χ^2 time-frequency discriminator for gravitational wave detection

Bruce Allen

Department of Physics, University of Wisconsin–Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201, USA
(Received 1 June 2004; published 18 March 2005)

Searches for known waveforms in gravitational wave detector data are often done using matched filtering. When used on real instrumental data, matched filtering often does not perform as well as might be expected, because nonstationary and non-Gaussian detector noise produces large spurious filter outputs (events). This paper describes a χ^2 time-frequency test which is one way to discriminate such spurious events from the events that would be produced by genuine signals. The method works well only for broadband signals. The case where the filter-template does not exactly match the signal waveform is also considered, and upper bounds are found for the expected value of χ^2.

I. INTRODUCTION

Matched filtering is a common and effective technique used to search for signals with a known waveform in a data stream [1]. The output of a matched filter will be large if the data stream contains the desired signal. But it can also be driven to large values by spurious noise. This paper describes a χ^2 time-frequency discriminator statistic which has proven effective at distinguishing between these two possibilities.

The method was invented by the author in 1996 [2] for use on data from the LIGO 40m prototype gravitational-wave detector [3]. The method was subsequently used in the analysis of data from the Japanese TAMA detector [4–8] and the first analysis of science data from the full-scale LIGO detectors [9]. It has also been used in preliminary searches using VIRGO engineering data [10] and GEO-600 data [11]. Until now the only detailed description of the method was in the documentation for the GRASP software package [2], as referenced in the publications above. This paper describes the method in more detail, and analyzes its properties.

The χ^2 time-frequency discriminator is designed for use with broadband signals and detectors. The essence of the test is to "break up" the instrument's bandwidth into several smaller bands, and to see if the response in each band is consistent with what would be expected from the purported signal. This method can only be used to discriminate signals for which the gravitational waveform is *known*, meaning that it can be calculated in advance, with high precision. (Note: the word "known" is slightly misleading since the waveform typically still depends upon a few unknown parameters, such as the overall scale and initial phase.)

A new generation of broadband gravitational wave detectors is now undergoing commissioning [12] and more sensitive instruments are in the planning and design stages [13,14]. We expect that this test will prove useful for those instruments as well.

In searching for signals and setting upper limits, the primary use of the χ^2 time-frequency discriminator is as a *veto*. This means that events which might otherwise be used, analyzed or studied in more detail are rejected because they have a χ^2 value which is too large. In general terms, the χ^2 time-frequency discriminator may be thought of as a method for reducing contributions from the non-Gaussian tails that characterizes most gravitational wave detectors [15]. Substantial efforts have been made to characterize these tails in the TAMA [16,17], Explorer [18], and Nautilus [19] detectors. Other methods for reducing the effects of these tails have also been proposed and/or used. For example Creighton [20] proposes a simple analytic model for non-Gaussian tails and uses it to characterize a network detection algorithm which is insensitive to this non-Gaussian tail. Other filtering methods, based on locally-optimal statistics which are less sensitive or insensitive to non-Gaussian tails have been proposed [21,22] for matched filtering and stochastic background searches. Shawhan and Ochsner [23] have developed a heuristic veto method for matched filtering for binary inspiral, based on counting threshold crossings in a short-time window. When tuned for the LIGO S1 data set with a half-second window, the method is effective, and provides a veto which is complementary to the χ^2 time-frequency discriminator presented here. Some related ideas have also been explored by Guidi [24].

The principal source that will serve as an example here is the gravitational radiation back-reaction driven inspiral of pairs of compact stars, also known as "binary inspiral". If each of the two stars (typically neutron stars or black holes) has masses smaller than a few solar masses, then the waveforms can be accurately calculated over the typical detector bandwidth (30–500 Hz) using post-Newtonian approximations [25–34]. In this case, the unknown signal parameters include an overall amplitude scale, the masses and spins of the two stars, a fiducial reference time (often taken to be the "coalescence time"), and the initial phase of the orbit. (The final unknown parameter, the orbital inclination, is degenerate with these other parameters, and may therefore be ignored.

---

*Electronic address: ballen@uwm.edu

The orbital eccentricity may also be neglected: by the time such systems are emitting in the detection band, they have radiated away any eccentricity and the orbit has circularized.) These signals are broadband, since the binary system is observed at the very end of its life when the signal frequency is increasing rapidly as the orbital period decreases and the stars spiral together.

The paper is structured as follows. Section II defines the notational conventions that are used. Section III derives the form of the optimal matched filter in the simplest case and describes its properties. Section IV defines the χ^2 time-frequency discriminator and derives its basic properties, also for the simplest case. Section V gives a brief illustrative example of this statistical test in action, computing and comparing the χ^2 values obtained for a simulated inspiral signal and a spurious noise event. Readers interested in acquiring a quick understanding of the method without seeing the technical details should start with this Section.

A significant problem in searching for gravitational wave signals is that the waveform depends upon a number of parameters (for example, masses) of the source. For this reason, data must be searched with a bank of filters designed to cover this parameter space [35,36]. Since this bank is discrete, and the source parameters are continuous, the match between signal and filter is never exact. The effects of this signal/template mismatch on the χ^2 discriminator are investigated and quantified in Sec. VI. One of the main results of this paper is an absolute upper limit on the expected value of χ^2 arising from template/signal mismatch.

Up to this point, the signals studied are of the simplest type, which can be completely described with only two unknown parameters: an overall amplitude, and an offset/arrival time. However this is insufficient for most cases of interest, where the signals are an (*a priori* unknown) linear combination of two different polarizations. Section VII treats this case, deriving two-phase results analogous to the single-phase results of the previous Sections.

Section VIII examines suitable thresholds on χ^2 for stationary Gaussian noise, and contrasts these with the heuristic thresholds used in published analysis of real detector data such as the LIGO S1 binary inspiral upper limit analysis [9].

Section IX examines a variation of the discriminator based on "unequal expected SNR" intervals, and shows that this discriminator still has most of the properties of the χ^2 discriminator defined in previous Sections.

There are an infinity of possible χ^2-like statistical tests and discriminators. Work by Baggio *et al.* [37] introduced a χ^2 test for use with resonant-mass gravitational wave detectors. In Sec. X the χ^2 time-frequency discriminator of this paper is compared to that test. While the tests share some similar features, they have quite different properties and behavior.

This is followed by a brief Conclusion, which summarizes the main results and some of the unanswered questions.

Appendix A contains a short calculation proving that the time-frequency discriminator defined in this paper has a classical χ^2 distribution if the detector's noise is Gaussian. Appendix B derives a simple mathematical result used in the body of the paper,

II. CONVENTIONS

The Fourier Transform of a function of time $V(t)$ is denoted by $\tilde{V}(f)$ and is given by

$$\tilde{V}(f) = \int e^{-2\pi i f t} V(t) dt. \qquad (2.1)$$

The inverse Fourier transform is

$$V(t) = \int e^{2\pi i f t} \tilde{V}(f) df. \qquad (2.2)$$

All integrals are from $-\infty$ to ∞ unless otherwise indicated, and $*$ denotes complex conjugate.

The detector output (typically a strain) is denoted by

$$s(t) = n(t) + h(t) \qquad (2.3)$$

where $n(t)$ is the (real) strain-equivalent noise produced by fluctuations within the detector and its environment, and $h(t)$ is a (real) gravitational waveform of astrophysical origin.

Since the detector's noise $n(t)$ can only be characterized statistically, one must introduce tools for determining the expected properties of quantities measured in the presence of this noise. There are several equivalent ways to do this. In this paper, we imagine that $n(t)$ is a random time-series drawn from a large ensemble of such time-series, whose statistical properties are those of the instrument noise [38]. If W is some functional that depends upon $n(t)$, then angle brackets $\langle W \rangle$ denote the average of W over the ensemble of different $n(t)$.

We assume that $\langle n(t) \rangle$ vanishes, which implies that $\langle \tilde{n}(f) \rangle = 0$. We also assume that the statistical properties of the detector's noise are second-order stationary [39] which implies that the expectation value $\langle n(t)n(t') \rangle$ depends only upon the time difference $t - t'$. It then follows that in frequency space

$$\langle \tilde{n}(f)\tilde{n}^*(f') \rangle = S_n(f)\delta(f - f'), \qquad (2.4)$$

where $\delta(f)$ is the Dirac delta-function. The *two-sided noise power spectrum* is a real non-negative even function $S_n(f)$ which from (2.4) can be shown to satisfy

$$\langle n^2(t) \rangle = \int S_n(f) df. \qquad (2.5)$$

This implies that $2S_n(f)df$ may be interpreted as the expected squared strain in the frequency band from f to $f + df$, for $f \geq 0$. (Note that much of the literature on

this subject, including publications of the LIGO Scientific Collaboration, uses a *one-sided* power spectrum $2S_n(f)$, because this is typically the quantity measured by standard instrumentation. Its use here would complicate many formulae with extraneous factors of two.)

As explained earlier, we are interested in the case of a known waveform. A prototypical example is a binary inspiral chirp waveform, which may be written as [28]

$$h(t) = \frac{D}{d}[\cos\phi T_c(t - t_0) + \sin\phi T_s(t - t_0)]. \qquad (2.6)$$

This waveform depends upon three *nuisance parameters.* These are the effective distance d to the source, a fiducial time t_0 (for example the coalescence time of the binary pair) and a phase ϕ which is determined by the orbital phase of the binary pair and its orientation relative to the detector.

The templates T_c and T_s are the signal waveforms that would be produced by a binary inspiral pair optimally-oriented with respect to the detector, at distance D, in the two possible polarization states (corresponding to rotating the detector axes by 45°). The waveform may depend upon additional nuisance parameters; we will return to this later.

For pedagogic purposes, we first consider the simpler case in which the phase ϕ of the waveform is known *a priori*, in advance,

$$h(t) = \frac{D}{d}T(t - t_0). \qquad (2.7)$$

In this case there are only two nuisance parameters: coalescence time to and effective distance d.

The quantity D is the canonical distance at which a source, optimally-oriented with respect to the detector, would produce the waveform T. Its value determines the overall normalization scale of the waveform T, since, for a given source type, the quantity $DT(t)$ is *independent* of D, and is determined by the laws of physics, specifically General Relativity.

III. MATCHED FILTERING

A matched filter is the optimal linear filter for detection of a particular waveform. Its form can be derived using a number of different techniques. Here we use one of the classical signal analysis methods.

For notational purposes it is useful to introduce a Hermitian inner product defined by

$$(A, B) = \int \frac{A^*(f)B(f)}{S_n(f)} df, \qquad (3.1)$$

for any pair of complex functions $A(f)$ and $B(f)$. The frequency dependence of A and B will usually be implied and not indicated explicitly.

A real detector functions only over a finite frequency band, and acquires data at a finite sample rate. In this case, the noise power spectrum S_n may be taken to be infinite outside the bandwidth of the instrument, effectively restricting the range of integration to lie between plus and minus the Nyquist frequency $f_N = 1/(2\Delta t)$, where Δt is the time between successive data samples.

The matched filter is a linear operator that maximizes the ratio of "signal" to "noise". We denote the filter by $\tilde{Q}^*(f)/S_n(f)$ and the output of the filter by z, so

$$z \equiv \int \frac{\tilde{Q}^*(f)\tilde{s}(f)}{S_n(f)} df = (\tilde{Q}, \tilde{s}). \qquad (3.2)$$

We require that z be real, which implies that $\tilde{Q}(f) = \tilde{Q}^*(-f)$, and also means that $\tilde{Q}(f)/S_n(f)$ corresponds to a real function (filter kernel) in the time-domain.

The expected value of z can be found from (2.7), and is given by

$$\langle z \rangle = \frac{D}{d}(\tilde{Q}, \tilde{T}e^{-2\pi i f t_0}) \qquad (3.3)$$

Here $\tilde{T}(f)$ denotes the Fourier transform of $T(t)$; the translation in time by t_0 is explicitly encoded in the exponential term. Note that in the absence of a source $(d \to \infty)$ the expected value of z vanishes since $\langle \tilde{n}(f) \rangle = 0$ then implies that $\langle \tilde{s}(f) \rangle = 0$.

