Best chirplet chain: Near-optimal detection of gravitational wave chirps

Éric Chassande-Mottin[1,*] and Archana Pai[2,†]

[1]*CNRS, Observatoire de la Côte d'Azur, ARTEMIS, BP 4229 06304 Nice Cedex 4 FRANCE*
[2]*INFN, Sezione Roma 1–P.le Aldo Moro, 2 00185 Roma ITALIA*
(Received 22 December 2005; published 27 February 2006)

The list of putative sources of gravitational waves possibly detected by the ongoing worldwide network of large scale interferometers has been continuously growing in the last years. For some of them, the detection is made difficult by the lack of a complete information about the expected signal. We concentrate on the case where the expected gravitational wave (GW) is a quasiperiodic frequency modulated signal i.e., a chirp. In this article, we address the question of detecting an *a priori* unknown GW chirp. We introduce a general chirp model and claim that it includes all physically realistic GW chirps. We produce a finite grid of template waveforms which samples the resulting set of possible chirps. If we follow the classical approach (used for the detection of inspiralling binary chirps, for instance), we would build a bank of quadrature matched filters comparing the data to each of the templates of this grid. The detection would then be achieved by thresholding the output, the maximum giving the individual which best fits the data. In the present case, this exhaustive search is not tractable because of the very large number of templates in the grid. We show that the exhaustive search can be reformulated (using approximations) as a pattern search in the time-frequency plane. This motivates an approximate but feasible alternative solution which is clearly linked to the optimal one. The time-frequency representation and pattern search algorithm are fully determined by the reformulation. This contrasts with the other time-frequency based methods presented in the literature for the same problem, where these choices are justified by "*ad hoc*" arguments. In particular, the time-frequency representation has to be *unitary*. Finally, we assess the performance, robustness and computational cost of the proposed method with several benchmarks using simulated data.

I. INTRODUCTION

The worldwide network [1] of large scale interferometric gravitational wave (GW) detectors have started to take data. The network includes the detectors GEO600, LIGO and TAMA. It will be completed soon by the upcoming Virgo. The overall sensitivity of these detectors is continuously improving. Interesting upper limits for the amplitude of GWs are being set and the first detection is hopefully not too far.

A large variety of astrophysical sources are expected to emit GWs in the observational frequency bandwidth of these detectors. From the data analysis viewpoint, the detection methodology for these sources depends on the availability of a reliable and complete model of the GW.

Generally speaking, the oscillations of the GWs are related to the orbital, rotational bulk motion of the constituents of the emitting system. Since the system loses energy by radiation, or because of some other physical process involved, its orbital period, and consequently the GW frequency can vary with time. In such case, the emitted GW is a frequency modulated signal i.e., a *chirp*. A detailed knowledge of the dynamics of the system is required to describe precisely the characteristics of the GW

chirps, in particular the phase evolution. This may not be always possible as described in the following examples.

The GW emitted by a coalescing binary of compact objects can be divided into three phases (inspiral, merger and ringing). Although the GW can be obtained accurately in the inspiral phase when the bodies are well separated [2] (using post-Newtonian expansions) and in the ringing phase after they have merged [3] (using perturbative methods), the in-between merger phase still defeats both the numerical and analytical efforts [4] for modeling its highly nonlinear regime. For large mass binaries, the merger phase contributes to a dominant fraction of the signal-to-noise ratio (SNR) [5]. In this case, the search method has to accommodate the significant lack of signal information.

Kerr black holes accreting matter from a surrounding magnetized torus are putative sources of the long gamma-ray bursts (GRBs) [6]. It is claimed that, the black hole spin energy is radiated away through GWs along with the GRB. The precise shape of the emitted waveform would need accurate hydrodynamical numerical simulations.

A third example is the GW emitted in the form of the quasinormal modes [7] by a newly born hot neutron star (during the cooling phase which follows the core collapse). Here, the characteristics of the GW depend on the equation of state of the proto-neutron star and various physical processes (like neutrino diffusion, thermalization and cooling) which are currently not known with accuracy.

All these three examples are expected to emit GW as an *unmodeled chirp*, the phase information being not (per-

*Electronic address: ecm@obs-nice.fr
†Electronic address: Archana.Pai@aei.mpg.de
Present affiliation: Max Planck Institut für Gravitationsphysik (AEI), Am Mühlenberg 1, D-14476 Potsdam, Germany

fectly) known. Its typical duration in the detector bandwidth is of the order of a few seconds.

While matched filtering is a well-known and efficient detection technique when a precise waveform model is available, the lack of waveform information prevents us from using the same approach. It is thus natural to advocate for *exploratory searches* (based on partial information or "good sense" models) as opposed to *targeted* ones (relying on a precise model).

Various strategies [8–14] have been designed following this viewpoint, for the detection of transients of short duration (tenth to hundredth of milliseconds) or GW *burst*. Such transients are typically from supernovae core collapses. The notion of a varying frequency is not adequate for such a small number of cycles. It is thus not meaningful to describe such transients as chirps. Their detection is a different issue than the one considered here.

Here, we are interested in exploratory search specifically for unmodeled chirps. In the past, this question has already been investigated yielding a detection method, the Signal Track Search (STS) [13]. The STS relies on the observation that, in a time-frequency (TF) representation, a chirp appears as a filiform pattern and this discriminatory signature can be searched for. A satisfactory implementation of this phenomenological argument calls for a proper TF representation (TFR) and pattern search algorithm. The STS results from "*ad hoc*" choices for the above mentioned points.

In this paper, we propose a new method for the detection of unmodeled chirps. It is based on the same general principles (pattern search in a TFR) as the STS. Its originality resides in the clear link we establish between the method (i.e., the choices of TFR and pattern search) and an optimality criterion.

The paper is organized as follows. We state the detection problem in Sec. II. We introduce the general chirp model referred to as *smooth chirp* and we assume that most physically realistic GW chirps can be described by this model. The phase of a smooth chirp is an arbitrary continuous and differentiable function with bounded first and second derivatives. In Sec. III, we derive the optimal statistic for detecting a given smooth chirp in noise, which is usually referred to as *quadrature matched filtering*. The idea is then to apply this statistic for any smooth chirp, and select the maximum which is associated to the individual that best fits the data. This maximization has to be done numerically. To do so, the set of smooth chirps being a continuous set, has to be discretized. In Sec. IV, we show that grids of templates can be constructed for smooth chirps using chains of small chirps, we call *chirplet chains* (CC). We further prove that the grid is *tight* i.e., any smooth chirp can be closely approximated by a chirplet chain. The maximization of the statistic over the set of smooth chirp can be reliably replaced by a maximization over the set of CCs. However, the number of CCs being very large, the

computation of the quadrature matched filter for all CCs is not tractable. In Sec. V, we propose a feasible (TF based) procedure for finding the best CC. We show that the quadrature matched filter can be reformulated approximately as a path integral computed in the TF representation given by the discrete Wigner-Ville (WV) distribution. As a result, the maximization of the statistic over the CCs amounts to obtaining the TF path of largest integral. We demonstrate that this kind of problem can be solved efficiently with dynamic programming. We detail our path search algorithm and we evaluate its computational cost. Finally, we compare the resulting algorithm with other methods in Sec. VI. Receiver operating characteristics obtained in several realistic situations demonstrate the superiority of the proposed approach.

II. SMOOTH CHIRPS IN GAUSSIAN NOISE

We introduce a general chirp model which we refer to as *smooth chirp*,

$$s(t) = A\cos(\phi(t) + \varphi_0) \quad \text{for } t_0 \le t \le t_0 + T, \quad (1)$$

and $s(t) = 0$ outside this interval.

A smooth chirp is characterized by the amplitude A, the initial phase φ_0 and a *smooth* phase evolution $\phi(t)$ (without loss of generality, we assume $\phi(t) = 0$ at the arrival time $t = t_0$). We define the term *smooth* as follows. A phase $\phi(t)$ is smooth if this function and its first three derivatives are continuous and we have

$$\left|\frac{df}{dt}\right| \le \dot{F} \qquad \left|\frac{d^2f}{dt^2}\right| \le \ddot{F}, \qquad (2)$$

for all t and where $f(t) \equiv (2\pi)^{-1}d\phi/dt$ is the instantaneous frequency. The chirping rate limits $\dot{F}$ and $\ddot{F}$ are chosen based on the allowed upper bounds obtained from general astrophysical arguments on the GW source of interest. The chirp model thus includes four parameters $\mathbf{p} \equiv \{A, \varphi_0, t_0, \phi(\cdot)\}$ which are not known *a priori* and need to be determined from the data.

Let the signal be correctly sampled at the Nyquist rate $f_s \equiv 1/t_s$ and let us assume that we acquire the data x_k by blocks of N samples. The GW signal is denoted by $s_k \equiv s(kt_s)$ for $k = 0, \cdots, N - 1$ with the duration $T = t_sN$. The noise n_k is assumed to be additive white and Gaussian with zero mean and unit variance. Since the noise of GW detectors is colored, this noise model implies that a whitening procedure has been already applied to the data. (Therefore, the signal s_k in Eq. (3) is a "whitened" version of the actual GW signal).

In this initial work, we restrict the smooth chirp model to have a constant envelope, although GW chirps are generally amplitude modulated. The constant envelope thus limits the descriptive power of the model. However, we

argue that the model is still reasonable for many cases[1] and that the phase information plays a major role for detection of chirps. We leave the problem of detecting amplitude modulated chirps for subsequent work.

III. OPTIMAL DETECTION OF A SMOOTH CHIRP

For each block of N data samples, the signal detection problem is to decide which statistical hypothesis suits best to the data among the following two:

$$(H_0) \quad x_k = n_k \qquad \text{noise only} \qquad (3a)$$

$$(H_1) \quad x_k = s_k + n_k \qquad \text{signal + noise} \qquad (3b)$$

In practice, this requires thresholding a functional of the data, commonly referred to as *statistic*. If the statistic crosses the threshold, H_1 is chosen as opposed to H_0 and *vice versa*.

Because of the presence of random noise, this decision may not be always the right one. There are two types of errors associated to this: false alarms (decide H_1 while H_0 is present) and false dismissals (the opposite). The probabilities of occurrence of these two errors fully quantify the performance of a given statistic. This information can be used to rank the large number of possible statistics and to identify the best one. This is the approach followed by the Neyman-Pearson (NP) criterion [15]: the NP-optimal statistic minimizes one error probability, while keeping the other fixed to a given value. To be precise, in the present case, it minimizes the false dismissal probability for a fixed false alarm probability.

It can be shown that the likelihood ratio (LR) defined by $\lambda \equiv \mathbb{P}(\{x_k\}|H_1)/\mathbb{P}(\{x_k\}|H_0)$ is NP-optimal [15]. For the detection problem described in Eq. (3), the LR can be easily obtained if we assume that the chirp parameters **p** are known in advance. When the parameters are not known *a priori* (which is the situation here), the ideal would be to have a statistic which is NP-optimal for all values of the parameters. This statistic is usually referred to as *uniformly most powerful*. However, it is not guaranteed that such statistic always exists, and even if it does, it is generally difficult to obtain.

A sensible solution consists in getting some kind of estimates for the unknown parameters and then use the LR assuming that the estimated value is the actual value. If we use *maximum likelihood* (ML) estimators of the unknown parameters, the resulting statistic is referred to as *generalized likelihood ratio test* (GLRT) [15] (or maximum likelihood test in the statistical community).

---

[1]It is important to stress here that the model applies to the whitened chirp. For inspiralling binary chirps crossing the entire detector bandwidth, the envelope of whitened chirp is flatter than the original GW signal. For the other cases, this fact depends on the location of the chirp within the detector band.

The GLRT can be shown to be uniformly most powerful in certain cases [15]. For our problem, up to our knowledge, this is an open question. Strictly speaking, it is thus not correct to qualify the GLRT as "optimal" (as is often done in the literature on GW data analysis). Nevertheless, we continue this misuse of language since the GLRT has proven to perform reasonably well and no better alternative appears to be available.

In the following subsections, we give the derivation of the GLRT statistic. We proceed with the maximization of likelihood ratio with respect to the parameters. Following [16], we note that out of the four parameters, A, φ_0 and t_0 are *extrinsic* parameters (known as kinematical or dynamical parameters) whereas $\phi(\cdot)$ is an *intrinsic* parameter (which determines the shape of the chirp waveform). On the basis of this distinction, the maximization over the extrinsic parameters can be treated in a simple manner whereas the computation of the ML estimate of the intrinsic parameter requires a more sophisticated numerical treatment.

A. Maximize the likelihood ratio: A and φ_0

In this subsection, we maximize the LR with respect to A and φ_0. In case of Gaussian noise, it is more convenient to use the log-likelihood ratio (LLR) which is expressed by

$$\Lambda(x; \mathbf{p}) \equiv \ln\lambda = \sum_{k=0}^{N-1} x_k s_k - \frac{1}{2}\sum_{k=0}^{N-1} s_k^2. \qquad (4)$$

We introduce $\bar{s}_k \equiv \cos(\phi_k + \varphi_0)$ (such that $s_k = A\bar{s}_k$) with the norm $\mathcal{N} \equiv \sum_{k=0}^{N-1} \bar{s}_k^2$.

The maximization of the LLR $\Lambda(x; \mathbf{p})$ over A is straightforward and gives the expression of the ML estimate of the amplitude, namely $\hat{A} = \sum_{k=0}^{N-1} x_k \bar{s}_k / \mathcal{N}$. Inserting this expression into Eq. (4), we obtain

$$\Lambda(x; \{\hat{A}, \varphi_0, t_0, \phi(\cdot)\}) = \frac{1}{2\mathcal{N}}\left(\sum_{k=0}^{N-1} x_k \bar{s}_k\right)^2. \qquad (5)$$

The analytical maximization of the LLR over φ_0 deserves a little more attention. The same calculation has been performed for the detection of chirps from inspiralling binaries [17,18] but it is based on the assumption that $\mathcal{N}$ is independent of φ_0 which is not valid in the context of arbitrary chirps. In Appendix A, we detail this calculation and discuss the validity of this assumption.

We express the resulting statistic $\ell(x; t_0, \phi) \equiv \Lambda(x; \{\hat{A}, \hat{\varphi}_0, t_0, \phi(\cdot)\})$ using the following notations for the cross-correlation of the data with the two quadrature waveforms,

$$x_c \equiv \sum_{k=0}^{N-1} x_k \cos\phi_k \qquad x_s \equiv \sum_{k=0}^{N-1} x_k \sin\phi_k, \qquad (6)$$

and for the norms and cross-products of $\cos\phi_k$ and $\sin\phi_k$,

$$n_c \equiv \sum_{k=0}^{N-1} \cos^2 \phi_k \qquad n_s \equiv \sum_{k=0}^{N-1} \sin^2 \phi_k, \qquad (7a)$$

$$n_x \equiv \sum_{k=0}^{N-1} \cos \phi_k \sin \phi_k. \qquad (7b)$$

We distinguish two cases. In the degenerate case where the two quadrature waveforms are linearly dependent (ϕ_k is a constant), $\mathcal{O} \equiv n_c n_s - n_x^2$ vanishes and we have

$$\ell(x; t_0, \phi) = (x_c^2 + x_s^2)/(2N). \qquad (8)$$

Otherwise $\mathcal{O} > 0$, the optimal statistic is

$$\ell(x; t_0, \phi) = (n_s x_c^2 - 2n_x x_c x_s + n_c x_s^2)/(2\mathcal{O}), \qquad (9)$$

and is commonly referred to as *quadrature matched filtering*, (see Appendix A).

