PHYSICAL REVIEW D **85,** 062003 (2012)

Loosely coherent searches for sets of well-modeled signals

V. Dergachev

LIGO Laboratory, California Institute of Technology, MS 100-36, Pasadena, California 91125, USA

(Received 17 October 2011; published 19 March 2012)

We introduce a high-performance implementation of a loosely coherent statistic sensitive to signals spanning a finite-dimensional manifold in parameter space. Results from full-scale simulations on Gaussian noise are discussed, as well as implications for future searches for continuous gravitational waves and related searches for slowly modulated signals, such as emitted by radio pulsars. We demonstrate an improvement of more than an order of magnitude in analysis speed over previously available algorithms. As searches for continuous gravitational waves are computationally limited, the large speedup results in gain in sensitivity.

DOI: 10.1103/PhysRevD.85.062003 PACS numbers: 07.05.Kf, 04.80.Nn, 95.55.Ym

I. INTRODUCTION

Loosely coherent algorithms [1] detect families of noise-dominated signals. Their development was prompted by the challenge of conducting a blind search for continuous gravitational waves. These signals are expected to be produced by rotating neutron stars with large mass quadrupole moment. Since the stars rotate in vacuum their frequency is expected to be stable (similar to radio pulsars) with slow modulation from energy loss, possible companion objects, Doppler shifts from changing detector velocity, and changes in detector orientation.

Traditional coherent approach requires construction of sufficiently dense bank of waveforms, which is used in a matched filter applied to acquired data. Because of large computational requirements this approach is impractical to pursue with the several years of data accumulated by LIGO detectors [2], while still covering a significant portion of the sky.

An alternative common approach is to use a semicoherent algorithm which breaks up data into short segments, each of which is analyzed with a coherent algorithm and the results are summed. The decrease in the coherent time base leads to great reduction in required computational power, at a cost of reduced sensitivity.

A loosely coherent algorithm straddles the middle of these extremes by considering families of signals with a limited phase drift relative to an ideal coherent template. For example, a signal family could consist of a set of waveforms with nearby parameters.

As we are still waiting for the first detection, we cannot rely on a natural source to verify correctness of the detector and search pipelines. Our algorithms must be resistant to possible imperfections of the detector, to faults in understanding of gravitation, or even to bugs in the search programs. It also helps to be sensitive to a wide family of signals, in case the loudest source is not a perfect sine wave, which can result, for example, from a companion object.

Consideration of a set of signals, rather than a particular template, leads to improvements in computational speed and controlled way to introduce robustness to deviations from ideal signal model.

The first implementation of a loosely coherent search [1] was designed for signals with a large amount of phase deviation over 30-minute intervals. This provided the gain in sensitivity needed for follow-up of outliers seen in the full data set of the LIGO detector's fifth science run S5 [3], while preserving the robustness of the underlying, semicoherent PowerFlux algorithm [4–6].

In this paper we explore the other end of the spectrum—an algorithm sensitive to coherent signals described by a small number of parameters, such as frequency or sky position. A number of coherent codes have been developed previously, in particular [7–16]. What is different in our approach is that, unlike previous algorithms, our sky templates are ''thick,'' sweeping small patches of parameter spaces. In particular, individual signals taken from the middle of nearby patches do not have a high degree of overlap. This property, together with careful attention to implementation particulars, provides for a very high-performance coherent code.

The implementation, discussed below, makes use of several new techniques:

(i) A local perturbative expansion of signal waveforms. A pair of complex valued parameters which conveniently describe polarization of incoming signals and the manifold of signals of unit strain.

(ii) An efficient method to compute $\mathcal{B}$-statistic [14]. A way to use a simplified and computationally efficient statistic to determine whether more complicated physically meaningful statistics need to be computed for a particular data sample.

(iii) Unlike previous implementations of $\widetilde{\mathcal{F}}$-statistic [7] this code derives upper limits as well as SNR, removing the need to run expensive Monte Carlo simulations to estimate upper limits.

(iv) *Pseudo*-convolutions are used as a core part of computational engine.

(v) An analysis of code efficiency uses actual running time with performance reported in CPU cycles per template.

II. SIMPLIFIED ALGORITHM

Suppose that we are interested in a family of signals described by the formula

$$s(t;a) = Ae^{2\pi i\nu[t+\Xi(t,a)]}, \tag{1}$$

where ν is the signal frequency and $\Xi(t, a)$ has a smooth dependence on time t and multidimensional parameter set a.

Our input data $z(t) = s(t; a) + \xi(t)$ consists of one signal from this family and, ideally, uncorrelated, but nonstationary Gaussian noise of variance $\sigma^2(t)$:

$$\langle \xi(t)\xi(t')\rangle = \sigma^2(t)\delta(t - t'). \tag{2}$$

In practice, this data is usually derived by partitioning input data into short segments, taking a Fourier transform of each segment, and then picking a single bin in each short Fourier transform (SFT). If the segments are non-overlapping this usually implies uncorrelated bins. Otherwise, a small amount of correlation between nearby bins is present, which we neglect in this paper. An alternative way to obtain $\xi(t)$ is to use a heterodyning filter in which case the correlation length will depend on the width of the filter and discretization step used during computation.

If we knew the frequency ν and parameter set a, we could form a matched filter that would return the amplitude of our signal:

$$A(\nu, a) = \frac{1}{\mathfrak{W}}\int_{t_0}^{t_1}\frac{z(t)}{\sigma^2(t)}e^{-2\pi i\nu[t+\Xi(t,a)]}dt. \tag{3}$$

Here $\mathfrak{W}$ is the total weight:

$$\mathfrak{W} = \int_{t_0}^{t_1}\frac{1}{\sigma^2(t)}dt.$$

If the signal parameters are not known, one can construct a bank of waveforms $s(t; a)$ and evaluate the integral separately for each template. This is, of course, computationally expensive.