The expected value of the square of z may be found using (2.4)

$$\langle z^2 \rangle = (\tilde{Q}, \tilde{Q}) + \left(\frac{D}{d}\right)^2 (\tilde{Q}, \tilde{T}e^{-2\pi i f t_0})^2 \qquad (3.4)$$

To estimate the error or uncertainty in a measurement of z, it is useful to define

$$\delta z = z - \langle z \rangle. \qquad (3.5)$$

The error or uncertainty in a measurement of z, due to noise in the detector, is

$$\sqrt{\langle (\delta z)^2 \rangle} = \sqrt{\langle (z - \langle z \rangle)^2 \rangle} = \sqrt{\langle z^2 \rangle - \langle z \rangle^2} = (\tilde{Q}, \tilde{Q})^{1/2}. \qquad (3.6)$$

From these quantities, we can now derive the properties of the optimal matched filter.

Under the assumptions that we have made about the detector output (2.3) the optimal choice of matched filter $\tilde{Q}$ is the one that maximizes the ratio of the expected filter output $<z>$ given by (3.3) to its expected uncertainty (3.6) due to detector noise. Hence the optimal choice of $\tilde{Q}$ maximizes

$$\frac{\langle z \rangle}{\sqrt{\langle (\delta z)^2 \rangle}} = \frac{(\tilde{Q}, A)}{(\tilde{Q}, \tilde{Q})^{1/2}}, \qquad (3.7)$$

where we have *defined*

$$A(f) \equiv \frac{D}{d}\tilde{T}(f)e^{-2\pi i f t_0}. \qquad (3.8)$$

Since the inner product is Hermitian, Schwartz's inequal-

ity states that

$$|(\tilde{Q}, A)|^2 \le (A, A)(\tilde{Q}, \tilde{Q}). \qquad (3.9)$$

The two sides are equal if and only if $\tilde{Q}$ is proportional to A. Hence the ratio (3.7) is maximized when $\tilde{Q}(f)$ is proportional to $A(f)$. Thus, the optimal filter is a time-reversed copy of the template, weighted by the expected noise in the detector [40].

The *Signal-to Noise Ratio* (SNR) is defined by the ratio of the *observed* filter output z to its (expected or observed) root-mean-square (r.m.s.) fluctuations

$$\text{SNR} = \frac{z}{\sqrt{\langle(\delta z)^2\rangle}} = \frac{(\tilde{Q}, \tilde{s})}{\sqrt{(\tilde{Q}, \tilde{Q})}}, \qquad (3.10)$$

and is independent of the normalization of the optimal filter $\tilde{Q}$. By definition, in the absence of a signal $\langle\text{SNR}\rangle = 0$ and $\langle(\text{SNR})^2\rangle = 1$.

It is convenient to choose the normalization of the optimal filter $\tilde{Q}$ so that $(\tilde{Q}, \tilde{Q}) = 1$. This may be achieved by choosing the filter $\tilde{Q}$ to be

$$\tilde{Q}(f) = (\tilde{T}, \tilde{T})^{-1/2}\tilde{T}(f)e^{-2\pi i f t_0}. \qquad (3.11)$$

With this normalization choice, the filter output z is equal to the SNR. Henceforth we will use z to denote both of these quantities.

While the optimal filter $\tilde{Q}$ is explicitly independent of the normalization scale of the template T, we showed earlier that the scales of D and T could be freely adjusted provided that their product DT was held fixed. For the purposes of interpreting the SNR z, it is convenient to set the distance scale D so that

$$(\tilde{T}, \tilde{T}) = 1. \qquad (3.12)$$

With this choice of normalization, the expected value of the SNR is

$$\langle z \rangle = \frac{D}{d}(\tilde{T}, \tilde{T})^{1/2} = \frac{D}{d}. \qquad (3.13)$$

This choice of normalization is thus equivalent to choosing the distance D at which the template is defined to be the distance at which an optimally-oriented source would have an expected SNR of unity: $\langle z \rangle = 1$.

Since the expected value of z is proportional to the inverse distance, one may use the actual measured value of z to estimate the distance. Since the actual measured value of z is affected by instrument noise, this estimator has some average error. One can easily estimate the error, since with our normalization choices

$$\langle z^2 \rangle = 1 + \left(\frac{D}{d}\right)^2, \qquad \text{and hence} \quad \langle(\delta z)^2\rangle = 1. \quad (3.14)$$

This means that the expected fractional error in estimating the inverse distance to the source is

$$\frac{\langle(\delta z)^2\rangle^{1/2}}{\langle z \rangle} = \frac{1}{\langle z \rangle} = \frac{d}{D} \qquad (3.15)$$

Thus, a measured SNR of $z = 10$ implies a fractional accuracy in distance determination of about 10%.

Up to this point, we have been assuming that the fiducial coalescence time t_0 is known. In practice, one searches a data stream for statistically significant values of z, for all possible choices of t_0. As a function of t_0 the SNR is

$$z(t_0) = \int \frac{\tilde{s}(f)\tilde{T}^* e^{2\pi i f t_0}}{S_n(f)} df. \qquad (3.16)$$

Because this is just an inverse Fourier transform, it is both practical and simple to compute this quantity from a data stream $s(t)$. For example the Fast Fourier Transform (FFT) algorithm allows the right hand side (r.h.s.) to be computed in order $N \ln N$ operations, where N is the number of data samples of h in the time or frequency domain.

IV. THE χ^2 DISCRIMINATOR TEST

In the previous section, we assumed only that the detector noise was second-order stationary. It is quite common in such studies to also assume that the noise is Gaussian. One may then show that the probability of the SNR exceeding some threshold falls exponentially with increasing threshold, and so large values of the SNR have low probability of being due to noise in the detector, and thus are a good indication that a real source is present.

Unfortunately, experience has shown that the noise in broadband gravitational-wave detectors is far from Gaussian. Typically it has a Gaussian or Gaussian-like component (arising from electrical, thermal and shot noise) but there is another "glitchy" component that could be described as Poisson-like impulse noise. There are many sources of this noise, including marginally stable servo systems and environmental anomalies. The effects of this noise on the filtering process described above can be dramatic. Whereas the matched filter is designed to give a large response when the signal waveform matches the template, it also can give a large response when the instrumental noise has a large glitch. Although the waveform of this glitch looks nothing like the template, it nevertheless drives the filter output to a large value.

The statistical test described here provides a way to determine if the output of the filter is consistent with what might be expected from a signal that matched the template. To be effective, both the signal and the detector must be broadband.

One way to understand how this test works is to imagine that instead of a single broadband detector, one is given p data streams from p different, independent narrow band detectors, each operating in a different fre-

quency band. For each detector, one can construct an optimal filter for the signal, and then one can ask if the results are consistent, for example, if the p (potentially different) fiducial times t_0 which maximize the output of each of the p independent detectors are consistent with a single value.

Begin by assuming that, using matched filtering as previously described, we have identified a time of arrival t_0 and an inverse distance D/d. The goal is to construct a statistic which indicates if the filter output is consistent with this signal.

We will do this by investigating the way in which $z(t_0)$ gets its contribution from different ranges of frequencies. To do this, we partition the frequency range $f \in [0, \infty)$ into a set of p distinct subintervals $\Delta f_1, \cdots, \Delta f_p$ whose union is $[0, \infty)$. The frequency intervals:

$$
\begin{aligned}
\Delta f_1 &= \{f \mid 0 \le f < f_1\}, \\
\Delta f_2 &= \{f \mid f_1 \le f < f_2\}, \\
&\cdots \\
\Delta f_{p-1} &= \{f \mid f_{p-2} \le f < f_{p-1}\}, \\
\Delta f_p &= \{f \mid f_{p-1} \le f < \infty\},
\end{aligned}
\tag{4.1}
$$

will be defined by the condition that the *expected signal contributions in each frequency band from a chirp are equal.* (Note that one may also pick intervals which do not satisfy this condition. In Sec. IX we show that when suitably defined, the resulting statistic still has a classical χ^2 distribution for the case of Gaussian detector noise.)

To define the frequency bands, it is helpful to introduce a set of p Hermitian inner products (for $j = 1, \cdots, p$) defined in analogy to (3.1) by

$$
(A, B)_j = \int_{-\Delta f_j \cup \Delta f_j} \frac{A^*(f) B(f)}{S_n(f)} df. \tag{4.2}
$$

In each of these integrals, the range of integration is over both the positive and negative frequencies. As discussed following (3.1), since $S_n(f)$ may be taken as infinite for $|f|$ greater than the Nyquist frequency f_N, the effective upper limit of the final frequency interval Δf_p is f_N rather than ∞.

Since the frequency intervals do not overlap, but cover all frequency values, the sum of these inner products

$$
(A, B) = \sum_{j=1}^{p} (A, B)_j \tag{4.3}
$$

yields the inner product (3.1) defined earlier. The p distinct frequency bands are uniquely determined by the condition that

$$
\text{choose } \Delta f_j \text{ so that } (\tilde{T}, \tilde{T})_j = \frac{1}{p}. \tag{4.4}
$$

A typical set of frequency intervals in shown in Fig. 1.

FIG. 1. A typical set of frequency intervals Δf_j for the case $p = 4$. These intervals are narrowest where the detector is the most sensitive, and broadest where it is least sensitive.

For given instrumental noise $S_n(f)$ the frequency intervals Δf_j depend upon the template T. However it may be the case that many templates actually share the *same* frequency intervals Δf_j. A good example of this is the set of stationary-phase post-Newtonian templates, where the amplitude is calculated in the first post-Newtonian approximation, and the phase is calculated to higher order [41,42]. For these templates, the frequency intervals are determined by

$$
(\tilde{T}, \tilde{T})_j = \frac{1}{p}(\tilde{T}, \tilde{T}),
$$

$$
\int_{\Delta f_j} \frac{f^{-7/3}}{S_n(f)} df = \frac{1}{p} \int_0^\infty \frac{f^{-7/3}}{S_n(f)} df, \tag{4.5}
$$

provided that m_1 and m_2 lie in a range for which the stationary-phase approximation holds within the detector band [43]. For this family of templates, *all* the templates share the *same* bands Δf_j.

The SNR (3.2) is an integral over all frequencies, and can be written as a sum of contributions from each of the p different bands,

$$
z = \sum_{j=1}^{p} z_j \quad \text{with} \quad z_j \equiv (\tilde{Q}, \tilde{s})_j. \tag{4.6}
$$

The expected values of z_j and its square are computed using the same techniques as before, and give

$$
\langle z_j \rangle = \frac{1}{p} \frac{D}{d}, \quad \text{and} \quad \langle z_j^2 \rangle = \frac{1}{p} + \frac{1}{p^2} \left(\frac{D}{d} \right)^2 \tag{4.7}
$$

In the absence of a signal (take $d \to \infty$) one finds

$$
\langle z_j \rangle = 0 \quad \text{and} \quad \langle z_j^2 \rangle = \frac{1}{p}. \tag{4.8}
$$

This suggests an obvious statistical test to see if the signal is consistent with the model.

Consider the p quantities defined by

$$
\Delta z_j \equiv z_j - \frac{z}{p}. \tag{4.9}
$$

These are the differences between the SNR in the band Δf_j, and the SNR that would be anticipated [44] in that band, based on the total *measured* SNR in all bands. By definition, these differences sum to zero

$$
\sum_{j=1}^{p} \Delta z_j = 0 \tag{4.10}
$$

and their individual expectation values vanish

$$\langle \Delta z_j \rangle = 0. \tag{4.11}$$

To calculate the expectation values of their squares, first note that the quantity $\langle z_j z \rangle$ must, by symmetry, be j-independent [45]. Since the sum over j of $\langle z_j z \rangle$ yields $\langle z^2 \rangle$, one must have

$$\langle z_j z \rangle = \frac{\langle z^2 \rangle}{p} = \frac{1}{p}\left[1 + \left(\frac{D}{d}\right)^2\right]. \tag{4.12}$$

Thus the expectation value of the square of Δz_j is

$$\langle (\Delta z_j)^2 \rangle = \left\langle \left(z_j - \frac{z}{p}\right)^2 \right\rangle = \langle z_j^2 \rangle + \frac{\langle z^2 \rangle}{p^2} - \frac{2\langle z_j z \rangle}{p}$$
$$= \frac{1}{p}\left(1 - \frac{1}{p}\right) \tag{4.13}$$

Notice that these quantities Δz_j do not depend upon d. In fact *these quantities Δz_j and their second-order statistical properties, are independent of whether or not a signal is present*. This motivates the definition of a discrimination statistic.

