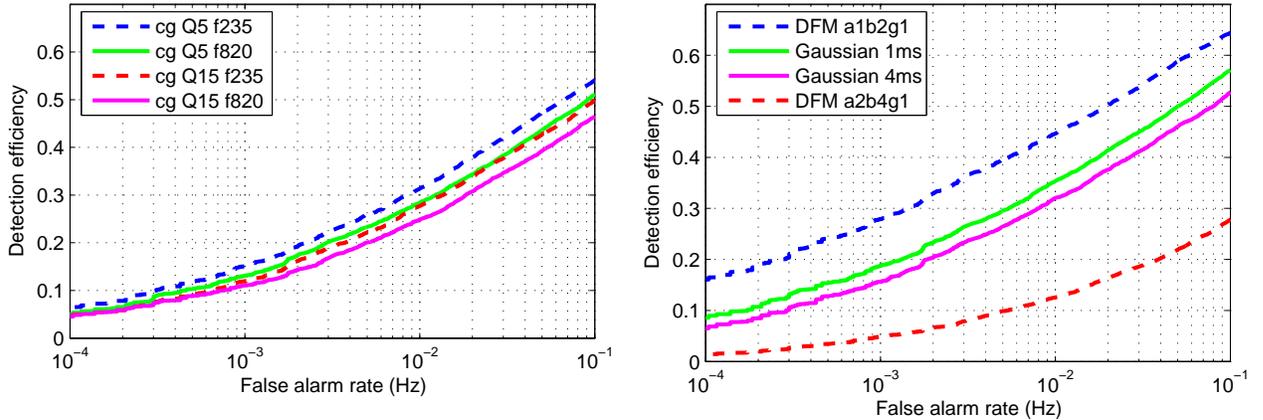FIG. 7: EGC ROC curves for Gaussian Cosine (*left*) and Gaussian plus DFM (*right*) waveforms, injected with SNR 5.

B. Detection performances

For injections at SNR 10, detection efficiency reaches 100% at a false alarm rate of 10^{-4} Hz, except for DFM A2B4G1 (97%). We discuss hereafter results for SNR 5 injections.

The optimal behaviour of the filter is obtained, unsurprisingly, with the Gaussian Cosine

signals, as illustrated by the similarity of the ROC curves in fig. 7. The same similarity exists for the Gaussian waveforms, which implies that for low Q values the Cosine-Gaussian is a sufficient approximation of the Gaussian shape.

The two DFM waveforms are very different from the templates and from each other. This difference is observed in the output of the EGC filter, one waveform (A1B2G1) being very well extracted, the other (A2B4G1) much less. The highest detection efficiency is actually achieved for A1B2G1. The particular sharpness of the waveform, the choice of templates and the event selection steps likely contribute together to this enhancement.

C. Events characteristics analysis

Considering the distribution of event parameters (introduced in section II C) helps to understand the behaviour of the filter. For the Central Frequency, Quality and Duration, average values and standard deviations are given in tab. I, along with expectations when meaningful for the waveforms.

1. Event frequency

The Gaussian Cosine provide a check on the possibility for frequency estimation. The small bias in the average estimate for the waveforms with quality factor 5, still compatible with the true value given the variance, disappears for $Q = 15$. It is introduced by the choice of templates and the natural limit on frequency resolution for shorter signals.

For A1B2G1, the large discrepancy between the frequency of the peak template in fig. 2 (713 Hz) and the event composite frequency in tab. I (593 Hz) suggests these estimations are not adequate. This is reasonable for signals with a large frequency bandwidth, like most waveforms from DFM. Still, A1B2G1 differs from the other selected waveforms by the narrowness of its single peak feature, that can only be properly matched by the few high frequency, low quality factor templates available, as illustrated in fig. 2. The performances for A1B2G1 cannot be directly compared to the available Gaussian waveforms, as its main peak is more than twice as narrow.

Gaussian Cos.	SNR 10		SNR 5	
Parameter (exp.)	Av.	Var.	Av.	Var.
F = 235 (Hz)	237.6	7.8	237.0	27.3
Q = 5	5.1	1.2	5.1	2.4
σ = 3.4 (ms)	3.4	0.8	3.4	1.7
F = 820 (Hz)	792.0	22.2	807.8	49.6
Q = 5	5.3	1.0	5.4	2.2
σ = 1.0 (ms)	1.0	0.2	1.0	0.4
F = 235 (Hz)	232.5	3.3	234.0	14.1
Q = 15	14.1	1.7	13.0	3.5
σ = 10.2 (ms)	9.5	1.3	8.8	2.4
F = 820 (Hz)	824.9	9.4	821.1	20.3
Q = 15	15.3	2.3	13.7	3.5
σ = 2.9 (ms)	3.0	0.4	2.7	0.7