B. Maximize the likelihood ratio: ϕ and t_0

The statistic ℓ results from a quadratic combining of the cross-correlations defined in Eq. (6). It can be seen as a "generalized dot-product" and can be related to a "distance" measuring the discrepancy between the data and *template* waveforms (or, in short, templates) defined by the phase ϕ [see Eq. (A8)]. Maximizing ℓ over ϕ is equivalent to minimizing this distance.

The expression in Eq. (9) is for a given known phase ϕ. If the phase is unknown but belongs to the set of smooth chirps, then we need to minimize the distance within this feasible set. In other words, we need to find that smooth chirp which best fits the data i.e., find

$$\ell_{\max}(x; t_0) = \max_{\text{all smooth chirps}} \{\ell(x; t_0, \phi)\}. \qquad (10)$$

This maximization is difficult to tackle analytically and has to be done numerically. The set of smooth chirps is a continuous set and hence not easy to manipulate numerically without discretizing it. For this purpose, we introduce *chirplet chains*, which we discuss in the next section.

As described earlier, we process the data stream blockwise. We compute the statistic independently for each block. The maximization over t_0 is obtained by comparing $\ell_{\max}$ for neighboring blocks and selecting the maximum. The ML estimate of t_0 is then given by the starting time of the corresponding block. The period separating two successive starting times thus defines the resolution of the estimate. If required, this resolution can be improved by increasing the overlap between two neighboring blocks.

We now concentrate on the maximization of $\ell(x; t_0, \phi)$ over ϕ in a given block. In the following, we remove t_0 from the arguments of ℓ to keep the notations simple.

IV. CHIRPLET CHAINS: A TIGHT TEMPLATE GRID FOR SMOOTH CHIRPS

In this section, we show that chirplet chains (CCs) can be used to construct template grids for smooth chirps. CCs are based on the simple geometrical observation: broken lines give good approximations of smooth curves. CCs are signals whose (instantaneous) frequency is a broken line. We verify that they are good approximation of the frequency curve of an arbitrary smooth chirp. We obtain the conditions ensuring that, for any smooth chirp, there always exists a sufficiently close CC. If these conditions are satisfied, the set of the CCs forms a tight template grid which can be used to search for an unknown smooth chirp. Finally, we examine the implementation of such grid for the toy (but realistic) model given by the inspiralling binary chirp.

A. Chirplet chains: piecewise linear frequency

All smooth chirps in Eq. (1) are supported in the TF domain $\mathcal{D}$, a rectangle of width T and of height equal to the Nyquist bandwidth $f_s/2$, as illustrated in Fig. 1. Let $\{(t_j = j\delta_t, f_m = m\delta_f); j = 0, \cdots, N_t, m = 0, \cdots, N_f\}$ be a regular TF grid led on $\mathcal{D}$ by splitting the time axis into N_t intervals of size $\delta_t \equiv T/N_t$, and the frequency axis into N_f bins of size $\delta_f \equiv f_s/(2N_f)$.

In the following, the subscripts j and m designate the index of the time interval and the frequency bin of the grid, respectively. The index $k \in \{0, \cdots, N-1\}$ denotes the time index of a sample.

A *chirplet* is a short piece of signal whose frequency varies linearly between two successive nodes of the grid. In the time interval j, we denote the time and frequency coordinates of the chirplet extreme points by (j, m_j) and $(j + 1, m_{j+1})$. In the TF plane, it is thus represented by a line joining the grid nodes (t_j, f_{m_j}) and $(t_{j+1}, f_{m_{j+1}})$ (see Fig. 1). Concretely, this means that the phase $\phi_k = \phi(t_s k)$ of a chirplet is a quadratic function of time, as follows, for $t_j \leq k t_s < t_{j+1}$

$$\phi_k \equiv a_j t_{j,k}^2 + b_j t_{j,k} + \theta_{j-1}, \qquad (11)$$

where $a_j = \pi(f_{m_{j+1}} - f_{m_j})/\delta_t$, $b_j = 2\pi f_{m_j}$ and $t_{j,k} = t_s k - t_j$.

We build the chirplet chain (CC) by enforcing chaining rules. The frequency and phase of this chain are continuous. Clearly, the continuity of the frequency is ensured by construction, while the phase continuity requires that

$$\theta_{j-1} = \pi \delta_t (f_{m_j} + f_{m_{j-1}}) + \theta_{j-2}, \qquad (12)$$

for $j \geq 1$, and fixing $\theta_{-1} = 0$. We also require that the slope of the chirplet frequency as well as the difference between the slopes of the frequencies of two consecutive chirplets are bounded absolutely. These bounds are given by the two parameters N_r' and N_r'' respectively such that
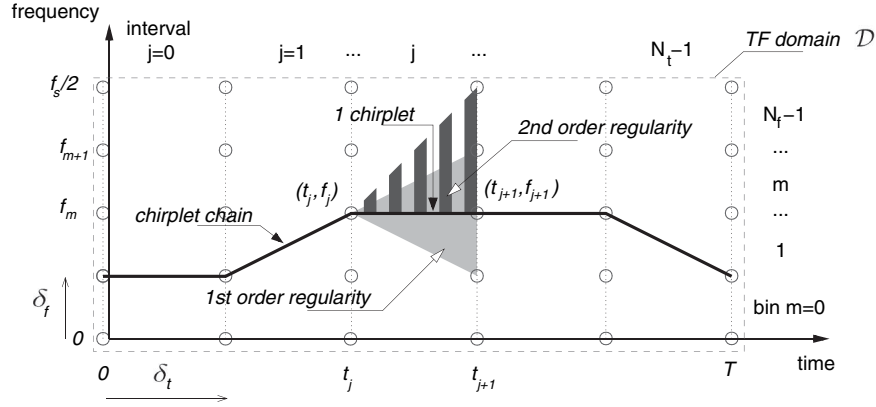
FIG. 1.   Chirplet chains—The TF domain of interest $\mathcal{D}$ is tiled by a regular TF grid of N_t time intervals and N_f frequency bins. A chirplet is a short piece of signal whose frequency varies linearly between two successive nodes of the grid. It is thus represented by a line joining the grid nodes. The slope of the chirplet frequency is limited (triangular region in light gray, here $N'_r = 1$). We chain the chirplets, imposing the continuity of the chain and limiting the difference between the slopes of two consecutive chirplets (triangular region with dark gray stripes, here $N''_r = 1$). Admissible chirplets in time interval j belong to the intersection of these two regions associated to the regularity constraints. Clearly, a chirplet chain is represented by a broken line in the TF plane.

(i) $$|m_{j+1} - m_j| \le N'_r \qquad \text{and}$$
(ii) $|m_{j+1} - 2m_j + m_{j-1}| \le N''_r$.

Clearly, a CC is represented by a broken line in $\mathcal{D}$. The two parameters N'_r and N''_r control the regularity of this line. Consistently, we will refer to (i) and (ii) as *regularity constraints*.

The instantaneous frequency of a smooth chirp is associated to a smooth curve in $\mathcal{D}$. In the same manner that broken lines are good approximations of smooth curves, CCs are good approximations of smooth chirps. Since CCs form a finite discrete set, they sample[2] the set of smooth chirps. In other words, they form a *template grid* of this set.

It is important to know whether this template grid is sufficiently tight i.e., whether for any smooth chirp, there always exists a sufficiently close CC. The template grid tightness is controlled by the choice of the four parameters defining the set of CCs, namely, the TF grid parameters N_t, N_f and the regularity parameters N'_r and N''_r. The first and preliminary step to address the tightness question is to define a distance measuring the "similarity" (or ambiguity) between two different chirps.

B. Distance in the set of smooth chirps

We follow the approach suggested in [18] and assume that we "receive" a chirp whose phase ϕ is different than the template phase ϕ^*. We set $x_k \hat{=} s_k = A\cos(\phi_k + \theta)$, and consider

$$\mathcal{L}(\phi, \phi^*) \equiv \frac{\ell(s; \phi) - \ell(s; \phi^*)}{\ell(s; \phi)}. \qquad (13)$$

---

[2]Strictly speaking, CCs do not *sample* the set of smooth chirps since they do not belong to this set (the second derivative of their frequency is not defined at the boundaries of the grid time intervals and it is thus not bounded).

Clearly, $\mathcal{L}$ measures the reduction factor of the "detection peak" due to the mismatch between the chirp present in the data and the chosen template. It is a relative measurement done with respect to the ideal case where the template matches exactly the considered chirp. In this sense, it can be interpreted as a SNR loss.

Since $\mathcal{L} \ge 0$ and equals 0 when $\phi = \phi^*$, it can be interpreted as a *distance* between the chirps. Note that $\mathcal{L}$ does not depend upon A. It depends only on the phases ϕ and ϕ^*, but this dependency is difficult to perceive intuitively from its definition in Eq. (13).

An approximated but much simpler expression can be obtained when the chirp and the template are close by Taylor expanding $\mathcal{L}$ for small $\Delta_k \equiv \phi^*_k - \phi_k$ and retaining the leading terms. The approximation is detailed in Appendix B and leads to the following expression

$$\mathcal{L}(\phi, \phi^*) \approx \frac{1}{N} \sum_{k=0}^{N-1} (\Delta_k - \Delta)^2, \qquad (14)$$

with $\Delta = 1/N \sum_{k=0}^{N-1} \Delta_k$.

Interestingly, we recognize in this expression the empirical estimate of the variance of the phase difference Δ_k. With this definition of the distance, two chirps are "identical" (their distance measured by $\mathcal{L}$ is zero) if and only if they have the same phase evolution up to an additive offset.

C. Is the CC grid tight?

In this section, we address the grid tightness problem and find the regularity and TF grid parameters which yield a tight template grid of CCs. We proceed as follows: we first consider an arbitrary smooth chirp of phase ϕ. Then we construct a CC "geometrically" close to this chirp. We check if this CC is admissible i.e., if it satisfies the regularity constraints. This imposes two conditions on the
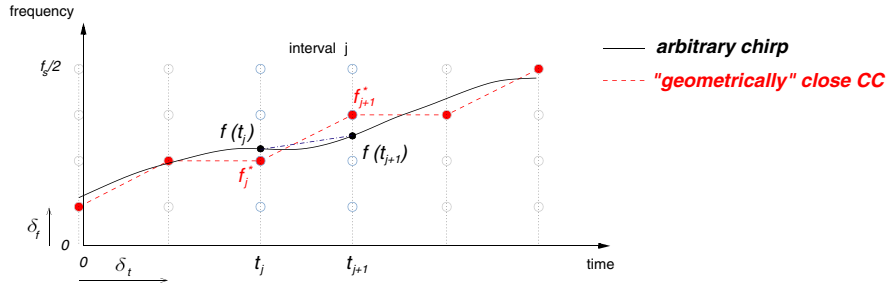
FIG. 2 (color online). Smooth chirp (solid curve) and its geometrically close CC (dotted broken line).

parameters. Finally, we check whether it is *effectively* close to the chirp (as measured by the distance). This yields the loss due to the approximation of the chirp by a CC in the worst case. For tight CC grid, this loss (homogeneous to a SNR loss) has to be small which imposes one more condition on the parameters.

1. Geometrically close CC

Let $\phi(t)$ and $f(t)$ be the phase and frequency of an arbitrary smooth chirp. The frequency evolution appears as a smooth curve in the TF plane.

We construct a CC geometrically close to this chirp as follows: for each $j = 0, \cdots, N_t$, we choose the node (t_j, f_j^*) of the jth column of the TF grid defined in Sec. IV A which is nearest to the point $(t_j, f(t_j))$. We draw the broken line which joins these nodes (see Fig. 2). The associated CC is the geometrically close CC to the chirp under consideration and we denote its phase ϕ^*.

2. Admissibility of the geometrically close CC

For given chirping rate limits $\dot{F}$ and $\ddot{F}$, the geometrically close CC may or may not satisfy the regularity constraints. This depends on the regularity and TF grid parameters. Below, we investigate this question.

a. 1st order regularity—Let us consider the chirplet of the interval j, we have

$$|f_{j+1}^* - f_j^*| \leq |f_{j+1}^* - f(t_{j+1})| + |f(t_{j+1}) - f(t_j)| + |f_j^* - f(t_j)|. \tag{15}$$

Using the mean value theorem[3] (see e.g., [19]), we get $|f(t) - f(s)| \leq \dot{F}|t - s|$ from which we deduce a bound on $|f(t_{j+1}) - f(t_j)|$. By construction, we have $|f_j^* -$

---

[3]Let the function $g(\cdot)$ be continuous in the open interval (a, b) and differentiable in the closed interval $[a, b]$. The mean value theorem states that there exists c in (a, b) such that $g(b) - g(a) = \dot{g}(c)(b - a)$ where $\dot{g}(\cdot)$ is the derivative of $g(\cdot)$. Consequently, we have $|g(b) - g(a)| \leq \dot{G}|b - a|$ with $\dot{G} = \sup_{x \in (a,b)}(\dot{g}(x))$.

$f(t_j)| \leq \delta_f/2$ and this leads to

$$|f_{j+1}^* - f_j^*| \leq \delta_f + \dot{F}\delta_t. \tag{16}$$

Thus, the geometrically close CC satisfies the regularity constraint (i) mentioned in Sec. IV A if

$$N_r' \geq \dot{F}\delta_t/\delta_f + 1. \tag{17}$$

We rewrite this condition in the following form

$$N_r' \geq 4\frac{N'}{N_t}\frac{N_f}{2N} + 1, \tag{18}$$

where $N' \equiv \dot{F}T^2$ is an adimensional quantity which depends only on the fundamental characteristics of the smooth chirp model.

b. 2nd order regularity—We consider two successive chirplets in intervals $j - 1$ and j. Using a similar method, the difference of their slopes can be bounded by

$$|f_{j+1}^* - 2f_j^* + f_{j-1}^*| \leq |f(t_{j+1}) - 2f(t_j) + f(t_{j-1})| + 2\delta_f. \tag{19}$$

Two consecutive applications of the mean value theorem to a function $f(\cdot)$ which satisfies Eq. (2) for all $s \in [r, t]$ with $0 \leq r < t \leq T$ yield the following result

$$|f(t) - 2f(s) + f(r)| \leq \ddot{F}((t - s)2 + (r - s)2)/2 + \dot{F}|t - 2s + r|. \tag{20}$$

Using $r = t_{j-1}$, $s = t_j$ and $t = t_{j+1}$, we get

$$|f_{j+1}^* - 2f_j^* + f_{j-1}^*| \leq \ddot{F}\delta_t^2 + 2\delta_f. \tag{21}$$

Therefore, the geometrically close CC satisfies the regularity constraint (ii) mentioned in Sec. IV A if

$$N_r'' \geq \ddot{F}\delta_t^2/\delta_f + 2. \tag{22}$$

We rewrite the above condition as

$$N_r'' \geq \frac{4}{3}\left(\frac{N''}{N_t}\right)^2\frac{N_f}{2N} + 2, \tag{23}$$

where $N'' \equiv \sqrt{3\ddot{F}T^3}$ is an adimensional quantity which depends only on the fundamental characteristics of the smooth chirp model, and thus from the astrophysical input. It is related to the maximum overall curvature of the chirp

frequency or more precisely to the largest number $\ddot{F}T^3$ of Fourier bins that the chirp frequency can sweep, the linear trend being removed.

3. Is the geometrically close CC effectively close?

We obtain a worst case estimate of the distance between the smooth chirp and its geometrically close CC by bounding the variations of their phase difference. We begin with bounding the frequency discrepancy.