We define the χ^2 time-frequency discriminator statistic by

$$\chi^2 = \chi^2(z_1, \cdots, z_p) = p \sum_{j=1}^{p} (\Delta z_j)^2. \tag{4.14}$$

This choice of statistic is one of the main results of the paper: in what follows we will study its properties in detail.

It follows immediately from (4.13) that the expected value of χ^2 is

$$\langle \chi^2 \rangle = p - 1 \tag{4.15}$$

Up to this point, the only assumption we have made is that the noise in the instrument is second-order stationary, specifically that $\langle n(t)n(t') \rangle$ depends only upon $t - t'$. To further analyze the properties of this statistic, additional assumptions are needed.

In the design of signal processing algorithms, it is common to analyze the performance of a method in the case where the instrument noise is both stationary and Gaussian. In this case, Appendix A shows that probability distribution function of χ^2 is a classical χ^2-distribution with $p - 1$ degrees of freedom. The (cumulative) probability that $\chi^2 < \chi_0^2$ is

$$P_{\chi^2 < \chi_0^2} = \int_0^{\chi_0^2/2} \frac{u^{p/2 - 3/2} e^{-u}}{\Gamma(\frac{p}{2} - \frac{1}{2})} \, du \tag{4.16}$$

$$= \frac{\gamma(\frac{p}{2} - \frac{1}{2}, \frac{\chi_0^2}{2})}{\Gamma(\frac{p}{2} - \frac{1}{2})} \tag{4.17}$$

where γ is the incomplete gamma function. In this case, where the noise is assumed to be stationary and Gaussian, the expected distribution of χ^2 values is quite narrow. One

has

$$\langle (\chi^2)^2 \rangle = p^2 - 1 \tag{4.18}$$

which implies that the "width" of the χ^2 distribution is

$$[\langle (\chi^2)^2 \rangle - \langle \chi^2 \rangle^2]^{1/2} = \sqrt{2(p-1)} \tag{4.19}$$

Thus, if the noise were stationary and Gaussian, we would expect to find χ^2 values in the range $[p - 1 - \sqrt{2(p-1)}, p - 1 + \sqrt{2(p-1)}]$. Since the fractional width of this range decrease with increasing p, one might expect that large values of p are desirable, since they appear to give high discriminating power.

Practice and experience have shown that large values of p do not, in fact, work very well [11]. Partly this is because the detector noise is neither stationary nor Gaussian, and partly this is because the signal is not a perfect match to the template. Large values of p tend to spread nonstationary glitch noise over many frequency bands, diluting its effect on χ^2. This is difficult or impossible to model analytically, and can best be understood (as in [11]) by Monte-Carlo studies of simulated signals added into real detector noise. However the effects of a signal-template mismatch can be studied analytically; this is done in Sec. VI.

V. HOW DOES THE χ^2 TEST WORK?

The χ^2 test was invented based on experience filtering data from the LIGO 40m prototype instrument [2,3]. It was observed that a binary inspiral filter bank registered many events that (when converted to audio) did not sound like inspiral signals. In particular, the low frequency component of the signal did not arrive first, followed by the midrange and high frequency components. The χ^2 test first arose from considering a set of matched filters in different bands, and testing to see if the filter outputs all peaked at the correct time. The signal z_1 was constructed from the lowest frequency band, z_2 from the next frequency band, and so on. This is illustrated in Fig. 2 for a single-phase test with $p = 4$ bands.

It is instructive to compare the values of the filter outputs (single-phase test) for the two cases shown in Fig. 2. For the simulated chirp, the signal-to-noise ratio was $z = 9.2$ and the signal values in the different bands were

$$\begin{aligned} z_1 &= 2.25 & z_2 &= 2.44 \\ z_3 &= 1.87 & z_4 &= 2.64 \\ z &= z_1 + z_2 + z_3 + z_4 = 9.2 \\ \chi^2 &= 4 \sum_{j=1}^{4} (z_j - z/4)^2 = 1.296 \end{aligned} \tag{5.1}$$

$$P_{\chi^2 \geq 1.296} = 1 - \frac{\gamma(3/2, 0.648)}{\Gamma(3/2)} = 73\%.$$
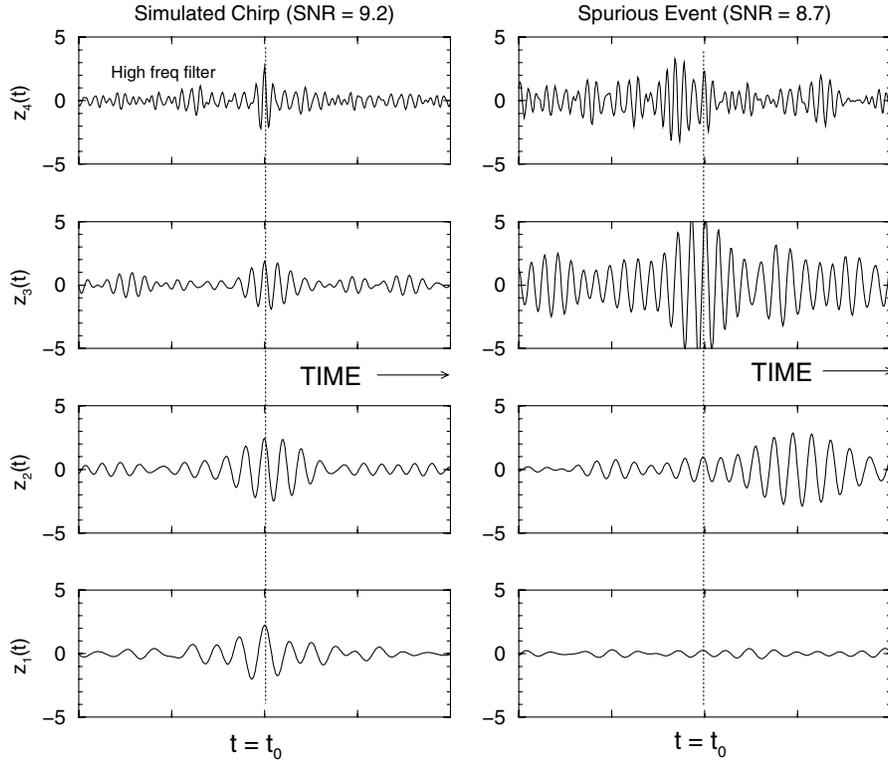
FIG. 2. The output of $p = 4$ single-phase filters for a simulated chirp signal added into a stream of detector noise (left set of figures) and a transient burst present in detector noise (right set of figures). For the simulated chirp, the filters in the different frequency bands all peak at the same time offset t_0: the time offset which maximizes the SNR. At this instant in time, all of the z_j are about the same value. However when the filter was triggered by the transient burst, the filters in the different frequency bands peak at different times. At time t_0 they have very different values (some large, some small, and so on).

This is quite consistent with the value of χ^2 that would be expected for a chirp signal in additive Gaussian noise.

For the spurious noise event shown in Fig. 2 the SNR $z = 8.97$ was quite similar but the value of χ^2 is very different:

$$z_1 = 0.23 \qquad z_2 = 0.84$$
$$z_3 = 5.57 \qquad z_4 = 2.33$$
$$z = z_1 + z_2 + z_3 + z_4 = 8.97$$
$$\chi^2 = 4 \sum_{j=1}^{4} (z_j - z/4)^2 = 68.4 \qquad (5.2)$$
$$P_{\chi^2 \geq 68.4} = 1 - \frac{\gamma(3/2, 34.2)}{\Gamma(3/2)} = 9.4 \times 10^{-15}.$$

The probability that this value of χ^2 would be obtained for a chirp signal in additive Gaussian noise is *extremely* small.

VI. EFFECT OF A SIGNAL/TEMPLATE MISMATCH ON χ^2

In the previous two Sections, we analyzed an optimal filter and constructed a χ^2 statistic for the case where the signal waveform was known exactly. In practice, this is not possible. Typically, signal waveforms come from a family characterized by a set of continuous parameters, such as masses and spins. Thus, in practice, to search for signals one uses a discrete set of templates, called a *template bank* [35,36]. Such banks can contain anywhere from dozens to hundreds of thousands of templates. Since each template in the bank is defined by a point in parameter space, the template bank may be thought of as a grid, or mesh, in parameter space. Typically, this grid is laid out to ensure that any signal from the continuous family of waveforms is "near" some point in the grid. In this section, we analyze the case where the signal waveform is "close" to the template waveform, but not a perfect match.

We begin by assuming that the signal is perfectly described by a waveform T', so that the detector's output is

$$s(t) = n(t) + \frac{D'}{d'} T'(t). \qquad (6.1)$$

Adopting the same conventions as before, we assume that D' is chosen so that T' obeys $(\tilde{T}', \tilde{T}') = 1$. For simplicity, and without loss of generality, we take $t_0 = 0$. Assume that this signal is close to that of the template T, and hence that the signal is detected in that template. The SNR is

$$z = (\tilde{Q}, \tilde{s}) = (\tilde{T}, \tilde{n}) + \frac{D'}{d'}(\tilde{T}, \tilde{T}'). \tag{6.2}$$

Using Schwartz's inequality, the inner product between the two templates must lie in the range $[-1, 1]$:

$$(\tilde{T}, \tilde{T}')^2 \le (\tilde{T}, \tilde{T})(\tilde{T}', \tilde{T}') \le 1. \tag{6.3}$$

One may think of the two templates as unit vectors separated by an angle θ and write this in the form

$$(\tilde{T}, \tilde{T}') = \cos\theta, \quad \text{for } \theta \in [0, \pi]. \tag{6.4}$$

This inner product is often called the *fitting factor*. The expected value of the SNR

$$\langle z \rangle = \frac{D'}{d'} \cos\theta, \tag{6.5}$$

is reduced by a factor of the fitting factor compared with the expected SNR D'/d' that would be obtained if the template bank contained a template perfectly matching the waveform T'.

The fractional difference between this "ideal case" expected SNR and the expected SNR in the mismatched template is called the *template mismatch* ϵ

$$\cos\theta = 1 - \epsilon. \tag{6.6}$$

The value of ϵ must lie in the range $\epsilon \in [0, 2]$, and may be restricted to the range $\epsilon \in [0, 1]$ by changing the sign of T' if needed. Hence, without loss of generality we will assume that $0 \le \cos\theta \le 1$ and that $0 \le \epsilon \le 1$. The case of most interest is when $\epsilon \ll 1$. Typically template banks are set up so that the worst-case mismatch corresponds to a loss of event rate (for a uniform source distribution) of 10%. Since the volume inside a sphere of radius r grows proportional to r^3 and the SNR is inversely proportional to distance, this corresponds to a typical worst-case template mismatch of $\epsilon = 0.033 = 3.3\%$.

Following the same procedures as in Sec. III one can find the expected SNR squared, which is

$$\langle z^2 \rangle = 1 + \left(\frac{D'}{d'}\right)^2 \cos^2\theta. \tag{6.7}$$

Thus, the first- and second-order statistics of z are indistinguishable from those that would be produced by a signal from a perfectly matched template T with SNR $D' \cos\theta/d'$.

To analyze the effects of the signal/template mismatch on the χ^2 statistic is slightly more involved. We begin by considering the way in which the templates overlap in each individual frequency band. Define a set of p real constants $\lambda_1, \cdots, \lambda_p$ by

$$(\tilde{T}, \tilde{T}')_j = \lambda_j \cos\theta. \tag{6.8}$$

It follows from (6.4) that these constants sum to unity,

$$\sum_{j=1}^{p} \lambda_j = 1. \tag{6.9}$$

Hence, the average value over frequency bins of the λ_j is $1/p$. The deviation away from this value is a measure of how close together (or far apart) the templates T and T' are in the frequency band Δf_j.