One way to gain a large speed-up is to introduce a new time variable t' that straightens out our signal into a sine wave:

$$t' = t + \Xi(t, a). \tag{4}$$

One then resamples [15] the input data $z(t)$ to be equally spaced in the new variable t' and uses a Fourier transform to compute amplitudes for a range of possible signal frequencies. It has also been proposed by B. F. Schutz that a change to nearby value a' can be accomplished with a kernel [17]. As far as we know this method of ''stepping around in the sky'' (named so because sky position is an important source of nonlinearity) does not yet have an implementation.

A simpler approach that combines the best features of both resampling and stepping around in the sky is to consider the following function of the input data:

$$F(\lambda;\nu_0,a_0) = \frac{1}{\mathfrak{W}}\int_{t_0}^{t_1}\frac{z(t)}{\sigma^2(t)}e^{-2\pi i\nu_0[t+\Xi(t,a_0)]}e^{-2\pi i\lambda t}dt, \tag{5}$$

which is easily computed with a fast Fourier transform. For $\lambda = 0$ it returns an amplitude estimate of the signal with parameters (ν_0, a_0). The values of F for $\lambda \neq 0$ carry slightly distorted information on nearby templates, which can be used to compute estimates of their signal amplitude with a convolution:

$$\begin{aligned}
A(\nu,a) &= \frac{1}{\mathfrak{W}}\int_{t_0}^{t_1} z(t)e^{-2\pi i\nu(t+\Xi(t,a))}\frac{1}{\sigma^2(t)}dt \\
&= \frac{1}{\mathfrak{W}}\int_{t_0}^{t_1} z(t)e^{-2\pi i\nu_0(t+\Xi(t,a_0))}\frac{1}{\sigma^2(t)}\cdot e^{-2\pi i(\nu-\nu_0)t-2\pi i(\nu-\nu_0)\Xi(t,a_0)-2\pi i\nu(\Xi(t,a)-\Xi(t,a_0))}dt \\
&= \int F(\nu-\nu_0-\mu;\nu_0,a_0)\cdot\int_{t_0}^{t_1} e^{-2\pi i\mu t-2\pi i(\nu-\nu_0)\Xi(t,a_0)-2\pi i\nu(\Xi(t,a)-\Xi(t,a_0))}dtd\mu.
\end{aligned} \tag{6}$$

The reader will notice that the last term is not quite the ordinary convolution—one of the convolved terms has a (slow) dependence on the convolution parameter. We call this a ''pseudo-convolution'' operator. We distinguish this case from the more general notion of integral operator, because in practical computation we do not have to update the slowly changing convolution with every data sample and the computational requirements of the operator are equivalent to the computational requirements of plain convolution.

In our case the pseudo-convolutions are of the form

$$A(\nu,a) = \int F(\nu-\nu_0-\mu;\nu_0,a_0)\times\int_{t_0}^{t_1} e^{-2\pi i\mu t-U(t,\nu,a)}dtd\mu, \tag{7}$$

where $U(t, \nu, a)$ is a *phase mismatch* function describing the difference in phase evolution between nearby templates. For smooth $U(t, \nu, a)$ the convolution operator is close to δ-function. In practical computation, using a

discrete Fourier transform, this means that our convolution can be approximated with an FIR filter that has small number of terms.

It is crucial to control the number of terms in the convolutions. There are two ways to achieve that, aside from simply using small increments in the parameters with a corresponding increase in the number of templates.

First, we can subtract a linear term from the argument of the exponent so that it has the same value at both ends of the segment $[t_0, t_1]$. The linear term is analogous to a Doppler shift correction and results in relabeling of frequency parameter ν.

Secondly, pseudo-convolutions have small or null commutators. One can then apply methods of linear algebra to change from initial set of operators (usually corresponding to individual parameters) to a set with progressively fewer convolution terms. The operators with the smallest number of terms are used in the innermost computational loop thus determining overall performance of the code.

We implement these techniques by representing $U(t, \nu, a)$ as a sum of a linear term and Fourier series with coefficients linear in ν:

$$U(t, \nu, a) \simeq \frac{U(t_1, \nu, a) - U(t_0, \nu, a)}{t_1 - t_0}(t - t_0) + \sum_{k=-\infty}^{\infty} (u_k^0(a) + u_k^1(a)\nu)e^{2\pi i k t}. \quad (8)$$

For many parameters we use (such as spin-down) the dependence on ν is linear and the equality is exact.

The linear term is folded into the frequency variable which thus acquires a shift of a fraction of frequency bin. The periodic terms are integrated with the help of the Jacobi-Anger identity:

$$e^{iz\cos\theta} = \sum_{n=-\infty}^{\infty} i^n J_n(z) e^{in\theta}. \quad (9)$$

Only a small number of terms are usually needed and recomputation is done rarely. The simulations presented in Sec. IV were carried out with innermost loops that used convolutions with only 11 terms—a number chosen to take advantage of vectorized arithmetic on modern CPUs.

The set of pseudo-convolution operators can then be transformed into a new basis by minimizing coefficients in the series. We then place the pseudo-convolution with the smallest number of terms into the innermost loop, which will dominate the scalability of the code.

III. IMPLEMENTATION DETAILS

While a fast engine to compute coherent sums is essential for our search code, it is only part of a whole. In particular, after computing coherent sums one needs to derive statistics such as maximum SNR or upper limit which can be expensive to compute. For example, if one were to use a rank-based method (which is nicely robust) to compute statistics on N bins, it would require sorting the data which has a scaling of $N \log N$—same as a fast Fourier transform used in the resampling method, and much slower than a convolution.

It is usually not practical to analyze the entire band of interest in one go, but rather one splits it into frequency bands of 1 Hz or smaller. The amount of loaded data can be greatly reduced by precomputing short discrete Fourier transforms of duration commensurate with the region of interest. It is convenient to have the SFT length be short enough that the signal frequency can be assumed to be stationary.

A. Polarization analysis

Continuous gravitational waves have a more complicated form than is given by Eq. (1)—there are two polarizations with detected strengths that vary with the orientation of the detector.

The following analysis is similar to one found in [7,18,19]; we prefer, however, to reduce the four real parameters to two complex parameters that have a symmetric role. We also derive a convenient equation for surfaces of constant h_0.