Waveforms	Parameters	SNR 10		SNR 5	
		Av.	Var.	Av.	Var.
	F (Hz)	507.0	79.4	592.6	110.2
DFM	Q	1.1	0.05	1.2	0.5
A1B2G1	σ (ms)	0.3	0.01	0.3	0.1
	F (Hz)	141.6	19.0	144.7	43.3
Gaussian	Q	1.3	0.02	1.3	0.2
	σ = 1 (ms)	1.6	0.3	1.6	0.6
	F (Hz)	40.6	3.2	37.1	6.2
Gaussian	Q	1.0	0.05	1.1	0.2
	σ = 4 (ms)	5.7	0.4	5.7	1.0
	F (Hz)	73.8	14.6	76.4	45.8
DFM	Q	1.3	0.1	1.3	0.3
A2B4G1	σ (ms)	3.4	0.5	3.4	1.0

TABLE I: Information (average value and standard deviation) for events associated to signal injections : Central Frequency, Quality factor and Duration. [28] For Gaussian and DFM waveforms, expectations are not given when a satisfying definition is not available.

2. Estimating event duration

Total event duration, whose distribution is shown in fig. 9 and characterized in tab. II, is consistently above the Gaussian width estimate given in table I, most clearly for DFM A1B2G1, with a factor 20 between the parameters at SNR 15, and 5.5 at SNR 5.

As expected, event duration increases with SNR. While at SNR 5 the ratio between Gaussian width and total duration varies between 2 and 3 for the other waveforms, they range between 3 and 8 at SNR 10. This comes from the shape of the templates : for a Gaussian-like signal, each extremum of the template will give an increased match, thus extending the time interval where the signal is actually detected. Besides, for the DFMs, total event duration is still below the real waveform duration, of the order of 0.01 and 0.1 seconds for A1B2G1 and A2B4G1. As for false alarm events, for the 10^{-2} Hz rate, 90 % of
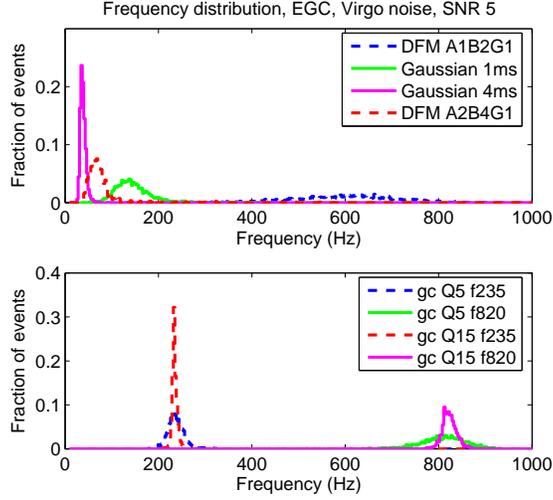
FIG. 8: Distributions of estimated frequency over events associated to signal injections (SNR 5).
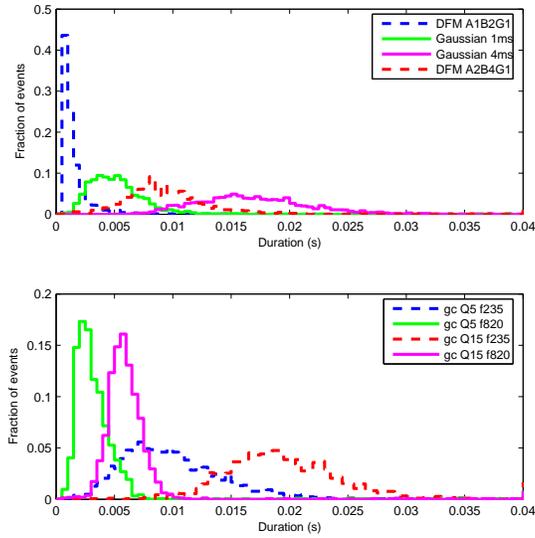

FIG. 9: Distributions of total event duration over events associated to signal injections (SNR 5).

them last less than 5.5 ms and 65 % less than 1 ms. The total event duration thus appears to contain little information on realistic events.

As no event lasted longer than 40 ms, the coincidence window (50 ms) does not affect the distributions.

Quality factor values (table I), when a reference is available, are accurate. For Gaussian Cosine Q15, the under-estimation of Q at lower SNR is reasonable considering that the tails

Gaussian Cos.	SNR 10		SNR 5		Waveforms	SNR 10		SNR 5	
dt (ms)	Av.	Var.	Av.	Var.	dt (ms)	Av.	Var.	Av.	Var.
F 235 Q 5	22.68	5.85	10.10	5.84	A1B2G1	6.03	5.04	1.65	3.90
F820 Q 5	7.28	4.55	3.60	5.15	G. 1 ms	12.92	5.14	5.45	4.94
F 235 Q 15	34.28	5.91	19.57	6.81	G. 4 ms	31.59	6.04	16.96	5.95
F 820 Q15	10.32	4.38	6.19	4.70	A2B4G1	18.67	8.66	9.39	5.29

TABLE II: EGC event duration (average value and standard deviation) for events associated to signal injections. [28]

of the waveform then disappear into the noise background.