The goal is to understand how the χ^2 statistic is affected by signal/template mismatch. To determine this, we first express the SNR in the j'th band as

$$z_j = (\tilde{Q}, \tilde{s})_j = (\tilde{T}, \tilde{n})_j + \frac{D'}{d'}(\tilde{T}, \tilde{T}')_j \tag{6.10}$$

$$= (\tilde{T}, \tilde{n})_j + \frac{D'}{d'} \lambda_j \cos\theta. \tag{6.11}$$

Using calculations identical to Sec. IV the expected value of the SNR and its square in the j'th band are

$$\langle z_j \rangle = \frac{D'}{d'} \lambda_j \cos\theta, \tag{6.12}$$

and

$$\langle z_j^2 \rangle = \frac{1}{p} + \left(\frac{D'}{d'}\right)^2 \lambda_j^2 \cos^2\theta. \tag{6.13}$$

As before, we define $\Delta z_j \equiv z_j - z/p$, giving

$$\Delta z_j = (\tilde{T}, \tilde{n})_j - \frac{1}{p}(\tilde{T}, \tilde{n}) + \frac{D'}{d'}\left(\lambda_j - \frac{1}{p}\right)\cos\theta. \tag{6.14}$$

The first difference between this analysis and the one for matching templates is that in the mismatched case, the expectation value of Δz_j does not vanish:

$$\langle \Delta z_j \rangle = \frac{D'}{d'}\left(\lambda_j - \frac{1}{p}\right)\cos\theta. \tag{6.15}$$

As before, we can use the assumption that the detector noise is second-order stationary to calculate

$$\langle z_j z \rangle = \left\langle \left[(\tilde{T}, \tilde{n})_j + \frac{D'}{d'}\lambda_j \cos\theta\right]\left[(\tilde{T}, \tilde{n}) + \frac{D'}{d'}\cos\theta\right]\right\rangle$$
$$= (\tilde{T}, \tilde{T})_j + \left(\frac{D'}{d'}\right)^2 \lambda_j \cos^2\theta = \frac{1}{p} + \left(\frac{D'}{d'}\right)^2 \lambda_j \cos^2\theta. \tag{6.16}$$

Using these results, it is straightforward to work out the expectation value of $(\Delta z_j)^2$:

$$\langle (\Delta z_j)^2 \rangle = \langle z_j^2 \rangle + \frac{\langle z^2 \rangle}{p^2} - \frac{2\langle z_j z \rangle}{p}$$
$$= \frac{1}{p}\left(1 - \frac{1}{p}\right) + \left(\frac{D'}{d'}\right)^2\left(\lambda_j - \frac{1}{p}\right)^2 \cos^2\theta. \tag{6.17}$$

The expectation value of the χ^2 discriminator statistic (4.14) is therefore

$$\langle\chi^2\rangle = p - 1 + \left(\frac{D'}{d'}\right)^2 \cos^2\theta \sum_{j=1}^{p} p\left(\lambda_j - \frac{1}{p}\right)^2$$

$$= p - 1 + \langle z\rangle^2 \sum_{j=1}^{p} p\left(\lambda_j - \frac{1}{p}\right)^2. \tag{6.18}$$

This is in sharp contrast to the case where the signal waveform matched the template perfectly. In that case, the expectation value of χ^2 was independent of the signal strength. Here, when the signal and template do not match perfectly, the expected value of χ^2 depends quadratically on the expected SNR $\langle z\rangle$ of the signal.

The dependence of the discriminator $\langle\chi^2\rangle$ on the square of the expected SNR $\langle z\rangle^2$ has a coefficient

$$\kappa = p \sum_{j=1}^{p} \left(\lambda_j - \frac{1}{p}\right)^2 = -1 + p \sum_{j=1}^{p} \lambda_j^2. \tag{6.19}$$

The quantity κ is manifestly non-negative; we now obtain an absolute upper bound on its value. Schwartz's inequality implies that

$$(\tilde{T}, \tilde{T}')_j^2 \leq (\tilde{T}, \tilde{T})_j (\tilde{T}', \tilde{T}')_j,$$

$$\lambda_j^2 \cos^2\theta \leq \frac{1}{p}(\tilde{T}', \tilde{T}')_j$$

$$\Rightarrow \tag{6.20}$$

$$\lambda_j^2 \leq \frac{1}{p\cos^2\theta}(\tilde{T}', \tilde{T}')_j.$$

Summing both sides over j one obtains

$$\sum_{j=1}^{p} \lambda_j^2 \leq \frac{1}{p\cos^2\theta}. \tag{6.21}$$

Combining this with the definition (6.19) of κ, one obtains the bound

$$0 \leq \kappa \leq \frac{1}{\cos^2\theta} - 1. \tag{6.22}$$

Note that this result does not assume *any* relationship between the signal waveform T and template waveform T'.

For the case of most interest (small template mismatch $\epsilon \ll 1$) this yields the bound

$$0 \leq \kappa \leq 2\epsilon \tag{6.23}$$

and hence one of the main results of this paper [46]

$$\langle\chi^2\rangle = p - 1 + \kappa\langle z\rangle^2 \quad \text{with} \quad 0 \leq \kappa \leq 2\epsilon. \tag{6.24}$$

This relationship *only* assumes that the fitting factor between the signal waveforms T and T' is close to 1.

We can obtain a different and tighter bound on κ if we assume that the frequency bands defined by (4.1) and (4.4) are the *same* for the waveforms T and T'. For example, as discussed earlier in the context of Eq. (4.5), this is true

for binary inspiral waveforms in the stationary-phase approximation. In this case, since $(\tilde{T}', \tilde{T}')_j = 1/p$, Eq. (6.20) implies that

$$\lambda_j^2 \cos^2\theta \leq \frac{1}{p^2}. \tag{6.25}$$

Thus one has

$$\frac{-1}{p\cos\theta} \leq \lambda_j \leq \frac{1}{p\cos\theta}. \tag{6.26}$$

It is convenient to define $\omega_j \equiv 1/p - \lambda_j$. In terms of these quantities

$$\kappa = p \sum_{j=1}^{p} \omega_j^2. \tag{6.27}$$

The values of ω_j are constrained by two relations:

$$\sum_{j=1}^{p} \omega_j = 0, \tag{6.28}$$

and

$$\frac{1}{p}\left(1 - \frac{1}{\cos\theta}\right) \leq \omega_j \leq \frac{1}{p}\left(1 + \frac{1}{\cos\theta}\right). \tag{6.29}$$

If we assume that the number of frequency bands $p > 2$ and that the signal/template mismatch

$$\epsilon = 1 - \cos\theta \leq \frac{2}{p} \tag{6.30}$$

then the maximum of κ is obtained when

$$\omega_1 = -\frac{p-1}{p}\left(1 - \frac{1}{\cos\theta}\right), \quad \text{and}$$

$$\omega_2 = \cdots = \omega_p = \frac{1}{p}\left(1 - \frac{1}{\cos\theta}\right). \tag{6.31}$$

The upper bound on κ is

$$\kappa \leq (p-1)\left(\frac{1}{\cos\theta} - 1\right)^2 \sim (p-1)\epsilon^2, \tag{6.32}$$

where in the final part of the relation we assume as before that the mismatch $\epsilon \ll 1$. This is one of the other main results of the paper. In the case where the signal waveform has the same bands Δf_j as the closest matching template, and $p \leq 2/\epsilon$, and ϵ is small, one has

$$\langle\chi^2\rangle = p - 1 + \kappa\langle z\rangle^2 \quad \text{with} \quad 0 \leq \kappa \leq (p-1)\epsilon^2. \tag{6.33}$$

This result is beautifully consistent with the previous limit (6.24) on $\langle\chi^2\rangle$. At the boundary of validity (6.30) of (6.32), one has
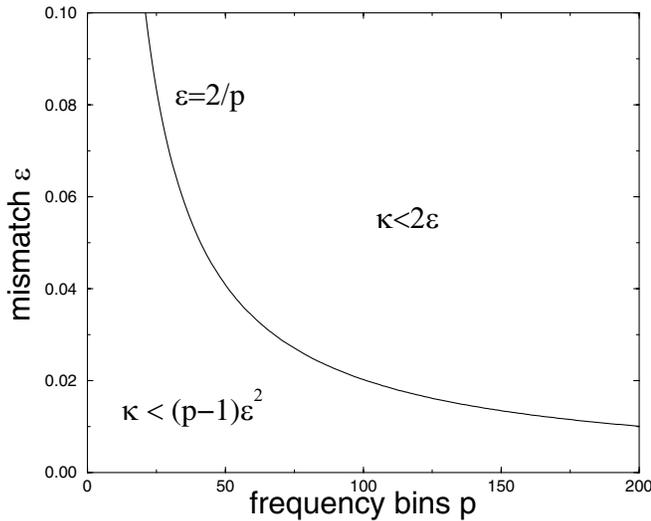
FIG. 3. The expected value of χ^2 for a signal that is not matched to the template satisfies $\langle \chi^2 \rangle = p - 1 + \kappa \langle z \rangle^2$ where $\langle z \rangle$ is the expected SNR. If signal and template share the same set of frequency bands, then below the curve $\epsilon = 2/p$ one has $\kappa < (p-1)\epsilon^2$. Above the curve $\kappa < 2\epsilon$. If signal and template do *not* share the same frequency bands, then $\kappa < 2\epsilon$ is the only limit that applies.

$$\kappa = (p-1)\left(\frac{1}{\cos\theta} - 1\right)^2 = \frac{1}{\cos^2\theta} - 1 \sim 2\epsilon \quad (6.34)$$

which agrees *exactly* with the previous limit (6.22). The limits that apply are summarized in Fig. 3. Note that when the two different limits (6.22) and (6.32) are expressed approximately to lowest order in terms of ϵ, they appear to differ slightly at the boundary $p = 2/\epsilon$. In fact they agree exactly: the approximate expressions differ at higher order in ϵ.

Appendix A shows that if the detector noise is Gaussian, then the χ^2 statistic (computing with a perfectly matching template) has a classical χ^2-distribution with $p - 1$ degrees of freedom. If the template does *not* match perfectly, the distribution becomes a noncentral χ^2 distribution, with a noncentrality parameter determined by the r.m.s. value of (6.15), which is $\kappa \langle z \rangle^2$. This is discussed in more detail at the end of Sec. VII. It follows from the fact that the variance of the terms that appear in the definition of χ^2,

$$\langle (\Delta z_j)^2 \rangle - \langle \Delta z_j \rangle^2 = \frac{1}{p}\left(1 - \frac{1}{p}\right) \quad (6.35)$$

are independent of the signal amplitude.

VII. SIGNAL OF UNKNOWN PHASE

As mentioned earlier, the signal from an inspiraling binary pair is a linear combination of two possible gravitational waveforms with an *a priori* unknown phase ϕ.

Here, we repeat the analysis done in the previous three sections, for this particular case of interest.

As in (2.6), the detector output is assumed to be of the form

$$s(t) = n(t) + \frac{D}{d}[\cos\phi T_c(t - t_0) + \sin\phi T_s(t - t_0)],$$

with $n(t)$ a random time-series drawn from a distribution appropriate to the detector noise, and both ϕ and d unknown. We assume (as is the case for binary inspiral) that the templates are orthonormal so that

$$(\tilde{T}_c, \tilde{T}_c) = (\tilde{T}_s, \tilde{T}_s) = 1, \quad (7.1)$$

and

$$(\tilde{T}_c, \tilde{T}_s) = 0. \quad (7.2)$$

Note that in the stationary-phase approximation, T_c and T_s are exactly orthogonal. Were they not, a Gram-Schmidt procedure could be used to construct an orthonormal pair spanning the same space of signals [47].

There are several (easy) ways to efficiently search for the unknown phase ϕ. In substance, all of these methods consist of filtering *separately* with the two templates T_c and T_s, and then combining the two filtered data streams. For our purposes a nice way to do this is to combine these separate (real) filter outputs into a single complex signal. Thus, we use the optimal filter

$$\tilde{Q} = (\tilde{T}_c + i\tilde{T}_s)e^{-2\pi i f t_0}. \quad (7.3)$$

Note that with this normalization the optimal filter is normalized so that $(\tilde{Q}, \tilde{Q}) = 2$.

The output of the filter is complex and is

$$\begin{aligned} z &= (\tilde{Q}, \tilde{s}) \\ &= (\tilde{Q}, \tilde{n}) + \left(\tilde{Q}, \frac{D}{d}[\cos\phi \tilde{T}_c + \sin\phi \tilde{T}_s]e^{-2\pi i f t_0}\right). \end{aligned}$$

Its expectation value is the complex number

$$\langle z \rangle = \frac{D}{d}(\cos\phi + i\sin\phi) = \frac{D}{d}e^{i\phi}. \quad (7.4)$$

The modulus of this complex number is the (expected) inverse distance, and its phase is the (expected) phase. Note that because the normalization of $\tilde{Q}$ has changed, the expected value

$$\langle |z|^2 \rangle = 2 + \left(\frac{D}{d}\right)^2. \quad (7.5)$$

is larger than in the single-phase case. The additional uncertainty about the phase ϕ means that the distance to the source can not be determined as accurately as in the single-phase case. Following conventional practice in the field, the modulus $|z|$ will be called the "Signal-to Noise Ratio" (SNR) although since in the absence of a source its

mean-square value is two, one might argue that $|z|/\sqrt{2}$ is the quantity that should legitimately carry this name.