We start by assuming that our signal consists of two polarizations:

$$h'_+ = A_+ \cos(\omega t + \phi) \qquad h'_\times = A_\times \sin(\omega t + \phi). \quad (10)$$

A generic pulsar signal can be represented as $A_+ = h_0(1 + \cos^2(\iota))/2$, $A_\times = h_0 \cos(\iota)$, with $h_0 = A_+ + \sqrt{A_+^2 - A_\times^2}$ and $\cos(\iota) = A_\times/(A_+ + \sqrt{A_+^2 - A_\times^2})$

We will assume that demodulation is performed for a fixed frame of plus and cross polarizations rotated at an angle β. In this coordinate system we have:

$$\begin{aligned} h_+ &= A_+ \cos(\omega t + \phi)\cos(\epsilon) - A_\times \sin(\omega t + \phi)\sin(\epsilon) \\ h_\times &= A_+ \cos(\omega t + \phi)\sin(\epsilon) + A_\times \sin(\omega t + \phi)\cos(\epsilon), \end{aligned} \quad (11)$$

where we introduced $\epsilon = 2(\psi - \beta)$, with ψ denoting the orientation angle of the source pulsar [5–7].

The signal amplitude in SFT bin corresponding to frequency ω is then

$$\begin{aligned} z &= \int (F_+ h_+ + F_\times h_\times) e^{-i\omega t} dt \\ &= \tfrac{1}{2} e^{i\phi}(F_+(A_+ \cos(\epsilon) + iA_\times \sin(\epsilon)) \\ &\quad + F_\times(A_+ \sin(\epsilon) - iA_\times \cos(\epsilon))) \\ &= F_+ w_1 + F_\times w_2, \end{aligned} \quad (12)$$

where F_+ and $F_\times$ denote amplitude response of the detector to plus and cross polarizations and we introduced complex amplitude parameters

$$w_1 = \tfrac{1}{2}e^{i\phi}[A_+ \cos(\epsilon) + iA_\times \sin(\epsilon)]$$
$$w_2 = \tfrac{1}{2}e^{i\phi}[A_+ \sin(\epsilon) - iA_\times \cos(\epsilon)]. \quad (13)$$

The complex amplitude parameters w_1 and w_2 are algebraically symmetric:

$$b = |w_1|^2 + |w_2|^2 = \tfrac{1}{4}(A_+^2 + A_\times^2)$$
$$= \tfrac{1}{16}h_0^2(1 + 6\cos^2(\iota) + \cos^4(\iota))$$
$$a = \mathrm{Im}(w_1\bar{w}_2) = \tfrac{1}{4}A_+A_\times = \tfrac{1}{8}h_0^2(1 + \cos^2(\iota))\cos(\iota) \quad (14)$$

and are otherwise unconstrained. They have a simple relation to more familiar real amplitude parameters $\mathcal{A}^\mu$ [20]:

$$w_1 = \mathcal{A}^1 - i\mathcal{A}^3 \qquad w_2 = \mathcal{A}^2 - i\mathcal{A}^4. \quad (15)$$

One easily finds the following equation of constant h_0:

$$\sqrt{|w_1 + iw_2|} + \sqrt{|w_1 - iw_2|} = \sqrt{h_0}, \quad (16)$$

the solutions of which form a singular surface enclosing a nonconvex solid. This complicated form is responsible for differences between PowerFlux style upper limits, which are always limited by sensitivity to linearly polarized signals, and SNR statistics, the outliers of which can have arbitrary polarization. A related issue is the difference between the $\mathcal{F}$-statistic and the $\mathcal{B}$-statistic [14].

It is easy to compute generators for rotations in ϕ and ϵ:

$$\frac{\partial}{\partial\phi}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = i\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \quad (17)$$

$$\frac{\partial}{\partial\epsilon}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix}. \quad (18)$$

This shows that the surface of $h_0 = 1$ is obtained by revolving the parabola

$$w_1' = \tfrac{1}{4}[1 + \cos^2(\iota)] \qquad w_2' = -\tfrac{i}{2}\cos(\iota) \quad (19)$$

along ϕ and ϵ.

B. Coherent sum

Assume we have data for many SFTs $\{z_i\}_{i=1}^N$. We use index i to denote SFT number (which thus corresponds to time). We assume that SFT bins are large enough so that the signal is contained in one bin per SFT, which is retained as z_i:

$$z_i = [F_+(t_i)w_1 + F_\times(t_i)w_2]e^{i\Phi(t_i)} + \xi_i. \quad (20)$$

In a coherent analysis we construct a weighted sum

$$Z = \sum_{i=1}^N \alpha_i \frac{z_i e^{-i\Phi(t_i)}}{F_+(t_i)w_1' + F_\times(t_i)w_2'} \quad (21)$$

which estimates signal amplitude. Here $\Phi(t_i)$ describes some assumed phase evolution due to changes in the source or detector, α_i are weights satisfying $\sum_{i=1}^N \alpha_i = 1$, and w_1' and w_2' are computed for polarization and phase of our signal, but assuming $h_0 = 1$, i.e. they satisfy

$$\sqrt{|w_1' + iw_2'|} + \sqrt{|w_1' - iw_2'|} = 1. \quad (22)$$

There are many ways to compute "optimal" weights α_i, in particular [7,14]. Here we use the variance of Z as the optimality measure. This choice allows analytic treatment (as opposed to L_1 or L_∞ norm) and ease of implementation, while minimizing the established upper limit in the common case of signal absence. This also leads to optimum average (over different noise realizations) signal SNR. A reader might wonder why do we not optimize $|Z|^2/\mathrm{Var}(Z)$ directly ? First of all, it is difficult to implement with required efficiency. Second, it has a disadvantage of fitting the received noise, which leads to data-mining effects, in particular, increased errors in parameter estimation and elevated SNR floor for background data.