The starting point is the following lemma (inspired from [20], p. 23) obtained from the application of the mean value theorem and some algebraic manipulations. If $f(\cdot)$ satisfies Eq. (2) for all $s \in [r, t]$ then

$$\left| f(s) - \left[f(r) + \frac{s - r}{t - r}(f(t) - f(r)) \right] \right| \leq \ddot{F}(t - r)^2. \tag{24}$$

However, this upper bound can be slightly improved: the term $(t - r)^2$ overestimates the more precise bound $g(s) \equiv \min\{(t - s)(t - (s + r)/2), (s - r)((s + t)/2 - r)\}$. [Note that $g(t) = 0$ and $g(r) = 0$ as expected.] In the worst case, we have $s = (t + r)/2$ and $g(s) = 3/8(t - r)^2$ as opposed to $(t - r)^2$. We include this gain in the following.

We apply the lemma in Eq. (24) with $t = t_{j+1}$ and $r = t_j$. The term inside the square brackets in the left hand side of Eq. (24) is equal to the frequency at time s of the chirplet obtained by joining $f(t_{j-1})$ to $f(t_j)$ (see the dot-dash line in Fig. 2). We denote this frequency $\tilde{f}(s)$, and then obtain

$$|f(s) - \tilde{f}(s)| \leq 3\ddot{F}\delta_t^2/8. \tag{25}$$

Since $|f(s) - f^*(s)| \leq |f(s) - \tilde{f}(s)| + |\tilde{f}(s) - f^*(s)|$ and $|\tilde{f}(s) - f^*(s)| \leq \delta_f/2$, we have

$$|f(s) - f^*(s)| \leq 3\ddot{F}\delta_t^2/8 + \delta_f/2 \equiv \Delta_f. \tag{26}$$

By definition $\phi(t) - \phi(r) = 2\pi \int_r^t f(s)ds$. Integrating both sides of the above inequality between two successive points $r = t_s(k - 1)$ and $t = t_s k$ for $k \in \{1, \cdots, N - 1\}$, we get

$$|\Delta_k - \Delta_{k-1}| \leq 2\pi\Delta_f t_s, \tag{27}$$

which constrains the variations of the phase difference $\Delta_k = \phi^* - \phi_k^*$.

We prove in Appendix C that the approximated distance $\mathcal{L}(\phi, \phi^*)$ as shown in Eq. (14) is maximum under this constraint, when $\Delta_k = \pm 2\pi\Delta_f t_s k$. In this case, the maximum is $\mathcal{L}(\phi, \phi^*) = (\pi\Delta_f T)^2/3 \equiv \mu'$.

We can finally state the following *tight template grid* theorem: for all smooth chirps of phase ϕ, there exists a CC of phase ϕ^* such that

$$\mathcal{L}(\phi, \phi^*) \leq \mu', \tag{28}$$

where $\mu' = \pi^2 T^2 (3\ddot{F}\delta_t^2/4 + \delta_f)^2/12$ is the maximum (i.e., in the worst case) *energy* SNR loss due to the mis-

match between the smooth chirp and the chosen template given by a close CC. The corresponding maximum *amplitude* SNR loss is $\mu = 1 - \sqrt{1 - \mu'} \approx \mu'/2$ for small μ'. Note that the *amplitude* SNR loss is linked to the *minimal match MM* defined in [18] by the relation $MM = 1 - \mu$. We express the maximum SNR loss μ in terms of the CC parameters as

$$\mu = \frac{\pi^2}{96} \left[\frac{1}{2}\left(\frac{N''}{N_t}\right)^2 + \frac{1}{2}\left(\frac{2N}{N_f}\right) \right]^2. \tag{29}$$

In principle, this loss can be made arbitrarily small by choosing N_t and N_f adequately. Therefore, *the grid of CC can sample the set of smooth chirps tightly.*

It is evident that two types of losses contribute to μ. The first one is related to the *geometrical error* due to the fact that the model is a broken line: within the time intervals of the TF grid, the model is a straight line which cannot perfectly follow the curvature of the smooth chirp frequency. The finer the grid along the time axis, the smaller the time interval, the better the line fits the smooth chirp frequency, thus reducing this error. The other is related to the *quantization error* as we require the node of the best broken line to belong to the TF grid: there is a difference between the best broken line we can possibly draw and the closest (quantized) one with vertices belonging to the grid. The finer the grid along the frequency axis, the closer the quantized line from the original, thus reducing this error.

When $N_t = N''$ and $N_f = 2N$, the maximum SNR loss is of order $\sim \pi^2/96 \approx 10\%$ and the two types of errors contribute equally. The same maximum SNR loss can be achieved with other choices for N_t and N_f. In the next section, we propose a criterion to solve this indetermination.

D. Smallest tight CC grid

As elaborated in the previous section, the tight template grid theorem gives the condition on the TF grid parameters N_t and N_f which ensures that the template set (the CCs) covers all the feasible set (the smooth chirps) with a given accuracy specified by the maximum SNR loss. The same accuracy can be achieved with several pairs of parameters, leading to a parameter indetermination.

For a small maximum SNR loss, the maximization of the LLR in Eq. (10) performed over the set of smooth chirps can be safely replaced by a maximization over the set of CCs, i.e.,

$$\ell_{\max}(x) \simeq \max_{\text{all CCs}} \{\ell(x; \phi)\}. \tag{30}$$

The statistic $\ell_{\max}$ in Eq. (30) results from the CC which maximizes the statistic or in other words, from the waveform of the template set which best fits the data. Generally speaking, when the data is noise *only*, the larger the number of (reasonably different) waveforms in the template set, the

larger the risk that one of the waveforms fits the noise and consequently, the larger the false alarm rate.

The TF grid parameters N_t and N_f influence very differently the number of CCs. The above argument suggests to select the parameters which minimize the numbers of CCs, for a given specified maximum SNR loss. We refer to the *smallest tight CC grid* as the set of CCs which results from this constrained optimization.

Let us first estimate the number of CCs. According to the regularity conditions, each of the number $N_c \sim N_f(2N_r' + 1)$ of possible chirplets in a given time interval can be chained to (at most) $2N_r'' + 1$ chirplets in the next time interval. Counting CCs is then a combinatorial problem. We have N_c chirplets in the first time interval, and $2N_r'' + 1$ possible choices for the $N_t - 1$ successive time intervals. Neglecting what happens at the lower and upper boundaries of the frequency axis (i.e., near DC and Nyquist), we obtain an upper-bound on (the logarithm of) the number N_{cc} of CCs as

$$\ln N_{cc} \lesssim \ln(2N_r'N_f) + (N_t - 1)\ln(2N_r'' + 1). \quad (31)$$

In practice, we have $N_t \gg 1$. The second term largely dominates the right-hand side and the first term can be neglected. We thus have $\ln N_{cc} \sim N_t \ln(2N_r'' + 1)$.

At this point, it is convenient to introduce $u \equiv N''/N_t$ and $v \equiv 2N/N_f$ and express the *smallest tight CC grid* problem with these variables. From the regularity constraints, we have $N_r'' = 4u^2/(3v) + 2$. We want to minimize the number of CCs

$$\ln N_{cc} \propto g(u, v) \equiv \frac{1}{u} \ln\left(\frac{8}{3}\frac{u^2}{v} + 5\right), \quad (32)$$

subject to a given maximum SNR loss i.e., $u^2 + v = C \equiv 8\sqrt{6\mu}/\pi$.

Combining the derivatives of the objective $dg = \partial_u g\, du + \partial_v g\, dv$ and of the constraint $dv = -2u\, du$, we obtain the equation giving the admissible point where the derivative dg/du vanishes, viz.

$$\frac{\ln y}{y - 5} + \frac{7}{4y} - \frac{3}{4} = 0 \quad (33)$$

where we defined $y \equiv 8u^2/(3v) + 5$. This equation can be solved numerically and gives $y \approx 8.95$. Let $\alpha \equiv u^2/v = 3(y - 5)/8$ be the ratio between the two errors contributing to μ. We obtain the smallest tight template grid when this ratio is $\alpha \sim 1.48$. For a required μ, we get the parameters of the resulting grid as follows. Using the constraint, we have $u = \sqrt{C\alpha/(1 + \alpha)}$ and $v = C/(1 + \alpha)$, from which we obtain the parameters,

$$N_t = 0.52\mu^{-1/4}N'', \qquad N_f = 0.78\mu^{-1/2}N. \quad (34)$$

Interestingly, this also implies that $N_r'' = 4\alpha/3 + 2 \approx 4$ is a constant (i.e., does not depend on μ). The last parame-

ter N_r' is directly determined by substituting Eqs. (34) in (18).

The parameters of the smallest tight template grid may not be always suitable in practice (see the later discussion on the implementation and numerical contingencies in Sec. V B 5) but they give interesting indications.

At this point, it is useful to see with an example if the proposed model and template grid sound tractable in a realistic case.

E. Toy model and CC parameters

We use the inspiralling binary chirps as a toy model to check whether the various parameters have reasonable order of magnitudes in this physically realistic situation. We consider the Newtonian approximation of the chirp whose frequency evolution is given by [17]

$$f(t) = f_0\left(1 - \frac{t - t_0}{T}\right)^{-3/8} \quad \text{for } t < t_0 + T, \quad (35)$$

where t_0 denotes the arrival time. In practice, the arrival time corresponds to the time at which the chirp enters the detector's bandwidth i.e., when its frequency reaches the low frequency (seismic) cutoff (denoted f_0) of the interferometric detectors. T defines the chirp duration, i.e. the time taken by the chirp from the arrival time until the binary coalescence.

The chirp duration can thus be estimated by

$$T \sim 1.3\mathrm{s}\left(\frac{f_0}{20}\right)^{-8/3}\left(\frac{M}{50M_\odot}\right)^{-5/3}, \quad (36)$$

where M is the total mass (objects of equal masses).

In this calculation, we assume the seismic cutoff frequency[4] of 20 Hz.

We fix $\dot{F}$ and $\ddot{F}$ to the corresponding values of the first and second derivatives of the chirp frequency, pertaining to the last stable circular orbit (LSCO[5]) viz.,

$$f_{\mathrm{LSCO}} \sim 88.4\ \mathrm{Hz}\left(\frac{M}{50M_\odot}\right)^{-1}, \quad (37)$$

$$\dot{F} \sim 1.33\ \mathrm{kHz/s}\left(\frac{M}{50M_\odot}\right)^{-2}, \quad (38)$$

$$\ddot{F} \sim 74\ \mathrm{kHz/s}^2\left(\frac{M}{50M_\odot}\right)^{-3}. \quad (39)$$

We note that T, f_{LSCO}, $\dot{F}$ and $\ddot{F}$ decrease with an increasing mass. When M increases, the chirp is thus shorter, less steep and curved, and it reaches only the lower part of the

---

[4]This is the seismic cutoff frequency targeted by the detector Virgo.

[5]For nonrotating stars, the LSCO is when the objects are at the distance $r = 6\ GM/c^2$.

frequency band. From the above equations, we deduce that

$$N' \sim 2.2 \times 10^3 \left(\frac{M}{50M_\odot}\right)^{-16/3}, \qquad N'' \sim 698 \left(\frac{M}{50M_\odot}\right)^{-4}.$$
(40)

The sampling frequency f_s is fixed by the width of the observational band of the GW detector, namely $f_s = 2048$ Hz. We thus have

$$N = f_s T \sim 2662 \left(\frac{M}{50M_\odot}\right)^{-5/3}.$$
(41)

Following Sec. IV D and fixing $\mu = 10\%$, the smallest tight CC grid has the following parameters for the TF grid

$$N_t \sim 645 \left(\frac{M}{50M_\odot}\right)^{-4},$$
(42)

$$N_f \sim 6566 \left(\frac{M}{50M_\odot}\right)^{-5/3},$$
(43)

and for the regularity, we have

$$N_r' \sim 17 \left(\frac{M}{50M_\odot}\right)^{-4/3}, \qquad N_r'' \sim 4.$$
(44)

The orders of magnitude for the various parameters appear to be reasonable. Since these parameters do not increase with M, the template grid defined with the above values remains acceptable and tight for higher masses $M \geq 50M_\odot$.

V. FIND THE BEST CHIRPLET CHAIN

In Sec. IV, we have shown that the SNR loss due to the use of a CC instead of the *ideal* template can be made small with an appropriate choice of parameters i.e., by making the CC grid tight. In other words, the problem of detecting a smooth chirp is equivalent to the one of detecting a CC as stated by Eq. (30). The maximization over the set of CCs—involved in the latter case—has the great advantage that it can be resolved numerically.

A. The exhaustive search is not feasible

Since CCs are in finite number, an obvious maximization procedure is to try them all and select the one which gives the maximum. To understand whether this solution is tractable, we need to know how many CCs are there. We consider that the search parameters N_t, N_f, N_r' and N_r'' are known and can be obtained from the physical and grid tightness requirements as discussed earlier.

We already presented an estimate of the number of CCs in Eq. (31) and saw that it grows exponentially with the number of time intervals of the TF grid. This estimate computed for the toy model example presented in the previous section gives $\log_{10} N_{cc} \approx 1400$. Clearly, this number is too large for an exhaustive search (i.e., computing ℓ for all possible CCs) to be carried out in real-time on

existing computers. Generally speaking, since the number of CCs increases exponentially with N_t, the cost of an exhaustive search scales exponentially with N_t and thus with the problem size N.

In the next section, we propose an algorithm which gives a good estimate for the optimal CC instead of the exact solution of the maximization problem described in Eq. (30). However, as opposed to the exhaustive search, the computational cost of this algorithm scales as a polynomial of the problem size N.

B. Near-optimal search

The maximization of $\ell(x; \phi)$ in Eq. (30) is a combinatorial maximization problem. The existence of an efficient solving algorithm for such problem is related to the structural properties of the "objective" function to be maximized, that is, ℓ in the present case. In this section, we show that ℓ can be reasonably approximated by a path integral computed over a time-frequency representation (TFR) of the data. The structure of the approximated statistic allows us to perform its maximization efficiently with dynamic programming. The approximation goes through two stages with an intermediate step for the reformulation of the statistic in the TF plane.

1. Approximation 1: for a CC, cosine and sine are almost orthogonal

As shown in Eq. (A8), the statistic ℓ can be expressed as

$$\ell(x; \phi) = \frac{1}{2} \left[\left(\sum_{k=0}^{N-1} x_k \tilde{c}_k \right)^2 + \left(\sum_{k=0}^{N-1} x_k \tilde{s}_k \right)^2 \right],$$
(45)

where the templates $\tilde{c}_k$ and $\tilde{s}_k$ are the orthonormalized counterparts of the waveforms in quadrature $\cos\phi_k$ and $\sin\phi_k$ obtained from the Gram-Schmidt procedure as given below

$$\tilde{c}_k = \frac{\cos\phi_k}{\sqrt{n_c}} \qquad \tilde{s}_k = \frac{n_c \sin\phi_k - n_x \cos\phi_k}{\sqrt{n_c \mathcal{O}}}.$$
(46)

Let $\{\cos\phi_k\}$ and $\{\sin\phi_k\}$ be the vectors in $\mathbb{R}^N$ associated to the quadrature waveforms. As it appears in the above expressions, these vectors are generally not orthonormal. Their deviation from orthonormality can be quantified with two parameters, defined by

$$\delta \equiv \frac{n_c - n_s}{n_c + n_s}, \qquad \epsilon \equiv \frac{2n_x}{n_c + n_s},$$
(47)

which are related to their vector lengths n_c, n_s and their scalar product n_x. The parameter δ measures the relative difference in the vector lengths while ϵ measures the angle between them. The vectors are orthonormal if and only if both δ and ϵ are zero.