To construct the χ^2 statistic, we choose frequency bands as before. We will assume that $\tilde{T}_c$ and $\tilde{T}_s$ have *identical* frequency bands and are orthogonal in each of these bands [48]. For inspiral sources, this is exactly true in the stationary-phase approximation where $\tilde{T}_s(f) = i\tilde{T}_c(f)$ for $f > 0$ and $\tilde{T}_s(f) = -i\tilde{T}_c(f)$ for $f < 0$. Thus

$$(\tilde{T}_c, \tilde{T}_c)_j = (\tilde{T}_s, \tilde{T}_s)_j = \frac{1}{p}, \tag{7.6}$$

and

$$(\tilde{T}_c, \tilde{T}_s)_j = 0. \tag{7.7}$$

We define the complex signal z_j in the j'th band as before

$$z_j \equiv (\tilde{Q}, \tilde{s})_j$$

and also define Δz_j as before

$$\Delta z_j \equiv z_j - \frac{z}{p}. \tag{7.8}$$

One then finds

$$\langle z_j \rangle = \frac{1}{p}\frac{D}{d}e^{i\phi}\cos\theta \qquad \langle |z_j|^2 \rangle = \frac{2}{p} + \frac{1}{p^2}\left(\frac{D}{d}\right)^2$$
$$\langle z_j^* z \rangle = \frac{2}{p} + \frac{1}{p}\left(\frac{D}{d}\right)^2 \qquad \langle |\Delta z_j|^2 \rangle = \frac{2}{p}\left(1 - \frac{1}{p}\right). \tag{7.9}$$

The χ^2 statistic is defined by [49]

$$\chi^2 = p\sum_{j=1}^{p}|\Delta z_j|^2 \tag{7.10}$$

and thus from (7.9) has expected value

$$\langle \chi^2 \rangle = 2p - 2. \tag{7.11}$$

In Appendix A we show that if the detector noise is Gaussian, then χ^2 has a classical χ^2 probability distribution. Because both the real *and* imaginary parts of Δz_j sum to zero, the number of (real) degrees of freedom is $2p - 2$.

We now consider the case where the astrophysical waveform $h(t) = \frac{D'}{d'}T'(t)$ does *not* exactly match any linear combination of the templates T_c and T_s. This is to be expected from real signals if the templates form a discrete finite grid in parameter space. As before, with no loss of generality we assume that $t_0 = 0$ and that D' is chosen so that

$$(\tilde{T}', \tilde{T}') = 1.$$

Consider the possible values, as $\psi \in [0, 2\pi)$ varies, of the inner product

$$(\cos\psi\tilde{T}_c + \sin\psi\tilde{T}_s, \tilde{T}').$$

Since both $\cos\psi\tilde{T}_c + \sin\psi\tilde{T}_s$ and $\tilde{T}'$ are unit length,

Schwartz's inequality implies that this inner product lies in the range $[-1, 1]$, and its maximum value must lie in the range $[0, 1]$. This maximum value (see Appendix B) is

$$\cos\theta \equiv \sqrt{(\tilde{T}_c, \tilde{T}')^2 + (\tilde{T}_s, \tilde{T}')^2}, \tag{7.12}$$

which defines $\theta \in [0, \pi/2]$. This in turn defines the mismatch $\epsilon = 1 - \cos\theta$ between the template T' and the one-parameter family of templates. Note that (in contrast to the single-phase case) the maximization over ψ *automatically* leads to $\cos\theta \in [0, 1]$ and hence $\epsilon \in [0, 1]$. We can also define $\phi \in [0, 2\pi)$ by

$$\cos\phi\cos\theta \equiv (\tilde{T}_c, \tilde{T}'), \quad \text{and} \quad \sin\phi\cos\theta \equiv (\tilde{T}_s, \tilde{T}'). \tag{7.13}$$

Thus one has

$$(\tilde{T}_c + i\tilde{T}_s, \tilde{T}') = e^{i\phi}\cos\theta. \tag{7.14}$$

This equation may be taken as the *definition* of ϕ and θ.

The filter output is given by

$$z = (\tilde{Q}, \tilde{s}) = (\tilde{Q}, \tilde{n} + \tilde{h})$$

and thus has expectation value

$$\langle z \rangle = \frac{D'}{d'}(\tilde{T}_c + i\tilde{T}_s, \tilde{T}') = \frac{D'}{d'}e^{i\phi}\cos\theta. \tag{7.15}$$

The expected square modulus of the filter output is

$$\langle |z|^2 \rangle = 2 + \left(\frac{D'}{d'}\right)^2\cos^2\theta = 2 + |\langle z \rangle|^2.$$

We now investigate the effect of the template/signal mismatch on the χ^2 statistic.

To begin, we need to characterize the overlap between the signal and the templates in the j'th frequency band. Define complex quantities λ_j by

$$(\tilde{T}_c + i\tilde{T}_s, \tilde{T}')_j = \lambda_j e^{i\phi}\cos\theta. \tag{7.16}$$

Using (7.14), these complex quantities are constrained by

$$\sum_{j=1}^{p}\lambda_j = 1. \tag{7.17}$$

The filter output in the j'th frequency band is given by

$$z_j = (\tilde{T}_c + i\tilde{T}_s, \tilde{n})_j + \frac{D'}{d'}\lambda_j e^{i\phi}\cos\theta.$$

and the various expectation values in the j'th band are

$$\langle z_j \rangle = \frac{D'}{d'} \lambda_j e^{i\phi} \cos\theta,$$

$$\langle |z_j|^2 \rangle = \frac{2}{p} + \left(\frac{D'}{d'}\right)^2 |\lambda_j|^2 \cos^2\theta,$$

$$\langle z_j^* z \rangle = \frac{2}{p} + \left(\frac{D'}{d'}\right)^2 \lambda_j^* \cos^2\theta,$$

$$\langle |\Delta z_j|^2 \rangle = \frac{2}{p}\left(1 - \frac{1}{p}\right) + \left(\frac{D'}{d'}\right)^2 \left|\lambda_j - \frac{1}{p}\right|^2 \cos^2\theta.$$

The expected value of χ^2 is then

$$\langle \chi^2 \rangle = 2p - 2 + \kappa |\langle z \rangle|^2 \qquad (7.18)$$

with

$$\kappa \equiv p \sum_{j=1}^{p} \left|\lambda_j - \frac{1}{p}\right|^2 = -1 + p \sum_{j=1}^{p} |\lambda_j^2|. \qquad (7.19)$$

To place an upper limit on κ, note that from Schwartz's inequality, for *any* value of the angle ψ, one has [50]

$$(\cos\psi \tilde{T}_c + \sin\psi \tilde{T}_s, \tilde{T}')_j^2 \le \frac{1}{p}(\tilde{T}', \tilde{T}')_j,$$

$$[\cos\psi (\tilde{T}_c, \tilde{T}')_j + \sin\psi (\tilde{T}_s, \tilde{T}')_j]^2 \le \frac{1}{p}(\tilde{T}', \tilde{T}')_j.$$

The *maximum* value of the left-hand side as 4 varies (see Appendix B) is

$$(\tilde{T}_c, \tilde{T}')_j^2 + (\tilde{T}_s, \tilde{T}')_j^2 =$$
$$(\Re[\lambda_j e^{i\phi}] \cos\theta)^2 + (\Im[\lambda_j e^{i\phi}] \cos\theta)^2 =$$
$$|\lambda_j|^2 \cos^2\theta,$$

where we have made use of (7.13) and (7.14). Thus

$$|\lambda_j|^2 \le \frac{1}{p\cos^2\theta}(\tilde{T}', \tilde{T}')_j.$$

Summing both sides over j and making use of (7.19) this implies that

$$0 \le \kappa \le \frac{1}{\cos^2\theta} - 1. \qquad (7.20)$$

This result makes no assumptions about the form of the mismatched signal $\tilde{T}'$. As in the single-phase case, if we assume that the mismatch is small, $\epsilon \ll 1$, we obtain

$$0 \le \kappa \le 2\epsilon. \qquad (7.21)$$

This result does not assume any relationship between the frequency bands of the signals T and T'.

If we assume that the bands Δf_j for the mismatched signal T' are the *same* as those for the templates T_c and T_s then we can obtain a much stronger upper bound. For example, this is the case if all three templates are drawn from a family of stationary-phase approximate inspiral chirps. In this case, $(\tilde{T}', \tilde{T}')_j = 1/p$ and the same logic as in the single-phase case can be used to establish that

$$\kappa \le (p - 1)\left(\frac{1}{\cos\theta} - 1\right)^2 \sim (p - 1)\epsilon^2, \qquad (7.22)$$

provided that $p > 2$ and $\epsilon = 1 - \cos\theta \le 2/p$. Thus, with some minor modifications, all the single-phase results apply to the unknown phase case.

In the case where the signal and template are not a perfect match, the expected value of Δz_j does not vanish:

$$\langle \Delta z_j \rangle = \frac{D'}{d'} e^{i\phi} \cos\theta\left(\lambda_j - \frac{1}{p}\right).$$

If the detector noise is stationary and Gaussian, then for a given astrophysical signal T' and filter-template $T_{c,s}$ this means that the probability distribution of χ^2 is a classical noncentral χ^2 distribution [51] with $2p - 2$ degrees of freedom and noncentrality parameter

$$\lambda = p \sum_{j=1}^{p} |\langle \Delta z_j \rangle|^2 = \kappa |\langle z \rangle|^2.$$

Unfortunately, for a set of candidate events, which correspond to different waveforms $D'T'/d'$ and ring off different templates T, the values of κ have different, unknown values, bounded only by (7.20) or (7.22). In this case, since the average of noncentral χ^2 distributions with different values of λ is *not* a noncentral χ^2 distribution, one can only bound the expected distribution, not determine it from first principles.

VIII. THRESHOLDING CONDITIONS

As described in Sec. I, the χ^2 time-frequency discriminator is most often used as a veto. For a given data set, the threshold value χ_*^2 is usually determined using Monte-Carlo simulation of signals, analytic guidance, and experience. If the signal and template were known to have identical form, then the threshold χ_*^2 would be a number. However since the signal and template are not expected to match perfectly, the threshold χ_*^2 is a function of the observed SNR.

It is helpful to understand the threshold that would be appropriate for stationary Gaussian noise. In this case, the optimal threshold is given by the inverse of the noncentral χ^2 cumulative distribution function [52]. A putative signal whose χ^2 value is smaller than this "Gaussian noise" threshold is likely to merit further examination, even if the noise is *not* Gaussian [53]. So in most cases a reasonable threshold will be greater than or equal to the threshold appropriate for Gaussian detector noise.