3. Timing accuracy

Exploitation of GW candidates will strongly rely on timing for coincidence studies with other instruments. In a network of gravitational wave detectors, delays in arrival time translate in constraints on source direction. This principle is applied easily when possible differences of the GW measured by separate instruments, from polarization effects, are neglected. Coherent analysis combining the data from all detectors, as discussed for instance in [29], are an important alternative when polarization is considered.

Gaussian Cos.	SNR 10		SNR 5		Waveforms	SNR 10		SNR 5	
δt (ms)	Av.	Var.	Av.	Var.	δt (ms)	Av.	Var.	Av.	Var.
F 235 Q 5	0.019	0.55	0.009	1.16	A1B2G1	-0.001	0.03	-0.004	0.08
F820 Q 5	0.019	0.16	0.018	0.33	G. 1 ms	0.020	0.15	0.01	0.42
F 235 Q 15	0.004	1.64	-0.020	2.93	G. 4 ms	0.019	0.47	0.03	1.67
F 820 Q15	0.024	0.45	0.030	0.88	A2B4G1	-1.320	0.51	-1.29	1.20

TABLE III: Arrival time error (average value and standard deviation) for events associated to signal injections. [28]

For all waveforms, arrival time is defined here from the location of the maximum absolute amplitude. The distributions of arrival time error, shown in figure 10 and summarized in
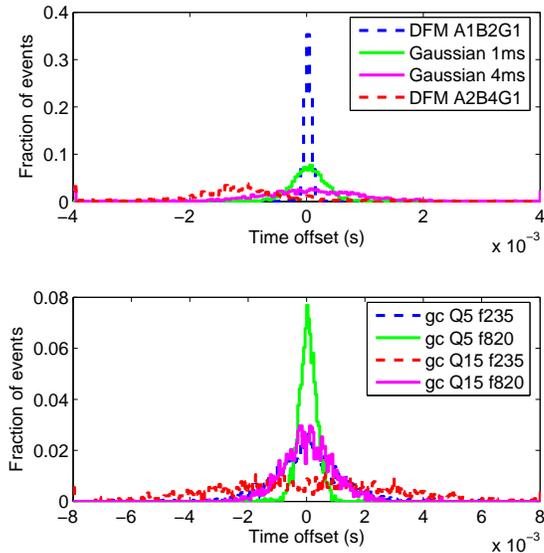
FIG. 10: Distributions of estimated arrival time error on recovered injections (SNR 5).

table III, have an average value of about a tenth of a millisecond and a standard deviation at least 20 times larger. This is to be compared to the sampling rate of the data (20 kHz or 0.05 ms) directly used in EGC. There is therefore no bias in the estimate, except for DFM A2B4G1. The delay for this waveform comes from a small asymmetry at the maximum of the main peak, visible in fig. 2, difficult to recover with symmetric templates.

The standard deviation is below the millisecond at SNR 10, except for Gaussian Cosine 235 Q 15, the longest duration waveform, and slightly above 1 ms at SNR 5 for the longer waveforms, the Gaussian Cosine at 235 Hz, Gaussian 4 ms and DFM A2B4G1. The spread for DFM A1B2G1 is the smallest, just as its main peak is the narrowest (less than a millisecond) of the testbed.

Shorter ($Q = 5$ compared to $Q = 15$) Gaussian-cosine are located in time with sharper distributions (fig. 10), but less so in frequency, as expected from the uncertainty relation for time-frequency transforms.

To translate these values into direction estimations, we refer to recent work [30] on source direction estimation from arrival times in a network of interferometer. It shows that these performances are not sufficient to guarantee an angular resolution of 1° over the sky (standard deviations below 0.1 ms would be required).

IV. CONCLUSION

Starting from the matched filtering technique, we developed the EGC algorithm for the extraction of impulsive gravitational wave signals. We have estimated its performances from simulations on Gaussian noise in terms of efficiency and timing accuracy. Elements of comparison of this algorithm with other burst detection methods can be found in [32] and [33]. These justify its use along with other filters.

Our simulations show that even for complex waveforms, arrival times are properly recovered. Frequency or, where applicable, quality factor estimates are consistent with injected parameters. Such characterization will hopefully contribute to the study of burst-like structures extracted from real data.