Assuming ξ_i are independent Gaussian variables with zero mean, we compute:

$$\mathrm{Var}\,(Z) = \sum_{i=1}^N \alpha_i^2 \frac{\mathrm{Var}(\xi_i)}{|F_+(t_i)w_1' + F_\times(t_i)w_2'|^2}. \quad (23)$$

One easily finds that $\mathrm{Var}(Z)$ is minimized for

$$\alpha_i = \frac{1}{\mathfrak{N}(w_1', w_2')}\frac{|F_+(t_i)w_1' + F_\times(t_i)w_2'|^2}{\mathrm{Var}(\xi_i)}, \quad (24)$$

where $\mathfrak{N}$ is the normalization weight:

$$\mathfrak{N}\,(w_1', w_2') = \sum_{i=1}^N \frac{|F_+(t_i)w_1' + F_\times(t_i)w_2'|^2}{\mathrm{Var}(\xi_i)}. \quad (25)$$

Substituting the optimal coefficients α_i, we compute:

$$Z(w_1', w_2') = \frac{1}{\mathfrak{N}(w_1', w_2')}\sum_{i=1}^N z_i e^{-i\Phi(t_i)}\frac{F_+(t_i)\bar{w}_1' + F_\times(t_i)\bar{w}_2'}{\mathrm{Var}(\xi_i)} \quad (26)$$

and

$$\mathrm{Var}\,[Z(w_1', w_2')] = \frac{1}{\mathfrak{N}(w_1', w_2')}. \quad (27)$$

We see that the total weight $\mathfrak{N}(w_1', w_2')$ can be interpreted as a measure of the amount of data used to compute $Z(w_1', w_2')$, as in the case of stationary input data it is proportional to the number of independent SFTs.

We now note that $\mathfrak{N}(w_1', w_2')$ depends quadratically on coefficients w_i' and on the estimate of variance of the data and can thus be computed once for all templates with similar detector response F_+ and $F_\times$. The unnormalized coherent sum $\mathfrak{N}Z$ is linear in w_i', and these coefficients are exactly the kind of sum that we learned to compute in Sec. II.

Our computation then results in the following representation of Z:

$$Z(n) = \frac{X_1(n)\bar{w}'_1 + X_2(n)\bar{w}'_2}{Y_{11}|w'_1|^2 + 2Y_{12}\mathrm{Re}(w'_1\bar{w}'_2) + Y_{22}|w'_2|^2}. \quad (28)$$

Here Y_{11}, Y_{12}, and Y_{22} are coefficients that determine the total weight of the coherent sum and $X_1(n)$ and $X_2(n)$ are two arrays (in frequency bin n) of coherent sums:

$$Y_{11} = \sum_{i=1}^{N} \frac{F_+(t_i)^2}{\mathrm{Var}(\xi_i)} \qquad Y_{12} = \sum_{i=1}^{N} \frac{F_+(t_i)F_\times(t_i)}{\mathrm{Var}(\xi_i)}$$

$$Y_{22} = \sum_{i=1}^{N} \frac{F_\times(t_i)^2}{\mathrm{Var}(\xi_i)}$$

$$X_1 = \sum_{i=1}^{N} z_i e^{-i\Phi(t_i)} \frac{F_+(t_i)}{\mathrm{Var}(\xi_i)} \qquad X_2 = \sum_{i=1}^{N} z_i e^{-i\Phi(t_i)} \frac{F_\times(t_i)}{\mathrm{Var}(\xi_i)}. \quad (29)$$

For convenience, we tabulate a few useful expressions using these coefficients:

$$\mathfrak{A}(w'_1, w'_2) = Y_{11}|w'_1|^2 + 2Y_{12}\mathrm{Re}(w'_1\bar{w}'_2) + Y_{22}|w'_2|^2$$

$$Z(w'_1, w'_2) = \frac{X_1\bar{w}'_1 + X_2\bar{w}'_2}{\mathfrak{A}(w'_1, w'_2)}$$

$$\mathrm{SNR}(w'_1, w'_2) = \frac{|X_1\bar{w}'_1 + X_2\bar{w}'_2|^2}{\mathfrak{A}(w'_1, w'_2)}$$

$$\mathrm{SNR}_R(w'_1, w'_2) = \frac{(\mathrm{Re}(X_1\bar{w}'_1 + X_2\bar{w}'_2))^2}{\mathfrak{A}(w'_1, w'_2)} \quad (30)$$

Here signal-to-noise ratio $\mathrm{SNR}(w'_1, w'_2)$ has been defined in power, rather than amplitude. This eases the analysis (no square roots) and has the same scaling for large signals as $\mathcal{F}$-statistic or PowerFlux SNR.

C. Efficient computation of coherent sum statistics

We now need to reduce the data to the SNR or some other statistic of the loudest outlier. The simplest method, employed by PowerFlux [4–6] and the large-δ loosely coherent search [1] is to scan different values of w_1 and w_2 looking for a maximum. The elements of our coherent sums are computed, however, with ~22 complex multiplications for each frequency bin, and even a modest grid of polarization parameters dominates computation.

An approach taken in [7] is to analytically maximize SNR over w_1 and w_2. A new $\mathfrak{B}$-statistic was introduced in [14] that was shown to have a better physically motivated prior. We describe an efficient method of computing it in Appendix .

There is an easy and elegant way to compute all of these statistics with minimal cost.

First we note, that our statistics are monotonic in λ for a signal family $\lambda s(t; a)$; they just disagree as to which signal parameters a are given preference. Once the signal strength is fixed (in any suitable statistic), picking a single member of each family, the rest of the parameters form a bounded manifold—and other statistics achieve a maximum and minimum value on it. We can thus infer an estimate of another statistic from knowing the maximum of some convenient, easy-to-compute measure of signal strength.

TABLE I. Statistics functions. The variables w'_1 and w'_2 are constrained by Eq. (22).