In the stationary Gaussian case, for fixed T and T', the expected value, variance σ^2, and standard deviation σ of the noncentral χ^2 distribution with $2p - 2$ degrees of freedom are [54]

$$\langle \chi^2 \rangle = 2p - 2 + \lambda = 2p - 2 + \kappa |\langle z \rangle|^2,$$
$$\sigma^2 = \langle (\chi^2)^2 \rangle - \langle \chi^2 \rangle^2 = 4p - 4 + 4\lambda, \quad \text{and}$$

$$\sigma = \sqrt{4p - 4 + 4\kappa |\langle z \rangle|^2}. \tag{8.1}$$

In the neighborhood of the maximum, for values of the noncentrality parameter λ significantly larger than $2p - 2$, the noncentral χ^2 distribution is approximately a Gaussian of width σ, centered about the mean value $2p - 2 + \lambda$. In this case, the optimal χ^2 veto threshold for Gaussian noise is well-approximated by

$$\chi_*^2 = \langle \chi^2 \rangle + \text{few}\sigma,$$

where σ are the expected statistical fluctuations in χ^2 evaluated for the "worst-case" value of κ. If we assume that the putative signals and templates do not share the same frequency bands Δf_j, so that the upper limit of (7.20) applies, then we obtain a χ^2 threshold of the form

$$\chi_*^2 = 2p - 2 + 2\epsilon |\text{SNR}|^2 + \text{few}\sqrt{4p - 4 + 8\epsilon |\text{SNR}|^2}, \tag{8.2}$$

where we have replaced the expected SNR by the measured SNR [55]. Although we have justified this approximation to the threshold for large noncentrality parameter λ, it turns out to be a reasonably good approximation even when the noncentrality parameter λ is small [56].
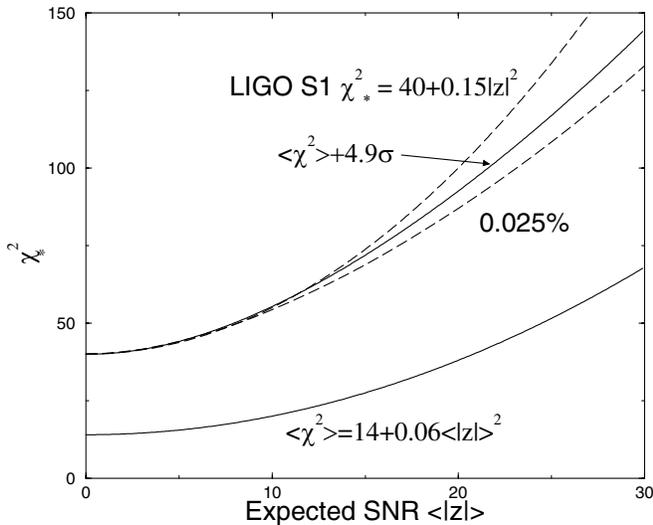

FIG. 4. A comparison of different thresholds χ_*^2 for the χ^2 time-frequency discriminator, for the LIGO S1 two-phase case with $p = 8$, $\epsilon = 0.03$, and worst-case $\kappa = 2\epsilon$. The bottom solid line shows the expected value of χ^2 for stationary Gaussian noise given by (8.1). The top solid line shows a threshold (8.4) set 4.9σ above this expected value. The upper dashed line shows the heuristic threshold (8.3) used for the LIGO S1 analysis, which was determined from Monte-Carlo studies. The lower dashed line shows the threshold which would be exceeded with probability 0.025% in stationary Gaussian noise.

It is interesting to compare the threshold appropriate for Gaussian noise to the χ^2 threshold used in the LIGO S1 analysis for $p = 8$ and $\epsilon = 0.03$, which is equation (4.7) of Ref. [9]

$$\text{LIGO S1} \chi_*^2 = 40 + 0.15|\text{SNR}|^2, \tag{8.3}$$

shown as the upper dashed line in Fig. 4. The lower solid curve shows the expectation value of χ^2 given by (7.18), and the upper solid curve shows a 4.9 standard deviation threshold given by (8.2), which is

$$\chi_*^2 = 14 + 0.06|\text{SNR}|^2 + 4.9\sqrt{28 + 0.24|\text{SNR}|^2}. \tag{8.4}$$

As is clear from the graph, the heuristic threshold is reasonably well matched by the sort of threshold that one might set based on a worst-case analysis for Gaussian detector noise.

For *very* large SNR, the Gaussian threshold condition (8.2) consists of two terms. The dominant term (quadratic in SNR) comes from the mean value $\langle \chi^2 \rangle$ and has coefficient exactly κ. The subdominant term (linear in SNR) comes from a few times σ. Hence, a threshold like the LIGO S1 choice would *not* veto high SNR events that *could* be confidently vetoed in Gaussian noise. This is illustrated in Fig. 5.

In the LIGO S1 analysis, which sets an upper limit on the Galactic inspiral rate, the probability of observing a close inspiral (very large SNR) is far smaller than the probability of observing a more distant (low SNR) event from near the Galactic center. Thus, applying the more
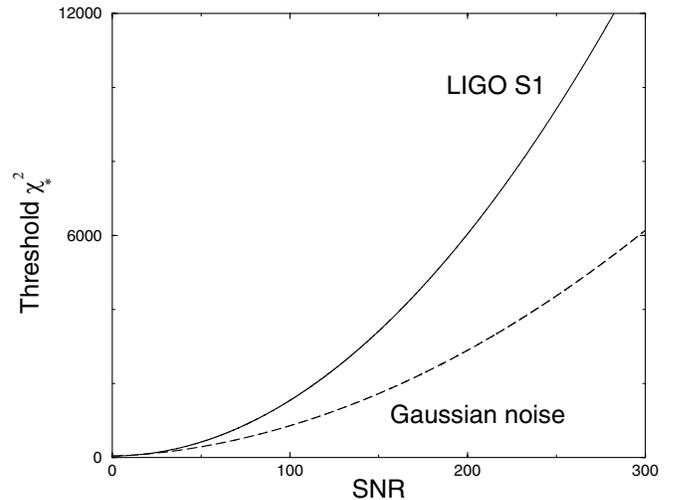

FIG. 5. A comparison of different thresholds for the χ^2 time-frequency discriminator, at large SNR. The upper solid curve is the LIGO S1 threshold (8.3). The lower (dashed) curve is the threshold that would be exceeded with probability 0.025% in stationary Gaussian noise. (On the scale of this graph the approximation (8.4) to the dashed curve is difficult to distinguish from the exact result obtained from the inverse of the noncentral χ^2 distribution.)

stringent thresholding condition at high SNR would probably not have had a significant detrimental effect on the analysis: it would not have significantly decreased the detection efficiency. However it also would not have improved the analysis, since the highest SNR events that passed the χ^2 threshold had SNR less than 16.

IX. UNEQUAL EXPECTED SNR FREQUENCY INTERVALS

One may also define a χ^2 time-frequency discriminator using frequency intervals which do *not* make equal expected contributions to the SNR, and almost all of the previous results hold [58]. For simplicity, in this Section we treat only the single-phase case.

Begin by making the same initial assumptions as in Sec. VI, but choose frequency intervals which do *not* make equal expected contributions to the SNR. Thus

$$\text{choose} \quad \Delta f_j \quad \text{so that} \ (\tilde{T}, \tilde{T})_j = q_j, \qquad (9.1)$$

where the $q_j \in (0, 1]$ do not necessarily equal $1/p$. Template normalization $(\tilde{T}, \tilde{T}) = 1$ implies that they satisfy $\sum_j q_j = 1$.

Define the SNR in the j'th band as previously

$$z_j = (\tilde{Q}, \tilde{s})_j, \qquad (9.2)$$

and define Δz_j as the difference between observed SNR in the j'th band, and the value that would be *anticipated* [59] based on the total SNR observed. Thus we define

$$\Delta z_j \equiv z_j - q_j z, \qquad (9.3)$$

where as before the observed SNR is $z = \sum_j z_j = (\tilde{Q}, \tilde{s})$. Note that by definition the sum $\sum_j \Delta z_j$ vanishes.

Consider the case of a signal waveform T' which may be mismatched to the template T, so $(T, T') = \cos\theta$ and (without loss of generality) $0 \leq \cos\theta \leq 1$. Within the j'th band, the template T' has overlap

$$(\tilde{T}, \tilde{T}')_j = \lambda_j \cos\theta, \qquad (9.4)$$

which may be taken to define the quantities λ_j. As before, the sum $\sum_j \lambda_j = 1$. Taking the signal to be

$$s(t) = n(t) + \frac{D'}{d'} T'(t), \qquad (9.5)$$

we can now compute the various expectation values.

It follows immediately from the above definitions that these expectation values are given by

$$\langle z \rangle = \frac{D'}{d'} \cos\theta, \qquad \langle z_j \rangle = \frac{D'}{d'} \lambda_j \cos\theta,$$

$$\langle \Delta z_j \rangle = \frac{D'}{d'} (\lambda_j - q_j) \cos\theta,$$

$$\langle z^2 \rangle = 1 + \left(\frac{D'}{d'}\right)^2 \cos^2\theta,$$

$$\langle z_j z_k \rangle = q_j \delta_{jk} + \left(\frac{D'}{d'}\right)^2 \lambda_j \lambda_k \cos^2\theta, \qquad (9.6)$$

$$\langle z_j^2 \rangle = q_j + \left(\frac{D'}{d'}\right)^2 \lambda_j^2 \cos^2\theta, \quad \text{and}$$

$$\langle (\Delta z_j)^2 \rangle = q_j(1 - q_j) + \left(\frac{D'}{d'}\right)^2 (\lambda_j - q_j)^2 \cos^2\theta,$$

where (the Kronecker symbol) $\delta_{jk} = 1$ if $j = k$ and vanishes otherwise. The formulae of Sec. VI correspond to the special case in which $q_j = 1/p$.

Define the χ^2 time-frequency discriminator in the unequal expected SNR interval case to be

$$\chi^2 = \sum_{j=1}^{p} (\Delta z_j)^2 / q_j. \qquad (9.7)$$

To characterize the statistical properties of χ^2, it is helpful to express it in terms of a different set of variables. Begin by writing χ^2 as

$$\chi^2 = \sum_{j=1}^{p} (z_j - q_j z) \Delta z_j / q_j = \sum_{j=1}^{p} z_j \Delta z_j / q_j$$

$$= \sum_{j=1}^{p} z_j^2 / q_j - \left(\sum_{j=1}^{p} z_j\right)^2, \qquad (9.8)$$

where we have made use of the fact that the Δz_j sum to zero. Define new variables $u_j = z_j / \sqrt{q_j}$, which have variance unity and are uncorrelated:

$$\langle u_j u_k \rangle - \langle u_j \rangle \langle u_k \rangle = \delta_{jk},$$

In terms of these variables, the statistic is

$$\chi^2 = \sum_{j=1}^{p} u_j^2 - \left(\sum_{j=1}^{p} \sqrt{q_j} \, u_j\right)^2. \qquad (9.9)$$

To characterize the probability distribution of χ^2, it is convenient to change variables once again.

We introduce new variables v_j which are linear combinations of the u_j and are most conveniently written in matrix form as

$$\begin{bmatrix} v_1 \\ \vdots \\ v_p \end{bmatrix} = M \begin{bmatrix} u_1 \\ \vdots \\ u_p \end{bmatrix} \qquad (9.10)$$

where M is a $p \times p$ square matrix. Choose M to be an orthogonal matrix, which thus satisfies $M^t M = M M^t = I$, where t denotes transpose and I is the $p \times p$ identity

matrix. These linear transformations (9.10) are rotations [60] that map $\mathbb{R}^p \to \mathbb{R}^p$.

Since the transformation is orthogonal, the new variables v_j also have variance unity and are uncorrelated:

$$
\begin{aligned}
\langle v_j v_k \rangle - \langle v_j \rangle \langle v_k \rangle &= \sum_{\ell=1}^{p} \sum_{m=1}^{p} M_{j\ell} M_{km} [\langle u_\ell u_m \rangle - \langle u_\ell \rangle \langle u_m \rangle] \\
&= \sum_\ell \sum_m M_{j\ell} M_{km} \delta_{\ell m} \\
&= \sum_m M_{jm} M_{km} \\
&= \sum_m M_{jm} (M^t)_{mk} \\
&= (MM^t)_{jk} = I_{jk} = \delta_{jk}.
\end{aligned}
\quad (9.11)
$$

Moreover

$$
\sum_{j=1}^{p} u_j^2 = \sum_{j=1}^{p} v_j^2, \quad (9.12)
$$

since rotations do not change the length of a vector.

The rotation M may be chosen so that any given orthonormal basis is mapped onto any other orthonormal basis of the same orientation (handedness). Thus one may choose the rotation so that the last of the new variables is

$$
v_p = \sqrt{q_1} u_1 + \sqrt{q_2} u_2 + \cdots \sqrt{q_p} u_p. \quad (9.13)
$$

This corresponds to constraining the final row of M to be $(\sqrt{q_1}, \cdots, \sqrt{q_n})$, or equivalently to requiring that M map the vector on the left hand side (l.h.s.) below to the final (p'th) basis vector.

$$
\begin{bmatrix} \sqrt{q_1} \\ \vdots \\ \sqrt{q_n} \end{bmatrix} \xrightarrow{M} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}
$$

In terms of these new variables, χ^2 given in (9.9) may be written using (9.12) and (9.13) as

$$
\chi^2 = \left(\sum_{j=1}^{p} v_j^2 \right) - v_p^2 = \sum_{j=1}^{p-1} v_j^2. \quad (9.14)
$$

This form makes it easy to characterize the statistics of χ^2 if the detector noise is Gaussian [61].