Intuitively, if the quadrature waveforms oscillate sufficiently, they should be close to orthonormality and δ and ϵ are expected to be small. This intuition is examined in

detail in Appendix D, in which we exploit the fact that ϕ is not arbitrary but it is the phase of a CC. We show that if ϕ is the phase of a CC whose node frequencies are in the bandwidth $f_l \leq f_{m_j} \leq f_s/2 - f_l$ for all j with

$$f_l \approx 2.5\delta_f\sqrt{N'}\left(\frac{N_f}{N}\right)^{3/2}\left(\frac{0.1}{\eta}\right)^{1/2}, \tag{48}$$

then $|\delta| \lesssim \eta$ and $|\epsilon| \lesssim \eta$.

In the following, we assume that this condition is satisfied. This imposes the CC frequency not to approach arbitrarily close to the DC nor to the Nyquist frequencies. Since the amplitude of the instrumental noise of GW interferometers diverges rapidly when going close to DC, it is not expected to detect GWs at low frequencies. Therefore the reduction of the bandwidth in the low frequency region should not be a problem as long as f_l remains small. We will check later with examples that the reduction of the useful bandwidth is indeed sufficiently small.

Using Eqs. (46) and (47), we can write $\tilde{c}_k$ and $\tilde{s}_k$ in terms of δ and ϵ as

$$\tilde{c}_k = \left(\frac{2}{N(1+\delta)}\right)^{1/2}\cos\phi_k, \tag{49}$$

$$\tilde{s}_k = \left(\frac{2}{N(1+\delta)}\right)^{1/2}\frac{(1+\delta)\sin\phi_k - \epsilon\cos\phi_k}{\sqrt{1-(\delta^2+\epsilon^2)}}, \tag{50}$$

noting that $n_c = N(1+\delta)/2$, $n_s = N(1-\delta)/2$ and $n_x = N\epsilon/2$.

Inserting this expression in Eq. (45) and taking the limit for small δ and ϵ, we find that $\ell(x;\phi) \to \hat{\ell}(x;\phi) \equiv (x_c^2 + x_s^2)/N$, the reminder $R(\delta,\epsilon) \equiv \hat{\ell}(x;\phi) - \ell(x;\phi)$ being given by

$$R(\delta,\epsilon) = \frac{1}{N}\frac{\delta(x_c^2 - x_s^2) + 2\epsilon x_c x_s - (\delta^2 + \epsilon^2)(x_c^2 + x_s^2)}{1-(\delta^2+\epsilon^2)}. \tag{51}$$

Considering that we have $\delta^2 + \epsilon^2 \leq \eta^2$ (see Appendix D) and

$$\frac{|x_c^2 - x_s^2|}{x_c^2 + x_s^2} \leq 1, \qquad \frac{|2x_c x_s|}{x_c^2 + x_s^2} \leq 1, \tag{52}$$

the relative error can be bounded as

$$\frac{|R(\delta,\epsilon)|}{\hat{\ell}(x;\phi)} \leq \frac{2\eta - \eta^2}{1-\eta^2} \approx 2\eta, \tag{53}$$

for small η.

Provided a good choice of η (and checking the consequences on f_l), this approximation error can be made small. We can safely replace ℓ by $\hat{\ell}$ which we express as the following complex sum :

$$\hat{\ell}(x;\phi) = \frac{1}{N}\left|\sum_{k=0}^{N-1}x_k\exp(i\phi_k)\right|^2. \tag{54}$$

2. Go to time-frequency: Moyal

The expression of $\hat{\ell}$ in Eq. (54) computes the canonical Hermitian scalar product between the data and a complex template waveform. While Parseval's formula allows an equivalent formulation of this scalar product in the frequency domain, *Moyal's formula* does the same in the TF domain, provided the use of a unitary TFR. One such TFR is the *discrete Wigner-Ville* (WV) distribution defined in [21] and given by

$$w_x(n,m) \equiv \sum_{k=-k_n}^{k_n} x_{p_{n,k}} x_{q_{n,k}}^* e^{-2\pi imk/(2N)}, \tag{55}$$

with $k_n \equiv \min\{2n, 2N-1-2n\}$, $p_{n,k} \equiv \lfloor n + k/2 \rfloor$ and $q_{n,k} \equiv \lfloor n - k/2 \rfloor$ where $\lfloor \cdot \rfloor$ gives the integer part. The arguments of w_x are the time index n and the frequency index m which correspond in physical units, to the time $t_n = t_s n$ and the frequency is $f_m = f_s m/(2N)$ for $0 \leq m \leq N$ and $f_m = f_s(N-m)/(2N)$ for $N+1 \leq m \leq 2N-1$. Thus, the frequency axis gets sampled at twice the usual rate (as performed by the FFT). The WV distribution is associated with a particular sampling of the TF plane. As discussed later in Sec. V B 5, this leads to some restrictions on the TF grid used for defining CCs.

Let $\{x_k\}$ and $\{y_k\}$ be two time series. Moyal's formula states that [21]

$$\left|\sum_{k=0}^{N-1}x_k y_k^*\right|^2 = \frac{1}{2N}\sum_{n=0}^{N-1}\sum_{m=0}^{2N-1}w_x(n,m)w_y(n,m). \tag{56}$$

Using Eqs. (54) and (56), we rewrite $\hat{\ell}$ as the inner-product of two TFRs, namely, the WV of the data w_x and the *template WV* w_e which is the WV of complex template waveform $e_k \equiv \exp i\phi_k$,

$$\hat{\ell}(x;\phi) = \frac{1}{2N^2}\sum_{n=0}^{N-1}\sum_{m=0}^{2N-1}w_x(n,m)w_e(n,m). \tag{57}$$

Qualitatively, we expect that the TFR of a chirp signal have large values essentially in the vicinity of a curve corresponding to their instantaneous frequency and vanishes elsewhere. The template WV w_e being the TFR of a chirp, it shares these characteristics. In the following section, we make use of this feature to simplify the statistic.

3. Approximation 2: the WV of a CC is almost Dirac

With continuous time and frequency variables, it is well-known that ([22], p. 130 and also 217) the WV of a *linear chirp* (i.e., a chirp whose frequency is a linear function of time) is a Dirac distribution along the TF line associated to the chirp frequency.

We assume that this remains reasonably true for discrete time and frequency and when the chirp is nonlinear (and, in particular, when it is a CC). More precisely, we consider that we have
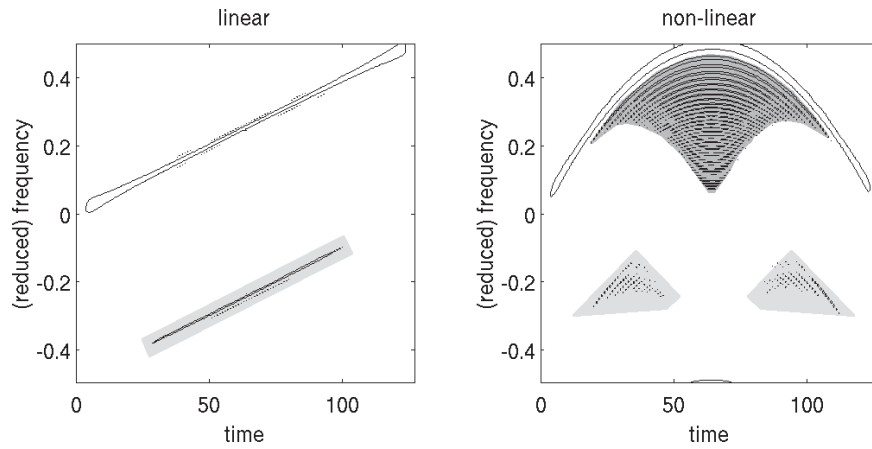
FIG. 3.　Discrete Wigner-Ville of two chirp signals—The signals are normalized to unit ℓ_2 norm and we show the contour at the level 1/8. Left: when the chirp frequency is a linear function of time, its WV is almost Dirac along the corresponding TF line in the TF half plane associated to positive frequencies. For the negative frequencies, the WV distribution exhibits *aliasing terms* (we highlight them by a background in light gray) which we neglect in the simplified model in Eq. (58). Right: when the chirp frequency is not linear (here, it is a parabolic chirp), *interference terms* appear (evidenced by a dark gray background). Their contribution are also disregarded in the simplified model. Note that the WV of the nonlinear chirp chosen for this illustration do present aliasing terms (with a light gray background), but they have a smaller amplitude than in the linear case.

$$w_e(n, m) \approx 2N\delta(m - m_n), \qquad (58)$$

where $m_n = [2Tf_n]$ where $[\cdot]$ denotes the nearest integer.

Here, f_n is the instantaneous frequency of the (possibly nonlinear) chirp. Equation (58) dissembles two approximations which we explain now.

For discrete time and frequency, the discrete WV of a linear chirp can be calculated analytically [21]. For the positive frequencies i.e., for $0 \le m \le N$, the model in Eq. (58) is an acceptable approximation of the exact result, as illustrated in Fig. 3. However, there is a significant difference in the negative frequencies i.e., for $N + 1 \le m \le 2N - 1$. In this region, the discrete WV exhibits *aliasing terms* (clearly seen in the left panel of Fig. 3) which are closely related to the unitarity property of the WV. In [21], the aliasing terms are shown to be oscillating terms (switching signs) with smaller amplitude than the preponderant terms modeled by Eq. (58). We can then expect their contribution to the summation in Eq. (57) to be negligible.

It is well known [22] that *interference terms* appear when computing (both continuous and discrete) WVs of nonlinear chirps. They can be related to the quadratic nature of this distribution (see [22] for a detailed analysis of the nature and geometry of these interferences). Interference terms change sign rapidly (see Fig. 3, right panel) and can be neglected for the same argument invoked for aliasing terms.

Inserting Eq. (58) into Eq. (57), we get the following approximation of $\hat{\ell}$:

$$\tilde{\ell}(x; \phi) = \frac{1}{N} \sum_{n=0}^{N-1} w_x(n, m_n). \qquad (59)$$

We see that this statistic results from the integral of the WV of the data along the TF path determined by the CC frequency f_n. In other words, this integral is the area under this TF path. We refer to this quantity as the *path length*.[6]

With this approximation, the maximization of the statistic in Eq. (30) amounts to finding the path giving the largest integral, or the longest path. Efficient methods exist for longest path problems [23]. These methods exploit the structural properties of path length (or integral) measurement, in particular, *additivity*. The length of this entire path can be measured by splitting the path and summing the length of its constituent parts. Thanks to this property, the maximization problem can be decomposed into a recursive series of small problems, each of them being solvable in polynomial time. This is the main principle of *dynamic programming* (DP), which we describe in the next section.

We remind the reader that, contrarily to the new statistic $\tilde{\ell}$, the exact statistic ℓ is not additive. DP *cannot* be applied to maximize ℓ.

4. Maximization with dynamic programming

DP is a classical method [23] for solving combinatorial optimization problems. As explained in the previous section, the idea is to decompose the problem into smaller ones that can easily be solved. In our context, the natural decomposition is given by the tiling of the time axis into chirplet intervals i.e., $t_j \le t_s n < t_{j+1}$ or equivalently $jb \le n \le (j + 1)b - 1$ where $b \equiv \delta_t/t_s$ is the number of

---

[6]If we see the WV as a Lebesgue measure (although this is an misuse of language since the WV can take negative values), the integral in Eq. (59) effectively defines a path length.

samples in an interval. The overall path integral is equal to the sum of the integrals computed in each chirplet interval marked with a superscript index as follows

$$\tilde{\ell}(x;\phi) = \sum_{j=0}^{N_t-1} \tilde{\ell}^j(x;\phi) \qquad (60)$$

with $\tilde{\ell}^j(x;\phi) \equiv \dfrac{1}{N} \sum_{n=jb}^{(j+1)b-1} w_x(n, m_n^j), \qquad (61)$

where $m_n^j = [2Nf_n^j]$ and the frequency f_n^j follows the line joining the grid points (t_j, f_{m_j}) and $(t_{j+1}, f_{m_{j+1}})$. We also denote with a subscript index j, the path integral up to interval j, viz.

$$\tilde{\ell}_j(x;\phi) = \sum_{j'=0}^{j} \tilde{\ell}^{j'}(x;\phi). \qquad (62)$$

DP relies on the *principle of optimality*. We elaborate this principle with the help of Fig. 4. We consider the chirplet in time interval j. In a chain passing through this chirplet, the regularity constraints limit the choice of preceding chirplets in the time interval $j-1$. We suppose that there are only three such chirplets; namely α, β and γ.

Now, consider the time interval $j-1$. We assume that we know the chain passing through the chirplet z (z being either α, β or γ) and giving the largest path integral summed up to the interval $j-1$. We denote this quantity by $\tilde{\ell}_{j-1}^{(z)}$. (In this discussion, the chirplet and its associated CC are designated by the same label).

We compute the path integral contribution in jth interval for the considered chirplet, and add the result to $\tilde{\ell}_{j-1}^{(z)}$ to obtain $\tilde{\ell}_j^{(z)}$ for all the three paths $z = \alpha, \beta$ and γ.

We mark with $(\star)$ the optimal chain associated to the *global* maximum of $\tilde{\ell}$ (i.e., summing from interval 0 to $N_t - 1$) which we denote $\tilde{\ell}^{(\star)}$. We further assume that this optimal chain $(\star)$ follows (α) up to interval $j-1$, continues following the considered chirplet in interval j and

proceeds to the last interval $j = N_t - 1$ with some chain (δ), hence $\tilde{\ell}^{(\star)} = \tilde{\ell}_j^{(\alpha)} + \tilde{\ell}^{(\delta)}$ where $\tilde{\ell}^{(\delta)}$ denotes the contribution of the chain (δ).

The principle of optimality states that the optimal chain $(\star)$ has the largest path integral $\tilde{\ell}_{j-1}^{(\star)}$ at interval $j-1$ as compared to *all* the other chains passing by the same chirplet in interval j. In particular, this means that $\tilde{\ell}_{j-1}^{(\star)} = \tilde{\ell}_{j-1}^{(\alpha)}$ is larger that $\tilde{\ell}_{j-1}^{(\beta)}$ and $\tilde{\ell}_{j-1}^{(\gamma)}$.

Proof by contradiction: Let us assume that $\tilde{\ell}_{j-1}^{(\beta)} > \tilde{\ell}_{j-1}^{(\alpha)}$. We construct the chain $(\triangle)$ formed by (β), the considered chirplet in interval j and the chain (δ). This CC is admissible. By construction, its path integral $\tilde{\ell}^{(\triangle)} = \tilde{\ell}_j^{(\beta)} + \tilde{\ell}^{(\delta)}$ is larger than $\tilde{\ell}^{(\star)}$. Therefore, the chain $(\star)$ is not optimal which contradicts our hypothesis—QED.

We apply this principle recursively starting from interval $j = 0$ and incrementing. For each chirplet interval and for all N_c chirplets of interval j, we keep only the CC maximizing the path integral up to this point and we discard the others. This procedure "prunes the combinatorial tree" and avoids to consider useless candidates before going to the next interval.

When the recursion reaches the last interval $N_t - 1$, we end up with a number N_c of CCs ending with a different last chirplet and having the maximum path integral among all chains with the same last chirplet. Finally, within these "short-listed" candidates, we select the chain with the largest $\tilde{\ell}$ which is the *global maximum*.