Statistic	Formula
Raw power	$\|X_1\|^2 + \|X_2\|^2$
Adjusted power	$Y_{22}\|X_1\|^2 + Y_{11}\|X_2\|^2$
SNR	$\max_{w'_1,w'_2} \frac{\|X_1\bar{w}'_1+X_2\bar{w}'_2\|^2}{\mathfrak{A}(w'_1,w'_2)}$
Upper limit	$\max_{w'_1,w'_2} \sqrt{\frac{\|X_1\bar{w}'_1+X_2\bar{w}'_2\|^2}{\mathfrak{A}(w'_1,w'_2)^2} + 2\frac{\|X_1\bar{w}'_1+X_2\bar{w}'_2\|}{\mathfrak{A}(w'_1,w'_2)^{3/2}} + \frac{\sigma-1}{\mathfrak{A}(w'_1,w'_2)}}$
$\mathfrak{F}$-stat	$\frac{Y_{22}\|X_1\|^2-2Y_{12}\mathrm{Re}(X_1\bar{X}_2)+Y_{11}\|X_2\|^2}{Y_{11}Y_{22}-Y_{12}^2}$
$\mathfrak{B}$-stat	$\int \mathbf{dw}' \frac{1}{\mathfrak{A}(w'_1,w'_2)} \Theta(\frac{\|X_1\bar{w}'_1+X_2\bar{w}'_2\|^2}{2\mathfrak{A}(w'_1,w'_2)})$

As a toy example, assume that our statistics vary by at most a factor of 4 for signals of the same power $|X_1|^2 + |X_2|^2$ and that our signal array consists of ≈ 300 000 complex Gaussian numbers with mean 0 and variance 1. Then on average the maximum power in this array is 12. But 95% of these samples have power below 3 = 12/4. Thus, to find the maximum of a more complicated statistics we only need to examine 5% of the samples, providing a factor

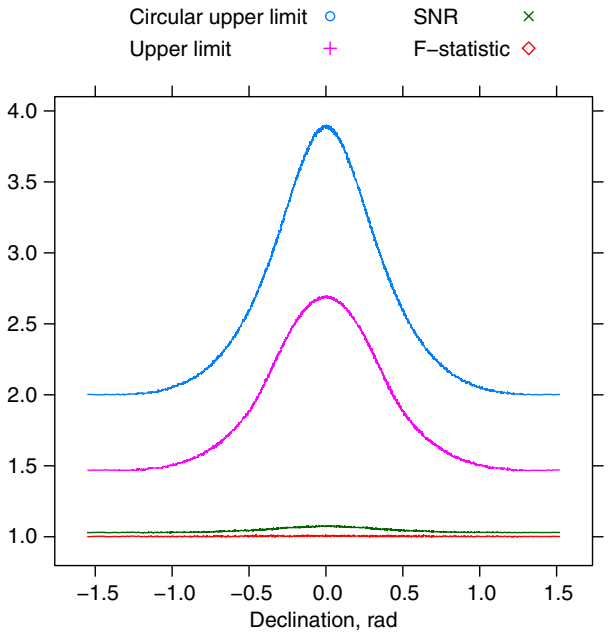

FIG. 1 (color online). Maximum variance of statistics for signals of constant adjusted power. The numbers come from simulations using Gaussian noise and describe the ratio between the maximum and minimum statistic values for a signal of constant adjusted power. Top curve—upper limit assuming circular polarization. Next curve below is upper limit statistic, followed by SNR and $\mathfrak{F}$-statistic (color online).

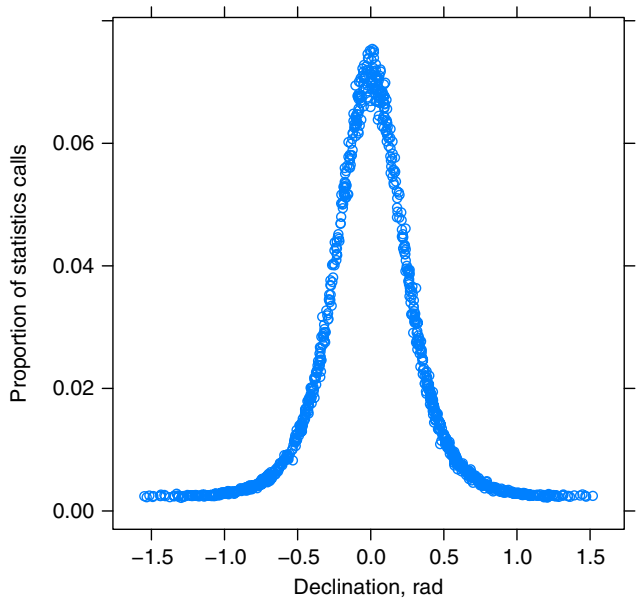

FIG. 2 (color online). Fraction of templates resulting in statistic calls. The underlying data was pure Gaussian noise (color online).

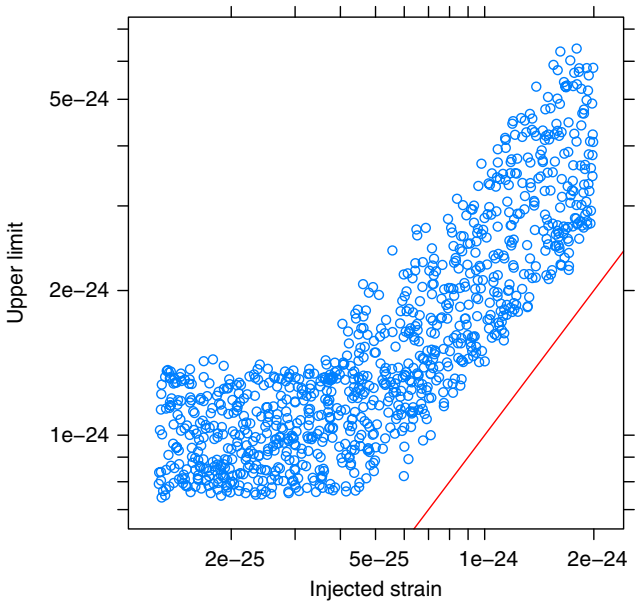

FIG. 3 (color online). Upper limit versus injected strain (color online).

of 20 speed up. If a signal is present our maximum is even higher excluding a larger amount of data.