If the noise is Gaussian, then the u_j are uncorrelated (and hence independent) Gaussian random variables with unit variance, and the v_k are also uncorrelated (and hence independent) Gaussian random variables with unit variance. Thus, in the case of Gaussian noise it immediately follows from (9.14) that χ^2 has a classical noncentral χ^2 distribution with $p - 1$ degrees of freedom. In the two-phase case, each of the v_j is a complex variable with independent real and imaginary parts and the resulting distribution has $2p - 2$ degrees of freedom.

The noncentrality parameter λ may be evaluated by calculating the expected value of χ^2. Using (9.6) and (9.7) this is

$$
\langle \chi^2 \rangle = p - 1 + \kappa \left(\frac{D'}{d'} \right)^2 \cos^2\theta = p - 1 + \kappa \langle z \rangle^2, \quad (9.15)
$$

and hence the noncentrality parameter is given by $\lambda = \kappa \langle z \rangle^2$. The constant κ, which is determined by the choice of intervals, the spectrum of the detector noise, and the frequency-dependence of the mismatch between templates, is given by

$$
\begin{aligned}
\kappa &= \sum_{j=1}^{p} (\lambda_j - q_j)^2 / q_j = \sum_{j=1}^{p} (\lambda_j^2 / q_j - 2\lambda_j + q_j) \\
&= -1 + \sum_{j=1}^{p} \lambda_j^2 / q_j.
\end{aligned}
\quad (9.16)
$$

Clearly κ is non-negative. One can easily obtain an upper limit on κ even in this case where the frequency intervals are not "equal SNR" intervals.

To obtain a limit on κ, begin with Schwartz's inequality, which implies that

$$
(\tilde{T}, \tilde{T}')_j^2 \le (\tilde{T}, \tilde{T})_j (\tilde{T}', \tilde{T}')_j \lambda_j^2 \cos^2\theta \le q_j (\tilde{T}', \tilde{T}')_j \Rightarrow \lambda_j^2 / q_j
$$

$$
\le \frac{1}{\cos^2\theta} (\tilde{T}', \tilde{T}')_j. \quad (9.17)
$$

Summing both sides over j and using (9.16) yields

$$
0 \le \kappa \le \frac{1}{\cos^2\theta} - 1, \quad (9.18)
$$

and hence for $\epsilon \ll 1$ one has $0 \le \kappa \le 2\epsilon$, just as in the "equal SNR interval" case.

One can also establish a stronger limit analogous to (6.32) for the case where the templates T and T' have the same values of q_j for a given set of frequency intervals. In this case, Schwartz's inequality implies that

$$
(\tilde{T}, \tilde{T}')_j^2 \le (\tilde{T}, \tilde{T})_j (\tilde{T}', \tilde{T}')_j, \qquad \lambda_j^2 \cos^2\theta \le q_j^2,
$$

and hence that

$$
-\frac{q_j}{\cos\theta} \le \lambda_j \le \frac{q_j}{\cos\theta}. \quad (9.19)
$$

Without loss of generality, relabel the frequency intervals so that

$$
q_1 \le q_2 \le \cdots \le q_p.
$$

The value of κ is maximized by setting:

$$
\begin{aligned}
\lambda_1 &= 1 + (q_1 - 1)/\cos\theta \\
\lambda_2 &= q_2/\cos\theta \\
&\cdots \\
\lambda_p &= q_p/\cos\theta.
\end{aligned}
\quad (9.20)
$$

This choice satisfies the constraint that $\sum_j \lambda_j = 1$, and the

r.h.s. of (9.19). In order that λ_1 satisfy the l.h.s. of the constraint (9.19) we need to have

$$1 - \cos\theta < 2q_1 \qquad (9.21)$$

or equivalently $\epsilon < 2q_1$. For the values of λ_j given in (9.20) one then obtains

$$
\begin{aligned}
\kappa &= \sum_{j=1}^{p} (\lambda_j - q_j)^2 / q_j \\
&= [(q_1 - 1)/\cos\theta + 1 - q_1]^2/q_1 + q_2\left(\frac{1}{\cos\theta} - 1\right)^2 \\
&\quad + \cdots + q_p\left(\frac{1}{\cos\theta} - 1\right)^2 \\
&= [(q_1 - 1)^2/q_1 + q_2 + \cdots + q_p]\left(\frac{1}{\cos\theta} - 1\right)^2 \\
&= \left[\frac{1}{q_1} - 1\right]\left(\frac{1}{\cos\theta} - 1\right)^2.
\end{aligned}
$$

For $p > 2$ this gives the upper bound on κ.

Denote the smallest value of q_j (q_1 just above) by q_{min}. Then, in the case where the templates T and T' have the same values of q_j for a given set of frequency intervals, and $1 - \cos\theta < 2q_{min}$ one has

$$0 \le \kappa \le \left(\frac{1}{q_{min}} - 1\right)\left(\frac{1 - \cos\theta}{\cos\theta}\right)^2. \qquad (9.22)$$

If $\epsilon = 1 - \cos\theta$ is much less than unity, then this may be written

$$0 \le \kappa \le \left(\frac{1}{q_{min}} - 1\right)\epsilon^2 \quad \text{if } \epsilon < 2q_{min}. \qquad (9.23)$$

These reduce to the previous results of Sec. VI when all the q_j (and hence q_{min}) equal $1/p$.

This "unequal expected SNR" χ^2 discriminator may be of practical use when it is impossible to construct equal SNR intervals. It may also permit the construction of discriminators which are specifically tuned to common types of detector noise.

X. OTHER TYPES OF χ^2 TESTS

There are many possible χ^2 tests that could be used to discriminate spurious signals from genuine ones. Here we compare the χ^2 time-frequency discriminator of this work with a standard χ^2 test used by Baggio *et al.* [37], which we denote by $\bar{\chi}^2$. This tests "goodness of fit" for a modeled signal embedded in stationary Gaussian noise, and is used in analyzing data from AURIGA, a narrow band resonant-bar gravitational wave detector.

It is instructive to express this standard test in the notation of this paper. $\bar{\chi}^2$ is constructed in two steps. First one picks a time interval $[t_1, t_1 + \tau]$ whose length τ is not less than the time duration of the signal-model template T convolved with S_n^{-1}, and which includes the support of $S_n^{-1} * T$ [62].

Then one constructs a function of a single real amplitude A (or, potentially, additional parameters describing the signal)

$$\bar{\chi}^2(A) = (\tilde{s} - A\tilde{T}, \tilde{s} - A\tilde{T}) = (\tilde{s}, \tilde{s}) - 2A(\tilde{s}, \tilde{T}) + A^2. \qquad (10.1)$$

Here, $\tilde{s}$ is computed from (2.1), but the integral is taken *only* over the time interval $[t_1, t_1 + \tau]$. As a function of A, $\bar{\chi}^2$ has an absolute minimum at

$$A = (\tilde{s}, \tilde{T}) = z.$$

The minimum value of $\bar{\chi}^2(A)$ defines $\bar{\chi}^2$:

$$\bar{\chi}^2 = (\tilde{s}, \tilde{s}) - (\tilde{s}, \tilde{T})^2 = (\tilde{n}, \tilde{n}) - (\tilde{n}, \tilde{T})^2. \qquad (10.2)$$

Thus, $\bar{\chi}^2$ measures the difference between the squared amplitude of the detector output and the squared SNR. It is clear from this equation that $\bar{\chi}^2$ is quite different from the χ^2 discriminator defined in this paper. In particular, if the detected SNR vanishes ($z = 0$) then $\bar{\chi}^2 = (\tilde{s}, \tilde{s})$, whereas $\chi^2 = \sum_{j=1}^{p}(\tilde{s}, \tilde{T})_j^2$. In this case $\bar{\chi}^2$ is measuring the "total length" of $\tilde{s}$, while χ^2 is measuring the sum of squares of the components of $\tilde{s}$ obtained by projecting it onto p orthonormal components of $\tilde{T}$.

It is also instructive to compute the expected value of $\bar{\chi}^2$ in our frequency-domain-based formalism [63]. Denote the instrument's data acquisition sample time by Δt, so that the Nyquist frequency is $f_N = 1/(2\Delta t)$ and the number of data samples is $N = \tau/\Delta t$. After setting $s(t)$ to zero outside of the time interval $[t_1, t_1 + \tau]$, one has

$$
\begin{aligned}
\langle |\tilde{n}(f)|^2 \rangle &= \int_0^\tau dt \int_0^\tau dt' \langle n(t + t_1)n(t' + t_1)\rangle e^{2\pi i f(t - t')} \\
&= \int df' \int_0^\tau dt \int_0^\tau dt' S_n(f') e^{2\pi i(f - f')(t - t')} \\
&= \int df' \int_0^\tau dt \int_{-\infty}^{\infty} dt' S_n(f') e^{2\pi i(f - f')(t - t')} \\
&= \int df' \int_0^\tau dt S_n(f') \delta(f - f') e^{2\pi i(f - f')t} \\
&= \int_0^\tau dt S_n(f) = \tau S_n(f). \qquad (10.3)
\end{aligned}
$$

In going from the second to the third line, we have assumed that τ is greater than the characteristic time over which the autocorrelation function of the noise falls off. From (10.3) it follows immediately that

$$\langle (\tilde{n}, \tilde{n}) \rangle = \int_{-f_N}^{f_N} \frac{\langle |\tilde{n}(f)|^2 \rangle}{S_n(f)} df = 2f_N\tau = N, \qquad (10.4)$$

where, as before, N is the number of data samples. And provided that the interval $[t_1, t_1 + \tau]$ includes the support of the template T, we have already shown that $\langle (\tilde{n}, \tilde{T})^2 \rangle = 1$. Combining this with (10.2) and (10.4) one finds the expectation value

$$\langle \bar{\chi}^2 \rangle = N - 1, \qquad (10.5)$$

corresponding to the fact that $\bar{\chi}^2$ has a classical χ^2 distribution with $N - 1$ degrees of freedom.

As this analysis and counting makes clear, the definition of $\bar{\chi}^2$ given in [37] includes the degrees of freedom associated with *every* pixel in the time-frequency plane. In contrast to this, the χ^2 time-frequency discriminator defined in this paper includes only blocks of pixels centered along the time-frequency track of the template T. In fact, when $\bar{\chi}^2$ is actually computed from data, the number of degrees of freedom is reduced to include only those degrees of freedom in the sensitive band of the detector [64].

XI. CONCLUSION

This paper defines a χ^2 time-frequency discriminator which is an effective veto for the output of a matched filter. The statistic looks along the time/frequency track of purported signal to see if the SNR accumulates in a way that is consistent with the properties of the signal and the second-order statistics of the detector's noise. Small values of χ^2 are consistent with the hypothesis that the observed SNR arose from a detector output which was a linear combination of Gaussian noise and the putative signal waveform. Large values of χ^2 indicate that either the signal did not match the template, or that the detector was producing very non-Gaussian noise. The method appears to work well for broadband detectors and signals, and may have wider applicability.

The main results of the paper are the definitions of χ^2 given in (4.14) and (7.10) and Eqs. (6.24), (6.33), and (7.18), which give upper bounds on the expected value of χ^2 if the signal and template are slightly mismatched. We also showed that the χ^2 time-frequency discriminator is distinct from the standard "goodness of fit" χ^2 test described in [37].

Recently the TAMA group has been experimenting with using $|z|^2/\chi_r^2$ as a thresholding statistic for detection purposes [7], where χ_r^2 is χ^2 divided by the number of degrees of freedom. In Monte-Carlo simulation studies, they have shown that this prevents simulated high SNR events from being rejected by the discriminator. This is one way to accommodate mismatch between templates and signals.

The construction of χ^2 requires the (*a priori* or posterior) choice of how many frequency bands to use. An outstanding research question is "what is the best way to set the value of p?" The correct answer to this question probably depends upon a number of factors. These include

(i) The ultimate goal of the analysis (i.e., setting upper limits, or detecting sources).
(ii) The statistical properties of the detector noise (both broadband background and transient glitches).