5. Numerical contingencies and computational cost

To summarize, we started with the initial problem in Eq. (30) of finding the CC with the largest statistic. We rephrased this problem (using approximations) into a *longest path* problem in the TF plane. Here, *path* refers to the TF curve followed by the frequency of the CC, and the *length* is given by the integral of the WV of the data along the path. The maximization of the path length over the set
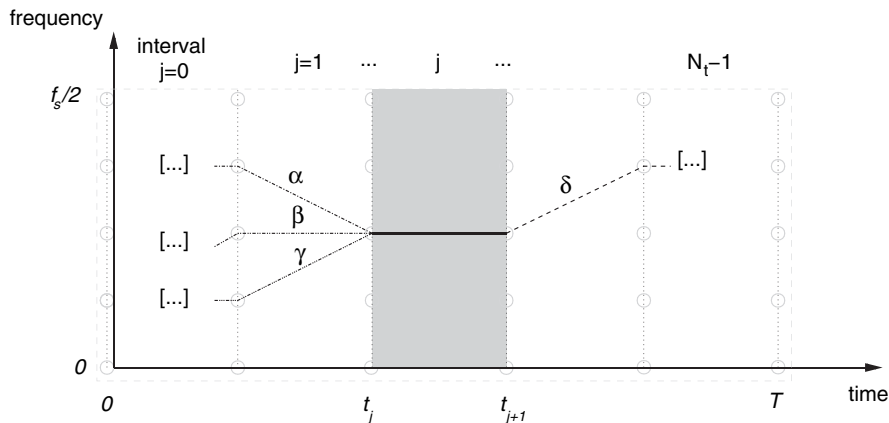

FIG. 4.   Principle of optimality of DP

of CCs can be performed efficiently using DP. The resulting algorithm is tractable numerically as shown by the estimate obtained in the second part of this section.

The definition of the CCs does not comprehend the fact that we only have access to discretized versions of the data and of their associated TF domain, denoted $\mathcal{D}$ in Sec. IV A. We begin this section by a discussion on these aspects.

a. Discretization issues—On one hand, the definition of the set of CCs relies on a TF grid sampling the *continuous* TF domain $\mathcal{D}$. Theoretically, this grid can be refined arbitrarily. On the other hand, the search operates effectively using the *discretized* version of $\mathcal{D}$, resulting from the sampling associated to the WV. This fixes a maximum TF resolution which cannot be surpassed.

It is useless to increase the resolution of the TF grid used for defining CCs beyond the one defined by WV. The WV divides the time axis into N intervals and the frequency axis into N bins.[7] Consequently, we have the following limitations, $N_t \leq N$ and $N_f \leq N$. Furthermore, in order to have time intervals (resp. frequency bins) of equal size , the TF grid parameters N_t (resp. N_f) must be divisors of N.

All these requirements limit the choice of N_t and N_f. It may happen that the parameters of smallest tight CC grid are not suitable because of that. Note that in the case we consider, we are generally led to adopt the finest resolution for the frequency axis i.e., $N_f = N$.

b. Estimate of the computational cost—

We estimate the computational cost by counting the floating point operations for all the primary subparts in the course of the procedure. The computation of the WV of the data involves N FFTs with time base $2N_f$ [21], such that the cost of this part is about $5NN_f\log_2 N_f$ (assuming a standard implementation with RADIX-2).

The number of operations required by DP is better estimated by grouping them by types, rather than by a sequential assessment. The path integral $\tilde{\ell}_j$ in Eq. (61) is computed (with b additions) only once for each N_c chirplets of all N_t intervals, with a corresponding cost equal to N_cN.

For each of the N_c chirplets in each interval, the algorithm selects among the (at most) $2N_r'' + 1$ possibly connected paths. This procedure is repeated $N_t - 1$ times, and thus requires $\sim N_t N_c (2N_r'' + 1)$ operations.

Knowing that the number of chirplets is $N_c \approx (2N_r' + 1)N_f$, the overall cost C thus scales with

$$C \propto 5NN_f\log_2 N_f + [N + (2N_r'' + 1)N_t](2N_r' + 1)N_f, \quad (63)$$

which is a polynomial of the problem size.

VI. APPLICATIONS

In this section, the proposed method is evaluated with several numerical tests and compared with two other TF based algorithms for the detection of unmodeled chirps, namely, the Signal Track Search (STS) [12] and Time-Frequency Clusters (TFC) [14]. The simulation code[8] of these tests uses the implementation of these algorithms provided by [24]. We first give a brief presentation of STS and TFC.

A. Existing algorithms

1. Signal Track Search

We have seen earlier that the TFR of a chirp signal can be essentially described in the TF plane as a regular alignment of large values forming ''ridges'' along the instantaneous frequency evolution. The STS uses this observation as a heuristic basis: detecting chirps amounts to finding ridges in a TFR.

In practice, the algorithm extracts the ridges from the WV distribution[9] of the data. Because of the presence of noise, image processing techniques are required to get a good ridge extraction. The authors chose an algorithm which is normally used for road extraction from aerial images. This algorithm is based on the fact that a ridge is a locus of points having a maximum curvature (as measured by the second derivative) in the transverse direction and a small gradient along the longitudinal direction. A hysteresis thresholding procedure is applied over the second derivative of the WV (smoothed by a low pass filter) to detect TF points which suffice the above condition, and to grow iteratively chains of TF points from these ridge precursors. In [12], the ridge length (number of TF points in a ridge) is then employed as the detection statistic. However, we do not use this definition here, but we rather consider the one given by the largest path integral computed along the detected ridges. We observed that this variation outperforms the original definition of STS.

2. TFClusters

TFClusters is initially thought to detect short oscillatory transients (and not specifically chirps). The TFR of such transient is sparse i.e., the TF contents is essentially described by few components of large amplitude. The basic idea of TFClusters is that, for reasonable SNR, the amplitude of the transient components is larger than the noise.

This motivates the thresholding of the TFR of the data, given by the spectrogram (modulus square of short-time

---

[7]It is possible to modify slightly the definition of WV in Eq. (55) to get a finer sampling of the frequency axis and keep unitarity. We reserve this possibility for later investigations.

[9]In [12], the authors use the standard definition of the discrete WV originally proposed by Claasen-Mecklenbräuker (see [21] for a definition and a detailed discussion). This definition differs from the one presented in Sec. V B 3. In particular, it does not satisfy unitarity.

Fourier transform), to retain the TF points with the largest values. A clustering algorithm is used to group the selected points. "Significant" clusters are chosen whose cardinals are greater than a threshold. "Insignificant" clusters are merged iteratively if they are sufficiently close to eventually form significant clusters. The statistic is then chosen to be the maximum sum of the TF powers over the clusters in the resulting list of significant clusters.

3. Discussion

It is important to stress a major difference between STS and TFClusters and the proposed method. By construction, the formers work well provided that the signal "stands above" the noise somewhere in the TF plane. If we define a local SNR in the TF plane (by computing at a given TF point, the squared difference of the mean values of the TFR under the hypotheses H_0 and H_1 divided by its variance under H_0), then this is equivalent to say that the *local* SNR has to be large at least for some TF points. However, just like the standard matched filtering, a detection with the best CC algorithm requires the *global* SNR (obtained by summing the local SNRs for all TF points) to be large. Clearly, this is a less stringent condition.

TF path integration is a central ingredient of the best CC search. This idea is also used for other methods developed for the detection of other GW sources. For instance, for inspiralling binaries, we can cite [25,26] and for the periodic GW sources, the Hough transform [27] and the stack slide searches [28].

Several distinctions must be stressed. First, the TF representation we use here (discrete WV) satisfies a specific and crucial property, namely, unitarity. This allows us to link the final statistic to the quadratic matched filtering. TF representations based on short-time Fourier or wavelet bases used by the above methods are not unitary. Second, the other methods require a precise model of the TF path (relying on the astrophysical source modeling) as opposed to our method.

For the problem addressed here i.e., the detection of unmodeled chirps, we have shown that CCs can be treated as an effective finite template grid. We could then imagine to apply one of the above methods and integrate along the entire set of TF paths associated to CCs. This is however computationally impossible because of the too large number of CCs, as already discussed in Sec. VA.

B. Newtonian chirps: illustrations and benchmark

For the illustration of the best CC search, we use the Newtonian chirp signal introduced in Sec. IV E. We recall that the frequency of such chirp is a power law given by Eq. (35). Normally, the Newtonian chirp also includes a prescribed evolution of the chirp amplitude. However, for simplicity and better match with our initial model, we decide not to take this into account and keep the chirp envelope to a constant.

The Newtonian chirp is completely defined by the total mass M of the binary (if we assume that the objects have equal masses) and its initial frequency f_0. Figure 5 presents an example of a typical Newtonian chirp signal, where we set $M = 7.3M_\odot$ and $f_0 = 96$ Hz. The chirp duration is $T = 0.5$ s. We fix the sampling frequency to $f_s = 2048$ Hz (therefore, the number of samples is $N = Tf_s = 1024$). A white Gaussian noise of unit variance is added to the signal.

Within the GW literature, it is customary to define the SNR through matched filtering (assuming the initial phase is known *a priori*). We follow this definition which gives in the present case,

$$\rho^2 \equiv \sum_{k=0}^{N-1} s_k^2 \approx A^2 N/2. \qquad (64)$$

We note that, with this definition we have $\rho^2 = 2\ell(s; \phi)$ (the factor of 2 accounts for the unknown initial phase).

We choose to scale the chirp amplitude to a SNR $\rho = 20$.

We apply the best CC search to this signal with the following search parameters. We arbitrarily fix the chirping rate limits to be $\dot{F} = 8192$ Hz/s and $\ddot{F} = 917.5$ kHz/s^2. These values are quite smaller than the ones expected at the LSCO (see Sec. IV E) but the time instant when these limits are reached is close (few tenths of milliseconds before) to the LSCO. In Fig. 5, the time instants when the chirp (the solid line on the right panel) reaches the chirping rate limits (with dotted vertical lines) and when the binary system reaches LSCO (with dashed-dotted horizontal line) are indicated. We fix the frequency axis sampling to the finest accessible resolution i.e., $N_f = N = 1024$. Similarly, we choose the smallest possible chirplet size with $N_t = N/2$. The rest of the parameters are derived from the regularity constraints. In this respect, it is useful to calculate the adimensional characteristics of the problem i.e., $N' = 2048$ and $N'' = 586.6$. The resulting parameters are $N'_r = 9$ and $N''_r = 3$, which gives a maximum SNR loss $\mu \approx .28$.

We recall that the best CC search relies on the approximation of the optimal statistic by a complex sum presented in Sec. V B 1. The parameter η controls the relative precision of this approximation. From the results of Sec. V B 1 and the above general chirp specifications, the approximation holds with a precision $\eta = 0.14$ in a frequency bandwidth $[f_l, f_s/2 - f_l]$ with $f_l = 96$ Hz which coincides (at least for the low frequencies, which are most important) with the frequency support of the present chirp.

Figure 5 presents the result of the best CC search with the above choice of parameters. The best CC closely matches the actual instantaneous frequency in the region where the regularity constraints are satisfied.

An example is obviously not sufficient to evaluate the method thoroughly. Receiver operating characteristics (ROC) gives a systematic assessment of the performance. The ROC of a given statistic l is the diagram giving the
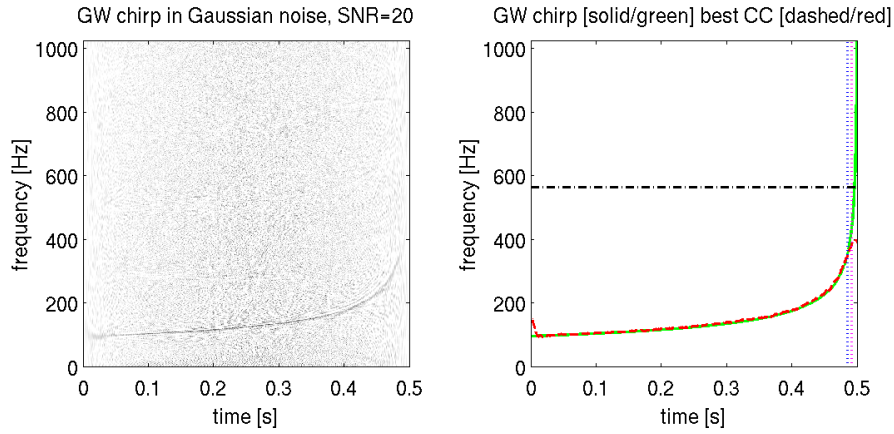
FIG. 5 (color online).   Newtonian chirp in white Gaussian noise—Left: WV distribution of the signal. Only positive contributions are displayed (negative ones are set to zero) with a gray-scaled color map going from white (minimum i.e, zero) to black (maximum). Right: the best CC in dashed/red closely matches the actual instantaneous frequency in solid/green in the region where the regularity constraints are satisfied. We indicate the instant when the chirp reaches the chirping rate limits with the dotted vertical lines and the frequency at LSCO with dashed-dotted horizontal line.

detection probability $P_d(l_0) \equiv \mathbb{P}(l \geq l_0|H_1)$ versus the false alarm probability $P_{fa}(l_0) \equiv \mathbb{P}(l \geq l_0|H_0)$ at a given SNR and for all thresholds l_0.

For this exercise, due to computing limitations, we prefer short signals with a small number of samples N. We choose a Newtonian chirp with total mass $M = 11M_\odot$ and initial frequency $f_0 = 96$ Hz which has a short duration $T \sim 250$ ms. Choosing the sampling frequency $f_s = 1024$ Hz, we have $N = 256$ samples. White Gaussian noise is added to the signal and the amplitude is scaled such that the SNR is $\rho = 10$.

We fix the chirping rate limits to $\dot{F} = 8.192$ kHz/s and $\ddot{F} = 1.05$ MHz/s^2. Like the above example, these limits are reached at a time instant close to the LSCO. We choose the finest TF grid parameters $N_t = 128$ and $N_f = 256$, and the regularity parameters $N_r' = 9$, $N_r'' = 4$. The resulting CC grid is tight with $\mu \approx 0.4$.

Concerning STS[10] and TFClusters,[11] we set their free parameters empirically using the recommendations available in the references, without a precise fine-tuning. Figure 6 displays a single trial and Fig. 7 presents the ROCs of the three methods presented previously. We see that the best CC search outperforms the two others as expected.

Here, we wish to add few remarks regarding the comparison between the best CC search and STS. The improve-

ment in the ROC of the best CC with respect to STS has two origins. First, the use of a unitary discrete WV instead of the standard WV helps in increasing the detection probability by few percent. The unitarity preserves the power in TF plane and hence improves the efficiency. Second, the major part of the improvement comes from the TF pattern search procedure. As explained in Sec. VI A, the use of a global search criterion instead of a local one is a crucial ingredient.

It is interesting to compare these ROCs with what could optimally achieve an imaginary observer which knows in advance the targeted chirp. Since this *clairvoyant* observer knows the chirp phase exactly, he can apply the optimal statistic i.e., the quadrature matched filter obtained in Appendix A. The ROCs of the quadrature matched filter can be obtained analytically (under Gaussian noise hypotheses). The false alarm and detection probabilities are given, respectively, by [29]:

$$P_{fa}(l_0) = \exp(-l_0), \tag{65}$$

$$P_d(l_0) = 1 - \exp(-\rho_c^2/2) \sum_{n=0}^{+\infty} \frac{(\rho_c^2/2)^n}{n!} I_{l_0}(n+1) \tag{66}$$

where $I_y(x) \equiv 1/\Gamma(x) \int_0^y e^{-u}u^{x-1}du$ is the incomplete Gamma function.