The template selection has a definitive influence on performances, so increasing the tiling parameter MM is an option, although it directly affects the computing cost. While the retained list gives satisfying results over the tested waveforms, dedicated searches focusing on some categories of signals, like the oscillating waveforms from [7] and [8], could use different template selection and thresholding rule, in exchange for reduced robustness to waveform type. However, the main goal of the event extraction process is to allow simple structures found in a signal to be grouped so that the whole waveform is detected. When successful, this allows EGC to detect any short waveforms where peaks, oscillations (within the range of the template bank) or a combination are prominent.

In the context of Virgo data analysis, the robustness offered by the matched filtering approach, useful to preserve most of the detection efficiency on real data in spite of difficult noise background is completed with the information extraction capabilities expected from time-frequency transforms. The EGC analysis method has has contributed the commissioning of Virgo, by helping to identify instrumental noises, thus highlighting the interest of additional event descriptors, notably frequency, for physical interpretations. It is currently applied, in addition to several other methods, on the data collected during the first Virgo Scientific Run.

Acknowledgements

We wish to thank the Virgo Collaboration for providing with the SIESTA engine the 24 hours of simulated data used in this work for Monte Carlo studies, and the LIGO-Virgo joint analysis group, whose purpose and initial findings are discussed in [32], for a number of useful discussions. We are grateful for the comments by the anonymous referees.

[1] http://www.ligo.caltech.edu

[2] http://www.virgo.infn.it

[3] M. Punturo, Virgo Note VIR-NOT-PER-1390-51 (2004).

[4] C.L. Frier and K.C.B. New in *Living Rev. Relativity*, **6**, (2003) http://www.livingreviews.org/lrr-2003-2

[5] H. Dimmelmeier, J.A. Font & E. Müller, A.A.**393**, 523 (2002)

[6] C. D. Ott *et al* , accepted in *Phys. Rev. Lett.***98**, 261101 (2007)

[7] M. Shibata & Y. Sekiguchi, *Phys. Rev.* D**71**, 024014 (2005)

[8] A. Burrows *et al* , *Astrophys. Jour.***640**, 878 (2006)

[9] E. Cappellaro *et al* , A.A.**351**, 459 (1999)

[10] P. Flandrin, "Time-frequency / time-scale analysis" (Academic Press, 1999)

[11] W.G. Anderson *et al* , *Phys. Rev.* D**63**, 042003 (2001)

[12] S. Klimenko and G. Mitselmakher, *Class. Quantum Grav.***21**, S1819 (2004)

[13] G.M. Guidi *et al* , *Class. Quantum Grav.***21**, 5 (S815)2004

[14] S. Chatterji *et al* , *Class. Quantum Grav.***21**, S1809-S1818 (2004)

[15] C.W. Therrien, "Discrete random signals and statistical signal processing" (Prentice Hall, 1992).

[16] L.A. Wainstein and V.D. Zubakov, "Extraction of signal from noise" (Prantice-Hall, 1962).

[17] Y. Pan *et al* , *Phys. Rev.* D**69**, 10 (2004)

[18] R.C. Pinnegar and L. Mansinha, Fluctuation and Noise Letters**3-3**, 357 (2003)

[19] For a flat PSD and a normalized Gaussian time series, choosing $Q = 1/2\pi$ (or $\sigma = 1/f_0$) gives the original S Transform.

[20] F. Acernese *et al* , *Class. Quantum Grav.***S197**, 23 (2006)

[21] P.D. Welch, IEEE Transactions on Audio and Electroacoustics**15**, 70 (1967)

[22] $w(i) = \frac{1}{2}(1 - \cos(\frac{2\pi i}{N-1}))$, $i \in [0, N-1]$

[23] N. Arnaud *et al* , *Phys. Rev.* D**67**, 102003 (2003)

[24] J. Hoshel and R. Kopelman, *Phys. Rev.* B**14**, 3438 (1976)

[25] The original algorithm was modified to avoid storage of the output for all templates and completed by a border-bridging step, to merge clusters artificially split over several maps due to data segmentation.

[26] For the PSD, a frequency resolution of $8\ 10^{-2}$ Hz was used and it was averaged on 20 segments.

[27] While this study aims at quantifying detection performance for a single event, the analyzed time series contains several, with an injection rate of 1/14 s. Tests showed however that at the selected SNR values, the multiplicity of events in the time series did not affect even a matched filter.

[28] Estimations based on 6000 injected events, over 24 hours of data, with a false alarm rate of 10^{-2} Hz.

[29] N. Arnaud *et al* , *Phys. Rev.* D**68**, 102001 (2003)

[30] F. Cavalier, *Phys. Rev.* D**74**, 082004 (2006).

[31] N. Arnaud *et al* , *Phys. Rev.* D**67**, 062004 (2003)

[32] F. Beauville *et al* , J. Phys.**S32**, 212 (2006)

[33] F. Beauville *et al* , submitted to *Class. Quantum Grav.*