In a practical implementation it is better to use adjusted power (see Table I), which results in a less-than-4 worst-case scale factor for upper limit statistics (evaluated for strong signals) and close to unity factors for plain SNR and $\mathfrak{F}$-statistics as can be seen on Fig. 1. The strong dependence on declination is due to changes in received power with daily rotation of Earth and nonuniform antenna pattern of the detector. The fraction of templates for which we computed the upper limit statistic during a Gaussian noise simulation run (discussed in more detail in the following section) is shown in Fig. 2.

IV. PERFORMANCE AND VALIDATION

An initial implementation of the ideas discussed above has been completed. It has a number of simplifications compared to an eventual production program—the input data is assumed contiguous, one spin-down is analyzed at a time, and there is no support for higher-order source frequency evolution parameters.

We have performed Monte-Carlo runs of 1000 injections each into Gaussian data spanning 4×10^6 s (approximately 1.5 months), assuming a detector located at LIGO Hanford observatory. The injection sky locations and source orientation were uniformly distributed. The spin-down parameter was set to be zero. In the first run, we injected signals of various strength (Fig. 3) to test signal detection and upper limit estimation. The second run had identical parameters and noise distribution, but the signal strengths were set to 0. This provided upper limits on pure noise alone (for relative comparison) as well as timing of worst-case computational performance, as the presence of strong signals makes computation of statistics faster. The injection frequency was uniformly distributed in ± 0.084 Hz interval centered on 400 Hz.

Each separate injection run included a search over a 1-arcminute disk around a nominal sky location that was obtained from rounding the true injection locations; hence the actual injection locations were uniformly distributed in relation to the sampled grid.

Each injection is analyzed independently, just as it would happen in a run using real signals. The upper limits are established using formulas given in Table I. As injected power rises above background the upper limits increase to be above injected signal values (Fig. 3). The gap between the reconstructed points and the diagonal line (red) marking injected values is due to a conservative correction for Hann-windowed input data.

When signals rise above background their frequencies are well localized (Fig. 4). We use this localization as a criterion for detection: a signal is considered to be found if its frequency is within 1×10^{-5} Hz of true value. This corresponds to false alarm ratio of $\approx 6 \times 10^{-5}$.

Figure 5 compares the efficiencies of various detection statistics. The SNR and $\mathfrak{F}$-statistics are mathematically equivalent; the only difference is that the SNR is computed by iterating over a grid of parameters w_1' and w_2', while the $\mathfrak{F}$-statistic has a much more efficient closed form. The slower SNR algorithm was used as a bridge to an implementation of the $\mathfrak{B}$-statistic, and will also be useful for future implementation of loosely coherent statistics.

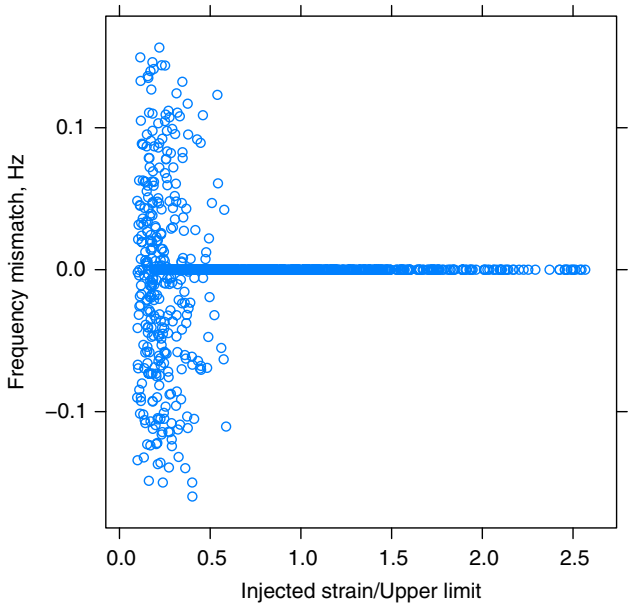

FIG. 4 (color online). Frequency reconstruction (color online).

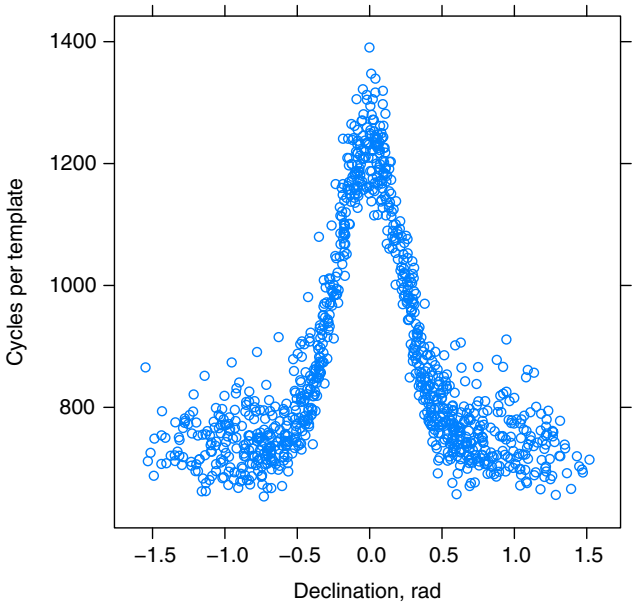

FIG. 6 (color online). CPU cycles spent per template as a function of declination of the injected signal (color online).

As seen from the plot we observe no difference in performance between $\mathfrak{F}$-statistic and $\mathfrak{B}$-statistic. We believe this is due to two factors:

(i) First, even for pure noise, maximizing over 2×10^6 bins results in high SNR values where the difference between $\mathfrak{F}$-statistic and $\mathfrak{B}$-statistic is smaller.

(ii) Second, the area searched is small which significantly reduces influence of the weight term $\mathfrak{A}(w_1', w_2')$ that appears in the $\mathfrak{B}$-statistic It might be possible to take advantage of the improved performance of the $\mathfrak{B}$-statistic at low SNR by analyzing multidetector data with consistency cuts to bring down maximum SNR.