This ROC depends only on one parameter, namely, the SNR ρ_c. The ROC curve of the clairvoyant statistic with $\rho_c = \rho$ provides an absolute upper bound on the detection probability. Obviously, having in hand all the information makes a very large difference with respect to the case where we only know that the incoming GW is a smooth chirp. The detection probability of the clairvoyant statistic is very close to 1 over the entire range of values chosen for the false alarm rate. This is why we do not show this curve. It is more interesting to compare the performances of the

_____

[10]Following the notations of [12], the size of the Gaussian kernel of the presmoothing filter is fixed to $\sigma = 2$. The low and high thresholds of the hysteresis are set to $3.3/$pixel^2 and $10/$pixel^2 resp.

[11]The TFR is given by the short-time Fourier transform computed over nonoverlapping blocks of 16 samples (i.e., intervals of ≈ 7.8 ms). The frequency axis is tiled into 32 bins (i.e., a resolution of 32 Hz). We use the nominal values given in [14] for the rest of the parameters namely $p = 0.1$, $\sigma = 5$, $\delta = [0, 0, 0, 0, 0, 0, 2, 3, 4, 4]$ and $\alpha = 0.25$.
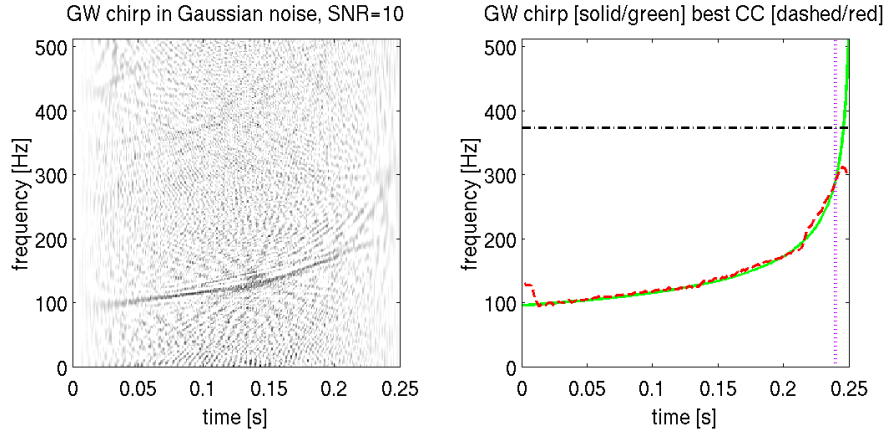
FIG. 6 (color online). Newtonian chirp in white Gaussian noise—Left: WV distribution of the signal (displayed similarly as in Fig. 5). Right: actual chirp frequency in solid/green and best CC in dashed/red. We indicate where the chirp reaches the chirping rate limits with dotted vertical lines and the frequency at LSCO with dashed-dotted horizontal line.

various methods with the ones of the clairvoyant observer for SNRs $\rho_c < \rho$. More precisely, we adjust ρ_c in such a way that the resulting curve matches reasonably well the ROC of the best CC search in the region of interest i.e., for false alarm probabilities in the range 10^{-5} to 10^{-4}. Since the SNR is inversely proportional to the distance of the GW source, the ratio of the actual SNR to the best-fit value ρ/ρ_c gives the reduction factor of the sight distance with respect to the ideal (and nonaccessible) situation where we have at our disposal all the information about the chirp we want to detect. We include the fitted clairvoyant ROC in

Fig. 7. The ratio in the sight distance can be estimated $\sim 10/6.15 \approx 1.6$.

C. Random CCs and robustness

While benchmarks based on Newtonian chirps are satisfactory for a comparison of several detection methods in a nominal situation, they do not provide a test for the robustness i.e., a measurement of the ability to detect reliably a large class of different chirps.

In this section, we present ROC curves computed using *random CCs*. Random CCs are generated by chaining chirplets which are randomly chosen in a range specified by regularity constraints. Therefore, the frequency of a random CC follows a kind of random walk in the TF plane. We generate a new random CC for each trial made to estimate the detection probability.

The detection of random CCs is obviously much more difficult than the detection of a single chirp. It is an effective test of the method robustness. No classical approaches (e.g., based on banks of quadrature matched filters as for inspiralling binary chirps) can be applied successfully in this case.

We assume the same general characteristics of the Newtonian chirp used in the first example in the previous section, namely $T = 0.5$ s, $f_s = 2048$ Hz, thus $N = 1024$ samples, $\dot{F} = 8192$ Hz/s and $\ddot{F} = 917.5$ kHz/s^2. We already computed satisfactory search parameters for this setup. Therefore, they remain unchanged ($N_t = 512$, $N_f = 1024$, $N_r' = 9$ and $N_r'' = 3$). The random CCs are generated on the same basis, but with a time interval slightly larger, the regularity parameters being increased accordingly i.e., $N_t = 64$, $N_f = 1024$, $N_r' = 65$ and $N_r'' = 57$. We use an additive white Gaussian noise.

Figure 8 presents an example of such signal (with SNR $\rho = 20$) and the result of the application of the best CC search. Figure 9 displays the ROC curve of the best CC
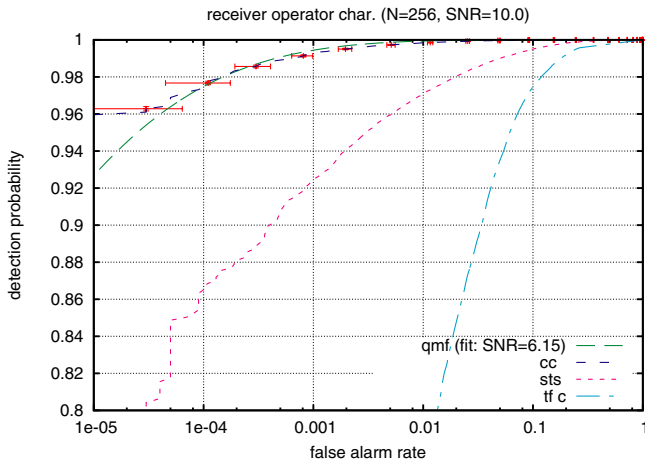

FIG. 7 (color online). Newtonian chirp in white Gaussian noise—Comparison of ROCs of the best CC search (dashed/blue, with error bars in solid/red) with STS (dotted/magenta) and TFC (dashed-dotted/cyan). The computation of each ROCs is perfomed over 2×10^5 trials (half for the false alarm probability and half for the detection probability). The diagram also includes the ROC of the clairvoyant quadrature matched filter (bold dashed/green) shown here with the SNR $\rho_c = 6.15$ adjusted to reasonably fit the ROC of the best CC search.
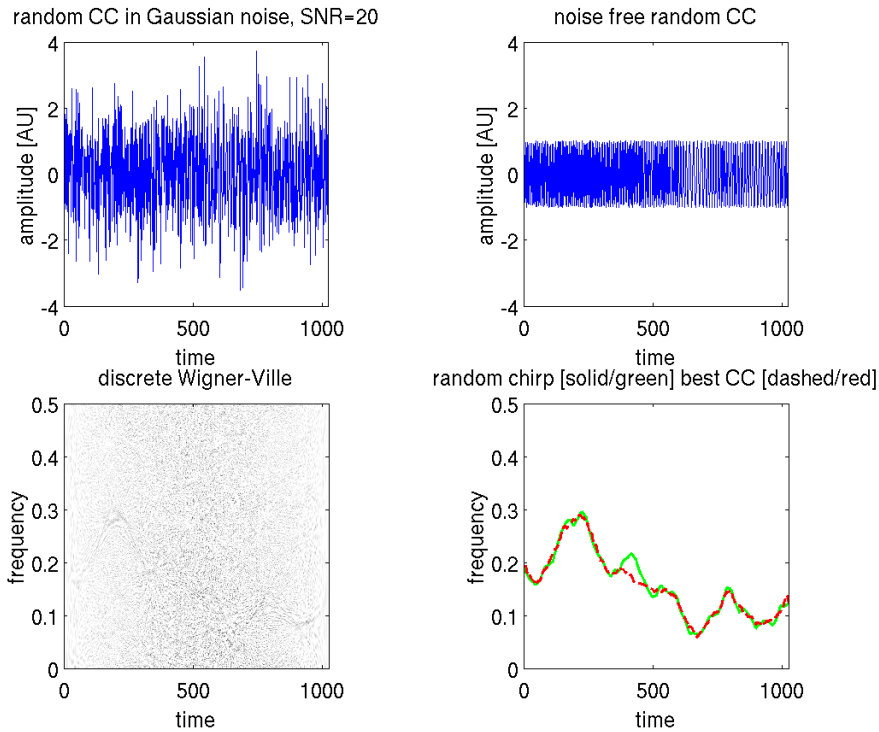
FIG. 8 (color online).   Random CC in white Gaussian noise—In these plots, we arbitrarily set $f_s = 1$. Top left: example of a random CC in white Gaussian noise. Top right: noise free random CC. Bottom left: WV distribution of the signal (displayed similarly as in Fig. 5). Bottom right: actual chirp frequency in solid/green and best CC in dashed/red. It is worthwhile to note that, although the best CC can lose track for some time because of noise fluctuations, it is able to recover the exact TF path.

search (with SNR $\rho = 12$) along with the one of the clairvoyant quadrature matched filter adjusted to an adequate SNR. We estimate a loss in the sight distance with respect to the clairvoyant case to be a factor of ~ 2.6. Best CC search "sees" to distances comparable to (in the sense,

with a reduction factor less than 1 order of magnitude) what classical methods achieve in other GW detection problems.

The computational cost of this search as estimated by Eq. (63) is about 1.42×10^8 of floating points operations for one block of duration $T = 0.5$ s. Assuming 10% overlap between successive blocks, real-time processing can be achieved with a computing power of 2.8 Gflops which is less than what a single standard workstation can handle today.
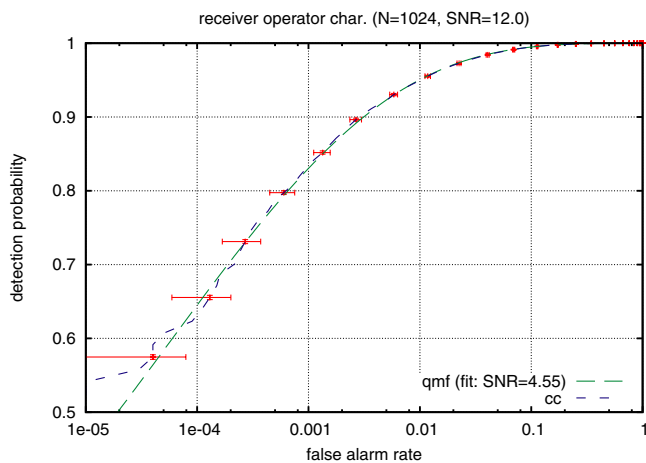
VII. CONCLUDING REMARKS

Smooth chirps define a general model of "nearly physical" GW chirps. Chirplet chains—chains of linear chirplets—allow the design of tight template grids for the detection of smooth chirps. The optimal detection requires these grids to be searched thoroughly to find the template which best matches with the data. Although the shear large number of templates prevents the use of an exhaustive search, near-optimal detection can be performed with the time-frequency based procedure presented here. Its originality lies in the clear link established between the optimal statistic and the proposed search algorithm. In particular, it justifies the choice of a specific time-frequency representation (the unitary discrete WV) and pattern search algorithm (TF path integral and dynamic programming). We

FIG. 9 (color online).   Random CC in white Gaussian noise—This diagram displays the ROC of the best CC search (dashed/blue, obtained from 2×10^5 trials) compared with the analytical ROC of the clairvoyant matched filter (bold dashed/green) with the SNR $\rho_c = 4.55$, adjusted to produce a reasonable fit.

have evaluated that best CC search is computationally tractable for detection of typical GW chirps.

It is important to emphasize several features which makes the proposed method attractive in practice. First, the free parameters (the chirp duration T and the chirping rate limits $\dot{F}$ and $\ddot{F}$) are few and directly related to physical characteristics. Second, the principle "*He who can do more can do less*" applies here: smooth chirps is a very general class of chirps. This model, and thus the search algorithm can be easily modified and adapted to incorporate additional astrophysical information. For instance, it is easy to search only chirps with an increasing (or decreasing) frequency. One may also want a more stringent constraint on the chirping rate at low frequencies than at high frequencies. The inclusion in the algorithm of a dependency of the chirping rate limit upon the frequency is straightforward. This leaves the possibility of a compromise between efficiency (since the restriction of the set of admissible waveforms due to additional constraints reduces the false alarm rate) and robustness, depending on the quantity and reliability of the information available on a specific GW source. Third and finally, it is simple to restrict the search to chirps starting and/or finishing at given time-frequency location. This feature could be used for partially known chirps whose waveforms is known only on a part of the total duration. Those signals could be detected with a hybrid approach combining a standard matched filtering where the waveform model is available, and best CC search for the rest.

ACKNOWLEDGMENTS

APPENDIX A: MAXIMIZING LLR $\Lambda(\cdot)$ OVER THE INITIAL PHASE φ_0

In this appendix, we maximize the statistic $\Lambda(x; \{\hat{A}, \varphi_0, t_0, \phi(\cdot)\})$ over the initial phase φ_0. We recall that

$$\Lambda(x; \{\hat{A}, \varphi_0, t_0, \phi(\cdot)\}) = \frac{1}{2\mathcal{N}}\left(\sum_{k=0}^{N-1} x_k \bar{s}_k\right)^2, \qquad (A1)$$

where $\mathcal{N} \equiv \sum_{k=0}^{N-1} \bar{s}_k^2$ is the norm of $\bar{s}_k = \cos(\phi_k + \varphi_0)$. To keep the notations simple, we do not mention all parameters explicitly in this appendix and set $\Lambda(x; \varphi_0) = \Lambda(x; \{\hat{A}, \varphi_0, t_0, \phi(\cdot)\})$.

In the literature concerning the detection of inspiralling binaries of compact objects [17,18], this maximization is usually performed assuming that $\mathcal{N}$ is independent of φ_0.

This assumption is correct when the two quadratures $\cos\phi_k$ and $\sin\phi_k$, viewed as vectors of $\mathbb{R}^N$, are orthonormal (i.e., orthogonal and of same norms). In this case, we have $n_c = n_s = N/2$ and $n_x = 0$ where n_c, n_s and n_x are the norms and cross-products of the quadratures as defined in Eqs. (7). Inserting this into

$$\mathcal{N} = n_c\cos^2\varphi_0 - n_x \sin(2\varphi_0) + n_s\sin^2\varphi_0, \qquad (A2)$$

we conclude that $\mathcal{N} = N/2$ is a constant.

However, for general phase evolution, the quadrature waveforms are not necessarily orthonormal. This is approximately true when the chirp oscillates sufficiently rapidly during many cycles (e.g., for inspiralling binaries of small mass). Since we are considering chirps with an arbitrary phase and of relatively short duration, such assumption is not realistic and we opt for the general case keeping the dependency of $\mathcal{N}$ upon φ_0.

Expanding $\bar{s}_k$ in terms of two quadratures and rewriting Eq. (A1), we get

$$\Lambda(x; \varphi_0) = \frac{(x_c \cos\varphi_0 - x_s \sin\varphi_0)^2}{2(n_c\cos^2\varphi_0 - n_x \sin(2\varphi_0) + n_s\sin^2\varphi_0)}, \qquad (A3)$$

where x_c and x_s are the cross-correlation of the data with $\cos\phi_k$ and $\sin\phi_k$ as defined in Eq. (6).