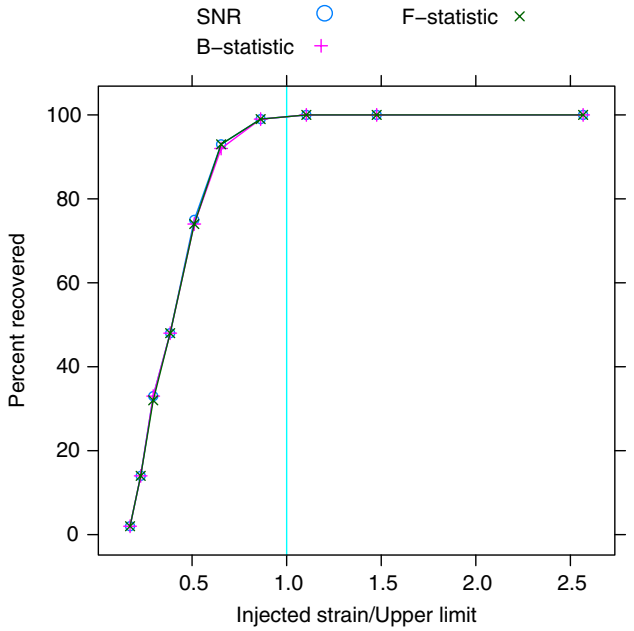

FIG. 5 (color online). Statistics efficiency versus strength of injected signal, relative to established upper limit (color online).

The computational performance of our code is shown in Fig. 6. The y-axis is in units of cycles per template, with each template computed once per frequency, sky position, and spin-down while sampling all possible alignments of the source. All statistics (SNR, upper limit, circular upper limit, $\mathfrak{F}$-statistic, and $\mathfrak{B}$-statistic) were computed during the run. The underlying data was purely Gaussian. The simulations were run on a cluster of 2.3 GHz AMD processors.

Our worst-case performance is below 1500 cycles which favorably compares with performance of resampling [15,21] that is estimated to be $\approx 20\,000$ cycles per template. The CPUutilization by different parts of the algorithm is shown in Table II. Only a third of the cycles is attributable to computation of the convolution, while at least 46% is spent gathering statistics, leaving much room for further improvement.

TABLE II. CPU cycles spent in different parts of the algorithm for sky position with declination of 1.0 rad.

CPU fraction	Code description
36%	Computation of upper limit
16%	Computation of 11-term convolution
10%	Computation of terms of convolutions
8%	General statistics function
7%	Computation of 63-term pseudo-convolution
2%	Computation of logarithm
Balance of 21%	Spread out over many parts of the algorithm

V. CONCLUSIONS

We have described an implementation of the loosely coherent statistic that searches over families of ideal continuous gravitational signals. The performance of the algorithm is more than an order of magnitude better than previously published algorithms, opening the way for exploring wider parameter spaces.

The algorithm is not specific to analysis of LIGO data and can be applied whenever one searches for narrow spectrum slowly modulated signals, for example, searches for binary systems in data from future space-based interferometers or discovery of radio pulsars.

There are several directions for further improvement:

(i) Making use of coherent or loosely coherent combination of data from two interferometers should improve sensitivity and provide better rejection of detector artifacts. This might also be an area where the $\mathcal{B}$-statistic will show its strength.

(ii) It is necessary to derive more efficient alternatives for the computation of upper limit.

(iii) It would be desirable to extend the algorithm to search over frequency evolution parameters to be able to detect binary systems with weak modulation.

(iv) The fine granularity of input data can be used to avoid high-intensity glitches, by excluding contaminated SFTs.

ACKNOWLEDGMENTS

This work has been done while being a member of LIGO laboratory, supported by funding from United States National Science Foundation. The author would like to thank Bruce Allen for the invaluable opportunity to stay at Max Planck Institute for Gravitational Physics (Albert Einstein Institute). The author has greatly benefited from suggestions and comments of his colleagues, in particular, Reinhard Prix and Keith Riles. The author is grateful to the anonymous referee for many helpful comments. This document has been assigned LIGO Laboratory document number LIGO-P1100104-v6.

APPENDIX A: EFFICIENT COMPUTATION OF $\mathcal{B}$-STATISTIC

$\mathcal{B}$-statistic was introduced in [14] as a Bayesian alternative to $\mathcal{F}$-statistic. It was shown that the $\mathcal{F}$-statistic is equivalent to a Bayesian statistic with a prior that favors linearly polarized signals and high signal strength. In contrast, the $\mathcal{B}$-statistic is isotropic in spin orientation.

The statistic starts with the likelihood function:

$$\begin{aligned}\mathcal{L}(x,\mathcal{A}) &= e^{\mathcal{A}^\mu x_\mu - (1/2)\mathcal{A}^\mu M_{\mu\nu}\mathcal{A}^\nu} \\ &= e^{\mathrm{Re}(\bar{w}_1 X_1 + \bar{w}_2 X_2) - (1/2)(Y_{11}|w_1|^2 + 2Y_{12}\mathrm{Re}(w_1\bar{w}_2) + Y_{22}|w_2|^2)}.\end{aligned} \tag{A1}$$

Instead of maximizing it, which is the approach of the $\mathcal{F}$-statistic we compute the integral:

$$\begin{aligned}\mathcal{B}(x) &= \int_{h<h_0} d\mathcal{A}\,\mathcal{L}(x,\mathcal{A}) \\ &= \int dw' \int_0^{h_0} dh \\ &\quad \cdot e^{h\mathrm{Re}(\bar{w}_1' X_1 + \bar{w}_2' X_2) - (1/2)h^2(Y_{11}|w_1'|^2 + 2Y_{12}\mathrm{Re}(w_1'\bar{w}_2') + Y_{22}|w_2'|^2)}.\end{aligned} \tag{A2}$$

The measures $d\mathcal{A}$ and dw' are chosen to be uniform in parameters ϕ, ψ, and $\cos(\iota)$. The integral with respect to h is not normalized, which allows to set its upper limit infinite and transition to an improper prior on $[0,\infty]$. This effectively changes $B(x)$ (as defined in [14]) to be in units of strain. It would be interesting to explore the possibility of deriving an upper limit estimator based on the same principles as $B(x)$.

The integral with respect to h can be shown to be a function of $\mathrm{SNR}_R(w_1', w_2')$ and total weight $\mathfrak{N}(w_1', w_2')$:

$$\int_0^\infty dx e^{ax - bx^2/2} = \frac{1}{\sqrt{b}} e^{a^2/2b} \int_{-\infty}^{a/\sqrt{b}} dx e^{-x^2/2} \tag{A3}$$

$$\mathcal{B}(x) = \int dw' \frac{e^{\mathrm{SNR}_R(w_1', w_2')/2}}{\sqrt{\mathfrak{N}(w_1', w_2')}} \int_{-\infty}^{\mathrm{SNR}_R(w_1', w_2')/2} e^{-x^2/2} dx. \tag{A4}$$

This still leaves a three-dimensional integral to carry out which is undesirable inside the inner loop. We note that $\mathfrak{N}(w_1', w_2')$ does not depend on phase ϕ, while

$$\mathrm{SNR}_R(w_1', w_2') = \cos(\phi)\mathrm{SNR}(w_1', w_2'). \tag{A5}$$

We can thus represent our statistic as

$$\mathcal{B}(x) = \int_{\phi=0} dw' \frac{1}{\sqrt{\mathfrak{N}(w_1', w_2')}} \Theta[\tfrac{1}{2}\mathrm{SNR}(w_1', w_2')], \tag{A6}$$

where $\Theta(x)$ is defined as

$$\Theta(x) = \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{\cos(\phi)x}}{\sqrt{2\pi}} \int_{-\infty}^{\cos(\phi)x} e^{-s^2/2} ds d\phi. \tag{A7}$$

We can now study $\Theta(x)$ as a new special function and find a means to compute it efficiently. We have the following easy identities:

$$\Theta(x) = \Theta(-x) \qquad \Theta(0) = \tfrac{1}{2}. \tag{A8}$$

It is also easy to compute approximations for small and large x:

$$\begin{aligned}\Theta(x) &= \frac{1}{2} + \left(\frac{1}{8} + \frac{1}{2\sqrt{2\pi}}\right)x^2 + O(x^4) \\ \Theta(x) &= \frac{e^{|x|}}{\sqrt{2\pi|x|}}[1 + O(1/x)].\end{aligned} \tag{A9}$$

Armed with these relations, we can spend some time in numerical experimentation and arrive at the following approximation:

$$\Theta(x) \approx \frac{\exp(\sqrt{0.25 + x^2})}{(4\pi^2 x^2 + 16e^2)^{1/4}} \frac{a_0 + a_2 x^2 + a_4 x^4 + x^6}{b_0 + b_2 x^2 + b_4 x^4 + x^6}, \tag{A10}$$

which has a 0.05% error over the entire range with the following values of constants:

$$\begin{aligned} a_0 &= 7.7199014890487 & b_0 &= 7.7201854234519 \\ a_2 &= 19.0337266315871 & b_2 &= 21.1533518190664 \\ a_4 &= 5.2017224760755 & b_4 &= 4.2818853782852. \end{aligned} \tag{A11}$$

Now we can compute the $\mathfrak{B}$-statistic with a simple sum over a uniform grid in ψ and $\cos(\iota)$, at a cost within an order of magnitude of computing the SNR statistic.

[1] V. Dergachev, Classical Quantum Gravity **27**, 205017 (2010).
[2] B. Abbott *et al.* (LIGO Scientific Collaboration), Rep. Prog. Phys. **72**, 076901 (2009).
[3] B. Abbott *et al.* (The LIGO and .Virgo Scientific Collaboration), arXiv:1110.0208v1.
[4] B. Abbott *et al.* (The LIGO Scientific Collaboration), Phys. Rev. D **77**, 022001 (2008).
[5] V. Dergachev, LIGO technical document LIGO-T050186, 2005, available in https://dcc.ligo.org/.
[6] V. Dergachev, LIGO technical document LIGO-T1000272, 2010, available in https://dcc.ligo.org/.
[7] P. Jaranowski, A. Królak, and B. F. Schutz, Phys. Rev. D **58**, 063001 (1998).
[8] P. Astone, K. M. Borkowski, P. Jaranowski, and A. Królak, Phys. Rev. D **65**, 042003 (2002).
[9] P. Astone *et al.*, Classical Quantum Gravity **20**, S665 (2003)
[10] A. Królak, M. Tinto, and M. Vallisneri, Phys. Rev. D **70**, 022003 (2004).
[11] C. Cutler and B. F. Schutz, Phys. Rev. D **72**, 063006 (2005).
[12] C. Cutler, arXiv:1104.2938v1.
[13] P. R. Brady, T. Creighton, C. Cutler, and B. F. Schutz, Phys. Rev. D **57**, 2101 (1998).
[14] R. Prix and B. Krishnan, Classical Quantum Gravity **26**, 204013 (2009).
[15] P. Patel, X. Siemens, R. Dupuis, and J. Betzwieser, Phys. Rev. D **81**, 084032 (2010).
[16] P. R. Brady and T. Creighton, Phys. Rev. D **61**, 082001 (2000).
[17] David. G. Blair, *The Detection of Gravitational Waves* (Cambridge University Press, Cambridge, England, 1991), p. 452, Chap. 16; B. F. Schutz, *Data Processing, Analysis and Storage for Interferometric Antennas* (Cambridge University Press, Cambridge, England, 1991).
[18] R. Prix, LIGO technical document Report No. LIGO-T0900149-v3, 2011, available in https://dcc.ligo.org/.
[19] V. Dergachev and K. Riles, LIGO Technical Document LIGO-T050187, 2005, available in https://dcc.ligo.org/.
[20] R. Prix and J. T. Whelan, Classical Quantum Gravity **24**, S565 (2007).
[21] Joseph Betzwiezer (private communication).