To proceed with the maximization, we first examine the special case where the quadratic waveforms are linearly dependent i.e., $\cos\phi_k \propto \sin\phi_k$ for all k. This implies that we are in the degenerate case where $\phi_k = \varphi_0$ is constant. Introducing the two angles $\varphi = \arg(x_c + ix_s)$ and $\eta = \arg(\sqrt{n_c} + i\sqrt{n_s})$, we can rewrite Eq. (A3) as

$$\Lambda(x; \varphi_0) = \frac{1}{2}\frac{x_c^2 + x_s^2}{n_c + n_s}\frac{\cos^2(\varphi + \varphi_0)}{\cos^2(\eta + \varphi_0)}. \qquad (A4)$$

The proportionality of the quadrature waveforms implies that $\sqrt{n_c}x_s \pm \sqrt{n_s}x_c = 0$ which gives $\sin(\eta - \varphi) = 0$, and hence $\eta = \varphi + \pi\mathbb{Z}$. We conclude that $\Lambda(x; \varphi_0)$ remains constant for all φ_0 and is equal to the statistic given by

$$\ell(x; t_0, \phi) = \frac{x_c^2 + x_s^2}{2N}. \qquad (A5)$$

In the non degenerate case, we compute the derivative of the statistic as given in Eq. (A3) with respect to φ_0. Its numerator turns out to be a second order polynomial of $\tan\varphi_0$. The root associated to the local maximum is

$$\hat{\varphi}_0 = \tan^{-1}\left(\frac{x_s n_c - x_c n_x}{n_x x_s - n_s x_c}\right) \qquad (A6)$$

which gives the ML estimator of the initial phase.

Inserting this expression in Eq. (A3) yields

$$\ell(x; t_0, \phi) \equiv \Lambda(x; \hat{\varphi}_0) = \frac{n_s x_c^2 - 2n_x x_c x_s + n_c x_s^2}{2\mathcal{O}}, \quad (A7)$$

where $\mathcal{O} \equiv n_c n_s - n_x^2 > 0$.

We can reexpress this statistic as

$$\ell(x; t_0, \phi) = \frac{1}{2}\left[\left(\sum_{k=0}^{N-1} x_k \tilde{c}_k\right)^2 + \left(\sum_{k=0}^{N-1} x_k \tilde{s}_k\right)^2\right], \quad (A8)$$

where $\tilde{c}_k$ and $\tilde{s}_k$ are the orthonormalized counterparts of the waveforms in quadrature $\cos\phi_k$ and $\sin\phi_k$ obtained from the Gram-Schmidt procedure as given below

$$\tilde{c}_k = \frac{\cos\phi_k}{\sqrt{n_c}}, \qquad \tilde{s}_k = \frac{n_c \sin\phi_k - n_x \cos\phi_k}{\sqrt{n_c \mathcal{O}}}, \quad (A9)$$

and referred to as templates of phase ϕ.

In practice, this orthonormalization is indeed performed for the detection of inspiralling binaries (see [30], p. 3046) and is justified with heuristic arguments. The derivation shows that it results directly from the maximization of the LLR.

APPENDIX B: TAYLOR APPROXIMATION OF THE DISTANCE BETWEEN CHIRPS

In this appendix, we detail the approximation of the statistic $\ell(s; \phi^*)$ with $s_k = A\cos(\phi_k + \varphi_0)$ and assuming that the template phase ϕ^* is close to the phase ϕ of the signal s present in the data. We start from the following Taylor expansion of $\ell(s; \phi^*)$ for small $\Delta_k \equiv \phi_k^* - \phi_k$

$$\ell(s; \phi*) = \ell(s; \phi) + \sum_{k=0}^{N-1} \partial_k \ell|_{\phi^*=\phi} \Delta_k$$
$$+ \frac{1}{2} \sum_{k,l=0}^{N-1} \partial_{kl}^2 \ell|_{\phi^*=\phi} \Delta_k \Delta_l + \cdots, \quad (B1)$$

where the partial derivatives $\partial_k \equiv \partial/\partial\phi_k^*$ and $\partial_{kl}^2 \equiv \partial^2/\partial\phi_k^* \partial\phi_l^*$ are taken with respect to the samples of the template phase ϕ^*. Next, we examine this expansion term by term and obtain analytical expressions as a function of the phase samples $\{\phi_k\}$ and $\{\phi_k^*\}$.

1. First derivative: local extremum

From Eq. (9), we write the statistic ℓ as the ratio $\ell = n/d$. The numerator is $n = n_s x_c^2 - 2n_x x_c x_s + n_c x_s^2$ and the denominator is $d = 2(n_c n_s - n_x^2)$. We thus have $\partial_k \ell = (\partial_k n - \ell \partial_k d)/d$.

We get the following general expressions of the derivative of the numerator[12]

---
[12]Here, we adopt the precedence rule $\partial_k ab = (\partial_k a)b$.

$$\partial_k n = \partial_k n_s x_c^2 + n_s 2x_c \partial_k x_c - 2(\partial_k n_x x_c x_s + n_x \partial_k x_c x_s$$
$$+ n_x x_c \partial_k x_s) + \partial_k n_c x_s^2 + n_c 2x_s \partial_k x_s, \quad (B2)$$

and of the denominator

$$\partial_k d = 2(\partial_k n_c n_s + n_c \partial_k n_s - 2n_x \partial_k n_x). \quad (B3)$$

We insert $s_k = A\cos(\phi_k + \varphi_0)$ and work out each of their component term. At the match (when $\phi^* = \phi$), we get

$$\partial_k n_s|_{\phi^*=\phi} = \sin 2\phi_k, \qquad \partial_k n_x|_{\phi^*=\phi} = \cos 2\phi_k, \quad (B4)$$

$$\partial_k n_c|_{\phi^*=\phi} = -\sin 2\phi_k = -\partial_k n_s|_{\phi^*=\phi}, \quad (B5)$$

$$\partial_k x_s|_{\phi^*=\phi} = A\cos\phi_k \cos(\phi_k + \varphi_0), \quad (B6)$$

$$\partial_k x_c|_{\phi^*=\phi} = -A\sin\phi_k \cos(\phi_k + \varphi_0). \quad (B7)$$

Combining all the above expressions, the derivative $\partial_k n$ can be factorized, yielding

$$\partial_k n|_{\phi^*=\phi} = \ell(s; \phi)\partial_k d|_{\phi^*=\phi}, \quad (B8)$$

where $\partial_k d|_{\phi^*=\phi} = 2(n_c - n_s)\sin 2\phi_k - 4n_x \cos 2\phi_k$ and $\ell(s; \phi) = A^2(n_c \cos^2\varphi_0 - n_x \sin 2\varphi_0 + n_s \sin^2\varphi_0)/2$. In conclusion, the first derivative $\partial_k \ell|_{\phi^*=\phi} = 0$ vanishes at $\phi^* = \phi$ which is thus a local extremum.

Using the parameters $\epsilon = 2n_x/N$ and $\delta = (n_c - n_s)/N$ as defined later in Sec. V B (and also discussed in Appendix D), the statistic and the denominator at the match can be expressed as functions of δ and ϵ as

$$\ell(s; \phi) = \frac{A^2 N}{4}(1 + \delta \cos 2\varphi_0 - \epsilon \sin 2\varphi_0) \quad (B9)$$

$$d|_{\phi^*=\phi} = \frac{N^2}{2}(1 - \delta^2 - \epsilon^2). \quad (B10)$$

2. Second derivative and distance

We show in the previous subsection that the first derivative at the match $\partial_k \ell|_{\phi^*=\phi}$ vanishes. Consequently, the second derivative at the match can be expressed simply in terms of the second derivatives of the numerator and denominator at the match, namely

$$\partial_{kl}^2 \ell|_{\phi^*=\phi} = \left[\frac{\partial_{kl}^2 n - \ell(s; \phi^*)\partial_{kl}^2 d}{d}\right]_{\phi^*=\phi}. \quad (B11)$$

We obtain the following general expressions for the second derivatives of the denominator

$$\partial_{kl}^2 d = 2[\partial_{kl}^2 n_c n_s + \partial_k n_c \partial_l n_s + \partial_l n_c \partial_k n_s + n_c \partial_{kl}^2 n_s$$
$$- 2(\partial_k n_x \partial_l n_x + n_x \partial_{kl}^2 n_x)]. \quad (B12)$$

and of the numerator

$$
\begin{aligned}
\partial_{kl}^2 n &= \partial_{kl}^2 n_s x_c^2 + 2(\partial_k n_s x_c \partial_l x_c + \partial_l n_s x_c \partial_k x_c \\
&\quad + n_s \partial_k x_c \partial_l x_c + n_s x_c \partial_{kl}^2 x_c) - 2(\partial_{kl}^2 n_x x_c x_s \\
&\quad + \partial_k n_x \partial_l x_c x_s + \partial_k n_x x_c \partial_l x_s + \partial_l n_x \partial_k x_c x_s \\
&\quad + \partial_l n_x x_c \partial_k x_s + n_x \partial_k x_c \partial_l x_s + n_x \partial_l x_c \partial_k x_s \\
&\quad + n_x x_s \partial_{kl}^2 x_c + n_x x_c \partial_{kl}^2 x_s) + \partial_{kl}^2 n_c x_s^2 + 2(\partial_k n_c x_s \partial_l x_s \\
&\quad + \partial_l n_c x_s \partial_k x_s + n_c \partial_k x_s \partial_l x_s + n_c x_s \partial_{kl}^2 x_s), \quad \text{(B13)}
\end{aligned}
$$

Similarly to the first derivative, we insert the expression of the signal $s_k = A\cos(\phi_k + \varphi_0)$ and evaluate each of the component terms of the above expressions. We have to

In Appendix D, we discuss the range of values taken by ϵ and δ depending on the phase ϕ. We show that these parameters are small $\epsilon, \delta \ll 1$ if the phase ϕ is a CC whose frequency does not come close to DC nor the Nyquist frequency. We assume that this remains true in the more general case, when ϕ is the phase of a smooth chirp. We retain the leading term (of order 0 in ϵ and δ) and get the following approximation

$$
\partial_{kl}^2 \ell|_{\phi^*=\phi} \equiv X_{kl} = \frac{A^2}{2N}((1-\hat{c}_k)(1-\hat{c}_l) + \hat{s}_k \hat{s}_l) \quad \text{(B16)}
$$

where $\hat{c}_k = \cos 2(\phi_k + \varphi_0)$ and $\hat{s}_k = \sin 2(\phi_k + \varphi_0)$.

b. Autoterms, k = l

We consider the case where $k = l$. Now, the second order derivatives do not vanish. In fact, we have

$$
\partial_k^2 n_c|_{\phi^*=\phi} = -2\cos 2\phi_k, \qquad \partial_k^2 n_x|_{\phi^*=\phi} = -2\sin 2\phi_k, \quad \text{(B17)}
$$

$$
\partial_k^2 n_s|_{\phi^*=\phi} = 2\cos 2\phi_k, \quad \text{(B18)}
$$

$$
\partial_k^2 x_c|_{\phi^*=\phi} = -A/2(\cos(2\phi_k + \varphi_0) + \cos\varphi_0), \quad \text{(B19)}
$$

$$
\partial_k^2 x_s|_{\phi^*=\phi} = -A/2(\sin(2\phi_k + \varphi_0) - \sin\varphi_0). \quad \text{(B20)}
$$

The consequence is an additional term D_{kl} to the second order derivative of the statistic $\partial_{kl}^2 \ell|_{\phi^*=\phi} = X_{kl} + D_{kl}$. With a direct calculation, we obtain its exact expression (no approximation needed):

$$
D_{kl} = \frac{A^2}{2}(-1 + \hat{c}_k)\delta_{kl}. \quad \text{(B21)}
$$

distinguish two cases i.e., the nondiagonal cross terms of the Hessian matrix when $k \neq l$ and the diagonal ones when $k = l$.

a. Cross terms, k ≠ l

When $k \neq l$, the above Eqs. (B12) and (B13) are significantly simplified because all the second order cross derivatives are zeros (namely $\partial_{kl}^2 n_c = \partial_{kl}^2 n_x = \partial_{kl}^2 n_s = 0$ and $\partial_{kl}^2 x_c = \partial_{kl}^2 x_s = 0$). We get

$$
\partial_{kl}^2 d|_{\phi^*=\phi} = -4\cos 2(\phi_k - \phi_l), \quad \text{(B14)}
$$

and combined with Eq. (B9),

$$
\begin{aligned}
[\partial_{kl}^2 - \ell(s;\phi^*)\partial_{kl}^2 d]_{\phi^*=\phi} &= \frac{A^2 N}{4}[1 + \cos 2(\phi_k - \phi_l) - \cos 2(\phi_k + \varphi_0) - \cos 2(\phi_l + \varphi_0) + \epsilon(\sin 2(\phi_k + \phi_l + \varphi_0) \\
&\quad - \sin 2\phi_k - \sin 2\phi_l - \sin 2\varphi_0) + \delta(\cos 2(\phi_k + \phi_l + \varphi_0) - \cos 2\phi_k - \cos 2\phi_l + \cos 2\varphi_0)].
\end{aligned}
$$
$$
\text{(B15)}
$$

where the Kronecker symbol is $\delta_{kl} = 0$ for $k \neq l$ and 1 for $k = l$.

3. Approximated distance

From Eqs. (B16) and (B21), and assuming that $|\epsilon|, |\delta| \ll 1$, we have $\ell(s;\phi) \approx A^2 N/4$. The distance defined in Eq. (13) can thus be written as

$$
\begin{aligned}
\mathcal{L}(\phi, \phi^*) &\approx \frac{1}{N}\sum_{k=0}^{N-1}(1-\hat{c}_k)\Delta_k^2 - \left(\frac{1}{N}\sum_{k=0}^{N-1}(1+\hat{c}_k)\Delta_k\right)^2 \\
&\quad - \left(\frac{1}{N}\sum_{k=0}^{N-1}\hat{s}_k\Delta_k\right)^2. \quad \text{(B22)}
\end{aligned}
$$

Considering that Δ_k and Δ_k^2 are slowly varying with respect to $\hat{c}_k$ and $\hat{s}_k$, we argue that, similarly to what is discussed in Appendix D, the positive and negative terms compensate when making the following sums $\sum_k \hat{c}_k \Delta_k$, $\sum_k \hat{s}_k \Delta_k$ and $\sum_k \hat{c}_k \Delta_k^2$. We neglect the small residual, which leads to the final approximation of the distance in Eq. (14).

APPENDIX C: CONSTRAINED MAXIMIZATION OF THE DISTANCE

We rewrite the constrained maximization problem described in Sec. IV C 3 of the distance in Eq. (14) under the constraint in Eq. (27) with simpler notations. We relate them to the initial problem at the end of this appendix.

Let $\{r_k\}$ a series of N real numbers. We want to maximize the empirical variance $V(r)$ expressed by

$$
V(r) = \frac{1}{N}\sum_{k=0}^{N-1} r_k^2 - \left(\frac{1}{N}\sum_{k=0}^{N-1} r_k\right)^2, \quad \text{(C1)}
$$

under the constraint that the increments $u_k \equiv r_k - r_{k-1}$ are

absolutely bounded by some constant $U > 0$ i.e., $|u_k| \leq U$ for $k > 0$.

The empirical variance $V(r)$ is invariant by the addition of an arbitrary constant C: let $r_k = y_k + C$, for all k, then $V(r) = V(y)$. We can thus assume with no loss of generality that $r_0 = 0$ (i.e., choose $C = -y_0$). Therefore, we have $r_k = \sum_{j=1}^{k} u_j$ for $k > 0$.

We want to maximize the convex function V in the set of feasible solution described by $\{r_k\}$ which is a polyhedron of $\mathbb{R}^N$. From a classical theorem of convex analysis (see [31], p. 187), we conclude that V reaches its maximum at one of the extreme points of this polyhedron. The extreme points are the points where the increments are either $u_k = +U$ or $u_k = -U$. There are 2^{N-1} extreme points and we need to identify the one which maximizes the convex function.

Let us rewrite the empirical variance $V(r)$ as a function of u_k. We leave the "autoterms" u_k^2 aside (for all extreme points, the autoterms are equal to U^2 independently of the sign of u_k. Their contribution is thus unimportant for the identification of the maximum) and concentrate on "cross terms" (i.e., terms in $u_j u_k$). A direct calculation leads to

$$V(r) = V_a + \sum_{j=1}^{N-2} \sum_{k=j+1}^{N-1} c_{jk} u_j u_k, \qquad (C2)$$

where $c_{jk} = 2j(N-k)/N^2$ and V_a is the contribution due to the autoterms.

Since all $c_{jk} > 0$, the maximum of V is reached when all u_k have the same signs, that is when u_k are all identically $+U$ or $-U$. Therefore, the empirical variance is maximum when $r_k = \pm kU$ and in this case $V(r) = U^2(N^2-1)/12$.

We recall that the distance between the smooth chirps is well approximated by the empirical variance of the phase discrepancy [see Eq. (14)]. We apply this result to the original maximization problem by setting $r_k \hat{=} \Delta_k$ and $U \hat{=} 2\pi \Delta_f t_s$ as given in Eq. (27).

APPENDIX D: BOUNDING δ AND ϵ OF A CC

The simplification of the statistic in Sec. V B is closely related to the orthogonality and length difference of the vectors $\mathbf{c} \equiv \{\cos\phi_k, k = 0, \cdots, N-1\}$ and $\mathbf{s} \equiv \{\sin\phi_k, k = 0, \cdots, N-1\}$ of $\mathbb{R}^N$.

Noting that their norms and scalar-product are, respectively, given by $n_c = \langle \mathbf{c}, \mathbf{c} \rangle$, $n_s = \langle \mathbf{s}, \mathbf{s} \rangle$ and $n_x = \langle \mathbf{c}, \mathbf{s} \rangle$ as defined in Eq. (7), the departure from "orthonormality" of $\mathbf{c}$ and $\mathbf{s}$ can be quantified by the two parameters

$$\delta = \frac{n_c - n_s}{n_c + n_s} \qquad \epsilon = \frac{2n_x}{n_c + n_s}. \qquad (D1)$$

The parameter δ measures the relative difference of the lengths of $\mathbf{c}$ and $\mathbf{s}$ while ϵ is related to the cosine of the angle between the two vectors.

When the vectors $\mathbf{c}$ and $\mathbf{s}$ are orthonormal i.e., orthogonal and of same lengths, both δ and ϵ are zero. By continuity, for nearly orthonormal vectors, δ and ϵ are then

expected to be small. Intuitively, this should be true for vectors with oscillating components like $\mathbf{c}$ and $\mathbf{s}$. Indeed, ϵ and δ can be rewritten in the form of oscillating sums, namely

$$\delta = \frac{1}{N} \sum_{k=0}^{N-1} \cos 2\phi_k \qquad \epsilon = \frac{1}{N} \sum_{k=0}^{N-1} \sin 2\phi_k. \qquad (D2)$$

The positive and negative contributions cancel in the summation, and thus leaves a small residual. In this appendix, we go beyond this intuitive rationale when the phase ϕ is a CC as defined in Eq. (11) and give a systematic investigation of the maximum value taken by δ and ϵ.

Equation (D2) motivates us to combine δ and ϵ is the following complex sum S

$$S \equiv \delta + i\epsilon = \frac{1}{N} \sum_{k=0}^{N-1} \exp i2\phi_k. \qquad (D3)$$

Bounding the modulus of S is equivalent to bounding δ and ϵ. Analytic number theory provides a large number of results concerning exponential sums like S, for improving upon the trivial bound $|S| \leq 1$. We use one of these, namely, the Kuzmin-Landau lemma, see [32] p. 7. We present a proof of this lemma pertaining to the present case where the phase ϕ is a CC.

The proof can be summarized as follows. A change of variables is introduced which allows us to put a bound on the modulus of S by a sum of the finite difference of complex variables. These new variables appear to be collinear in the complex plane. The sum of the modulus of their difference is thus equal to the distance between the extremes. The final bound on $|S|$ is then obtained by combining this property with the explicit expression of the phase of the CC, provided a constraint on the lower and higher frequencies reached by the CC.

Let us define for $1 \leq k \leq N-1$, the following variables

$$d_k \equiv 2(\phi_k - \phi_{k-1}) \qquad \zeta_k \equiv \frac{1}{1 - \exp(id_k)}. \qquad (D4)$$

We perform the above change of variables in the sum S using the relation

$$\exp(i2\phi_k) = [\exp(i2\phi_k) - \exp(i2\phi_{k+1})]\zeta_{k+1}, \qquad (D5)$$

and we get

$$NS = \zeta_1 \exp(i2\phi_0) + \sum_{k=1}^{N-2} (\zeta_{k+1} - \zeta_k) \exp(i2\phi_k) + (1 - \zeta_{N-1}) \exp(i2\phi_{N-1}). \qquad (D6)$$

By taking the modulus on both side, we obtain the following bound,

$$N|S| \leq |\zeta_1| + |1 - \zeta_{N-1}| + \sum_{j=0}^{N_t-1} \sum_{k=jb+1}^{(j+1)b-1} |\zeta_{k+1} - \zeta_k|$$

$$+ \sum_{j=1}^{N_t-2} |\zeta_{jb+1} - \zeta_{jb}|. \tag{D7}$$

where we split the sum in Eq. (D6) into smaller ones calculated over chirplet intervals i.e., $t_j \leq t_s k < t_{j+1}$ or equivalently $jb \leq k \leq (j+1)b - 1$ with $b = \delta_t/t_s$, the number of samples in a chirplet interval. In the last sum, we separate the terms corresponding to the transition between two consecutive chirplets.

We now obtain a bound on each term of the RHS of Eq. (D7), starting with the first sum. Equation (D4) can be rewritten as

$$\zeta_k = \frac{1}{2}[1 + i \cot(d_k/2)]. \tag{D8}$$

The variables ζ_k are all located on the line $\Re(\zeta) = 1/2$.

Within a chirplet interval, i.e. if $t_j \leq kt_s < t_{j+1}$, the phase difference is a linear function of k given by

$$d_k = 2\pi t_s[(2 - r)f_{m_j} + rf_{m_{j+1}}], \tag{D9}$$

where $r = (2t_{j,k} - t_s)/\delta_t$.

We assume that the node frequencies of the CC are constrained in the following bandwidth:

$$f_l \leq f_{m_j} \leq f_s/2 - f_l, \tag{D10}$$

where $f_l = f_s c/2$ and $0 < c < 1/2$. In other words, the CC cannot approach arbitrarily close to neither DC nor the Nyquist frequency.

Since $0 < r < 2$, we have $4\pi t_s f_{m_j} \leq d_k \leq 4\pi t_s f_{m_{j+1}}$ if $f_{m_j} \leq f_{m_{j+1}}$ (and the opposite in the other case) which implies that

$$0 < 2\pi c \leq d_k \leq 2\pi(1 - c) < 2\pi, \tag{D11}$$

for all k, hence $-\infty < \Im(\zeta_k) < +\infty$.

If $f_{m_j} \leq f_{m_{j+1}}$ (resp. $f_{m_j} \geq f_{m_{j+1}}$), the phase difference d_k and hence $\Im(\zeta_k)$, increases (resp. decreases) monotonically with k.

Since their imaginary parts are finite and monotonic, the variables ζ_k are associated to consecutive points on the line $\Re(\zeta) = 1/2$ of the complex plane. The sum of the lengths of the segments linking two nearby points is equal to the length between the extremes, thus

$$\sum_{k=jb+1}^{(j+1)b-1} |\zeta_{k+1} - \zeta_k| = |\zeta_{(j+1)b} - \zeta_{jb+1}|. \tag{D12}$$

Applying the mean value theorem to the function $g(x) = \cot(x/2)/2$, whose derivative is $\dot{g}(x) = 1/(4\sin^2(x/2))$ and using the constraint in Eq. (D11), we obtain the following

bound

$$|\zeta_{(j+1)b} - \zeta_{jb+1}| \leq \frac{|d_{(j+1)b} - d_{jb+1}|}{4\sin^2(\pi c)}. \tag{D13}$$

We carry on by bounding the numerator

$$|d_{(j+1)b} - d_{jb+1}| = 4\pi t_s|f_{m_{j+1}} - f_{m_j}| \leq 4\pi N'_r/N, \tag{D14}$$

and denominator with $2c \leq \sin(\pi c)$ (this is valid for $0 \leq c \leq 1/2$) and by summing over all j to finally obtain the bound on first summation term in Eq. (D7),

$$\sum_{j=0}^{N_t-1} \sum_{k=jb}^{(j+1)b-1} |\zeta_{k+1} - \zeta_k| \leq \frac{\pi N'_r N_t}{4Nc^2}. \tag{D15}$$

The second summation coming from the boundary points of the chirplet intervals can be bounded in a similar way, considering that

$$|\zeta_{jb+1} - \zeta_{jb}| \leq \frac{|d_{jb+1} - d_{jb}|}{4\sin^2(\pi c)}, \tag{D16}$$

and combining with

$$|d_{jb+1} - d_{jb}| = 2\pi t_s^2|f_{m_{j+1}} - f_{m_{j-1}}|/\delta_t \leq \frac{4\pi t_s N'_r}{N\delta_t}, \tag{D17}$$

we get the result

$$\sum_{j=1}^{N_t-2} |\zeta_{jb+1} - \zeta_{jb}| \leq \frac{\pi t_s N'_r N_t}{4Nc^2\delta_t}. \tag{D18}$$

Finally, from Eq. (D8), we have the following inequalities

$$|\zeta_k| = \frac{1}{2|\sin(d_k/2)|} \leq \frac{1}{2\sin(\pi c)} \leq \frac{1}{4c}, \tag{D19}$$

which, when applied with $k = 1$ and $k = N - 1$, set an upper limit to the remaining terms in the RHS of Eq. (D7), noting that $|1 - \zeta_{N-1}| = |\zeta_{N-1}|$.

Combining this result with Eqs. (D15) and (D18), we get

$$|S| \leq \frac{1}{2Nc} + \frac{\pi N'_r N_t}{4c^2 N^2}(1 + 1/b). \tag{D20}$$

The number of samples in a chirplet interval being an integer $b \geq 1$, and selecting the dominating contribution, we conclude that $|S| \lesssim \eta$ with

$$\eta = \frac{\pi N'_r N_t}{2c^2 N^2}. \tag{D21}$$

This bound is obtained from a worst case estimate. Generally, δ and ϵ are smaller than this value. With the choice of a small c, a more realistic estimate rather than a strict bound can be obtained replacing the inequality $2c \leq \sin(\pi c)$ by the first order Taylor approximation $\pi c \sim$

$\sin(\pi c)$ in the proof above, yielding the following estimate

$$\eta = \frac{N'_r N_t}{\pi c^2 N^2}. \tag{D22}$$

Summarizing, we obtained an upper bound η on $|S|$ by restricting the frequency of the CC in a bandwidth defined by c. We rather use the reciprocal i.e., we get the limits of the frequency bandwidth from an acceptable value for η. If we assume that $N'_r \approx 4(N'/N_t)(N_f/(2N))$ as given by the

regularity condition, the frequency bandwidth is $[f_l, f_s/2 - f_l]$ with

$$f_l = \frac{f_s c}{2} \approx 2.5\sqrt{N'}\delta_f\left(\frac{N_f}{N}\right)^{3/2}\left(\frac{0.1}{\eta}\right)^{1/2}, \tag{D23}$$

where the leading constant is obtained from $\sqrt{20/\pi} \approx 2.5$. We use this result in Sec. V B 1.

[1] (GEO600), www.geo600.uni-hannover.de; (LIGO), http://www.ligo.org; (Virgo), http://www.virgo.infn.it; (TAMA), http://tamago.mtk.nao.ac.jp.

[2] L. Blanchet, Living Rev. Relativity **5**, 3 (2002).

[3] K. D. Kokkotas and B. Schmidt, Living Rev. Relativity **2**, 2 (1999).

[4] B. Schutz, gr-qc/0410121.

[5] E. E. Flanagan and S. A. Hughes, Phys. Rev. D **57**, 4535 (1998).

[6] M. H. P. M. van Putten *et al.*, Phys. Rev. D **69**, 044007 (2004).

[7] V. Ferrari, G. Miniutti, and J. A. Pons, Classical Quantum Gravity **20**, S841 (2003).

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

[9] T. Pradier *et al.*, Phys. Rev. D **63**, 042002 (2001).

[10] E. Chassande-Mottin, Phys. Rev. D **67**, 102001 (2003).

[11] A. Vicerè, Phys. Rev. D **66**, 062002 (2002).

[12] W. Anderson, P. Brady, J. Creighton, and E. Flanagan, Phys. Rev. D **63**, 042003 (2001).

[13] W. Anderson and R. Balasubramanian, Phys. Rev. D **60**, 102001 (1999).

[14] J. Sylvestre, Phys. Rev. D **66**, 102004 (2002).

[15] S. Kay, *Fundamentals of Statistical Signal Processing: Detection theory* (Prentice Hall, Englewood Cliffs, NJ, 1998).

[16] B. J. Owen and B. S. Sathyaprakash, Phys. Rev. D **60**, 022002 (1999),.

[17] B. S. Sathyaprakash and S. V. Dhurandhar, Phys. Rev. D **44**, 3819 (1991).

[18] B. J. Owen, Phys. Rev. D **53**, 6749 (1996).

[19] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic Press, New York, 1980).

[20] E. Arias-Castro, D. Donoho, and X. Huo, IEEE Trans. Inf. Theory **51**, 2402 (2005).

[21] E. Chassande-Mottin and A. Pai, IEEE Signal Process. Lett. **12**, 508 (2005).

[22] P. Flandrin, *Time-frequency/Time-scale Analysis* (Academic Press, New York, 1999).

[23] D. P. Bertsekas, *Dynamic Programming and Optimal Control* (Athena Scientific, Belmont, MA, 2000), 2nd ed., Vol. 1.

[24] LSC Algorithm Library, http://www.lsc-group.phys.uwm.edu/daswg/projects/lal.html.

[25] E. Chassande-Mottin and P. Flandrin, Applied and Computational Harmonic Analysis **6**, 252 (1999).

[26] M. Morvidone and B. Torrésani, International Journal on Wavelet and Multiresolution Information Processing **1**, 19 (2003).

[27] B. Krishnan, A. Sintes, M. A. Papa, B. F. Schutz, S. Frasca, and C. Palomba, Phys. Rev. D **70**, 082001 (2004).

[28] P. Brady and T. Creighton, Phys. Rev. D **61**, 082001 (2000).

[29] N. L. Johnson and S. Kotz, *Distribution in Statistics– Continuous Univariate Distributions* (John Wiley & Sons, New York, 1970).

[30] R. Balasubramanian, B. S. Sathyaprakash, and S. V. Dhurandhar, Phys. Rev. D **53**, 3033 (1996).

[31] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization* (Athena Scientific, Belmont, MA, 2003).

[32] S. W. Graham and G. Kolesnik, *Van der Corput's Method of Exponential Sums* (Cambridge Univ. Press, Cambridge, 1991).