(iii) The maximum mismatch ϵ of the template bank.
(iv) The accuracy to which the putative signal waveforms can be calculated or predicted.

One possible answer comes from the behavior of χ^2 as a function of the template mismatch ϵ. We have shown that there are two possible types of behavior, depending upon whether or not the two templates have the same power spectrum (which implies that they share the same intrinsic frequency bands). In some situations, it may make sense to work along the boundary in the (p, ϵ) plane that separates these two types of behavior, as shown in Fig. 3.

Some interesting work on this topic has been done by Babak [11] who has found the optimal value of p for the GEO detector by studying the relative distributions of χ^2 in the presence and absence of simulated inspiral chirp signals.

A related issue concerns the construction of a template bank. The minimum number of required templates is fixed by physics and the behavior of the detector: one divides the volume of parameter space by the volume covered per template [35]. However within this constraint the actual locations of the templates and their precise parameters are quite arbitrary. It may be possible to break this degeneracy by constructing a template bank in such a way that the effects on χ^2 of a signal/template mismatch are minimized, or bounded significantly *below* the absolute limits that we have obtained. Roughly speaking this corresponds to placing the templates in such a way that the overlap $(\tilde{T}, \tilde{T}')_j$ is simultaneously maximized in each of the different bands $j = 1, \cdots, p$. This might also require varying the value of p as one moves across the template bank.

While the χ^2 test was specifically constructed for broadband signals, it may be generalized to signals that are normally thought of as "narrow band". One example is the Continuous Wave (CW) signals expected from a rapidly rotating neutron star (pulsar). In fact, these CW signals are not so narrow band. Typically, the Earth's motion around the solar system modulates such a signal by a part in 10^4 over a six-month-long observation. Since the intrinsic frequency of such a source is of order 1 kHz, and the frequency resolution during six months is of order 10^{-7} Hz, these signals are actually spread over approximately 10^6 frequency bins. Thus a χ^2 test could be employed for such signals.

In fact a corresponding χ^2 test could be implemented in the time-domain for any type of signal. In effect, one simply breaks the template (viewed as a function of time) into p contiguous and nonoverlapping sections, each of which gives an equal expected contribution to the total SNR [65]. One then forms the χ^2 statistic by seeing if these relative contributions are clustered around the expected SNR (which is a fraction $1/p$ of the total SNR). Note that an analysis like the one done in this paper shows that this quantity does *not* have a classical χ^2 distribution

if the detector noise is Gaussian and *colored*. This is because the noise in two nonoverlapping time intervals is correlated. However if the length of the time intervals is long compared to the characteristic correlation time of the noise, or if the detector output and template are whitened, then the resulting quantity would have a classical χ^2 distribution for Gaussian noise.

APPENDIX A: DISTRIBUTION OF χ^2 FOR STATIONARY GAUSSIAN DETECTOR NOISE

Here, we derive the probability distribution function (pdf) of the χ^2 discriminator under the assumption that the detector's noise is stationary and Gaussian. For simplicity we treat the single-phase case; the two-phase case corresponds to replacing p and $p-1$ by $2p$ and $2p-2$ respectively.

Since the different z_j are each constructed from different, nonoverlapping frequency bands, they are themselves Gaussian random variables. Hence their pdf is

$$P(z_1, \cdots, z_p) = \prod_{j=1}^{p} (2\pi\sigma)^{-1/2} e^{-[z_j - \alpha/p]^2/2\sigma} \quad \text{(A1)}$$

with $\sigma = 1/p$ and $\alpha = \langle z \rangle$.

We need to calculate the pdf of $\Delta z_j = z_j - z/p$. This is complicated by the fact that these variables are correlated [66] since their sum vanishes exactly. We denote the pdf of Δz_j by $\bar{P}(\Delta z_1, \cdots, \Delta z_p)$. It is defined by the relation that the integral of any function of p variables $F(u_1, \cdots, u_p)$ with respect to the measure defined by this probability distribution satisfies

$$\int du_1 \cdots \int du_p \bar{P}(u_1, \cdots, u_p) F(u_1, \cdots, u_p) = \int dv_1 \cdots \int dv_p P(v_1, \cdots, v_p) F\left(v_1 - \sum_{j=1}^{p} \frac{v_j}{p}, \cdots, v_p - \sum_{k=1}^{p} \frac{v_k}{p}\right). \quad \text{(A2)}$$

We can use this definition to find a closed form expression for $\bar{P}$.

Let $F(u_1, \cdots, u_p) = \prod_{j=1}^{p} \delta(u_j - \Delta z_j)$ in (A2). One obtains

$$\bar{P}(\Delta z_1, \cdots, \Delta z_p) = \prod_{j=1}^{p} \int dv_j \frac{e^{-[v_j - \alpha/p]^2/2\sigma}}{(2\pi\sigma)^{1/2}} \delta\left(v_j - \Delta z_j - \sum_{k=1}^{p} \frac{v_k}{p}\right). \quad \text{(A3)}$$

To evaluate the integral, change to new variables $w_1, \cdots, w_{p-1}, W$ defined by

$$v_1 = W/p + w_1$$
$$\cdots$$
$$v_{p-1} = W/p + w_{p-1}$$
$$v_p = W/p - w_1 - \cdots - w_{p-1}. \quad \text{(A4)}$$

The Jacobian of this coordinate transformation is

$$J = \det\left[\frac{\partial(v_1, \cdots, v_p)}{\partial(w_1, \cdots, w_{p-1}, W)}\right] = \det \begin{bmatrix} 1 & 0 & \cdots & 0 & 1/p \\ 0 & 1 & \cdots & 0 & 1/p \\ & & \cdots & & \\ 0 & 0 & \cdots & 1 & 1/p \\ -1 & -1 & \cdots & -1 & 1/p \end{bmatrix}. \quad \text{(A5)}$$

Using the linearity in rows of the determinant, it is straightforward to show that $J = 1$.

The integral may now be written as

$$\bar{P}(\Delta z_1, \cdots, \Delta z_p) = \int dw_1 \cdots \int dw_{p-1} \int dW \times (2\pi\sigma)^{-p/2} e^{-[(v_1 - \alpha/p)^2 + \cdots + (v_p - \alpha/p)^2]/2\sigma}$$
$$\times \delta(w_1 - \Delta z_1) \cdots \delta(w_{p-1} - \Delta z_{p-1}) \times \delta(w_1 + \cdots + w_{p-1} + \Delta z_p). \tag{A6}$$

The argument of the exponential may be expressed in terms of the new integration variables as

$$(v_1 - \alpha/p)^2 + \cdots + (v_p - \alpha/p)^2 = w_1^2 + \cdots + w_{p-1}^2 + (W - \alpha)^2/p + (w_1 + \cdots + w_{p-1})^2 \tag{A7}$$

and thus the integral yields

$$\bar{P}(\Delta z_1, \cdots, \Delta z_p) = \int dW (2\pi\sigma)^{-p/2} e^{-[\Delta z_1^2 + \cdots + \Delta z_p^2 + (W-\alpha)^2/p]/2\sigma} \delta(\Delta z_1 + \cdots + \Delta z_p)$$
$$= (2\pi\sigma)^{-p/2} (2\pi\sigma p)^{1/2} e^{-[\Delta z_1^2 + \cdots + \Delta z_p^2]/2\sigma} \delta(\Delta z_1 + \cdots + \Delta z_p). \tag{A8}$$

This pdf is easily visualized. In $\mathbb{R}^p$ it vanishes except on the $(p-1)$-plane $\Delta z_1 + \cdots \Delta z_p = 0$. On that hyperplane it is a spherically-symmetric function of the distance from the origin.

This probability distribution arises because we do not know the true expectation value $\langle z \rangle$ but can only estimate it using the single measured value of z. This issue arises whenever the mean of a distribution is not know but must be estimated (problem 14-7 of [67]). This probability distribution is "as close as possible to a Gaussian" subject to the constraint that the sum of the ΔS_j must vanish. It is significant that this pdf is completely independent of α, which means that if the detector noise is Gaussian then the properties of the Δz_j do not depend upon whether a signal is present or not.

We can now compute the probability distribution of $\chi^2 = p(\Delta z_1^2 + \cdots \Delta z_p^2)$ using (A8). The probability that $\chi^2 < \chi_0^2$ is the integral of (A8) inside a sphere of radius $\chi_0/\sqrt{p}$. To evaluate the integral, introduce a new set of coordinates $(x_1, \cdots, x_p)$ on $\mathbb{R}^p$ obtained from any special orthogonal $SO(p)$ transformation of the p coordinates $\sqrt{p}(\Delta z_1, \cdots, \Delta z_p)$ for which the *new* p'th coordinate is orthogonal to the hyperplane $\Delta z_1 + \cdots + \Delta z_p = 0$. For example take $x_p = \Delta z_1 + \cdots + \Delta z_p$. Let r denote the radius from the origin $r^2 = x_1^2 + \cdots + x_p^2$, and note that $\chi^2 = r^2$. The probability is then

$$P(\chi < \chi_0) = \int_{r < \chi_0} \bar{P}(\Delta z_1, ..., \Delta z_p) p^{-p/2} d^p x$$
$$= (2\pi\sigma p)^{1/2 - p/2} \int_{r < \chi_0} e^{-r^2/2} \delta(x_p) d^p x.$$

The integral over the coordinate x_p is trivial, yielding a spherically-symmetric integral over $\mathbb{R}^{p-1}$:

$$P(\chi < \chi_0) = (2\pi\sigma p)^{1/2 - p/2} \int_{r < \chi_0} e^{-r^2/2} d^{p-1} x. \tag{A9}$$

Since this is spherically-symmetric, we can write the volume element $d^{p-1} x = \Omega_{p-2} r^{p-2} dr$ where $\Omega_n = \frac{2\pi^{(n+1)/2}}{\Gamma[(n+1)/2]}$ is the $n-$volume of the unit-radius $n-$sphere S^n. One then has

$$P(\chi < \chi_0) = (2\pi\sigma p)^{1/2 - (p/2)} \Omega_{p-2} \int_0^{\chi_0} r^{p-2} e^{-r^2/2} dr.$$

Changing variables to $u = r^2/2$ this takes the form

$$P(\chi < \chi_0) = \frac{1}{\Gamma(\frac{p}{2} - \frac{1}{2})} \int_0^{\chi_0^2/2} u^{(p/2) - 3/2} e^{-u} du$$
$$= \frac{\gamma(\frac{p}{2} - \frac{1}{2}, \frac{\chi_0^2}{2})}{\Gamma(\frac{p}{2} - \frac{1}{2})}$$

which is the classical χ^2 cumulative distribution for $p - 1$ real degrees of freedom, expressed in terms of the incomplete γ-function.

APPENDIX B: MAX OF A $\cos(\psi)$ + B $\sin(\psi)$

Twice in this paper, we require the maximum of $f(\psi) = A \cos\psi + B \sin\psi$, for fixed values of A and B, as ψ varies in the interval $[0, 2\pi]$. This is trivial to obtain. Setting the derivative $df/d\psi$ to zero gives $B \cos\psi_0 - A \sin\psi_0 = 0$, implying that at the maximum $\tan\psi_0 = B/A$. Thus one has $\sec^2\psi_0 = 1 + \tan^2\psi_0 = 1 + B^2/A^2$ and hence $\cos\psi_0 = A/\sqrt{A^2 + B^2}$ and $\sin\psi_0 = B/\sqrt{A^2 + B^2}$. Substituting these into f yields the maximum value $f(\psi_0) = \sqrt{A^2 + B^2}$.

[1] C.W. Helstrom, *Statistical theory of signal detection* (Pergamon, Oxford, 1968).

[2] Bruce Allen, in *GRASP: A data analysis package for gravitational wave detection* (2000) Version 1.9.8, p. 180, http://www.lsc-group.phys.uwm.edu/~ballen/grasp-distribution/. The χ^2 test was first publicly described in Version 1.02 of GRASP, released in April 1997.

[3] B. Allen, J. K. Blackburn, P. R. Brady, J. D. E. Creighton, T. Creighton, S. Droz, A. D. Gillespie, S. A. Hughes, S. Kawamura, T. T. Lyons, J. E. Mason, B. J. Owen, F. J. Raab, M. W. Regehr, B. S. Sathyaprakash, R. L. Savage, Jr., S. Whitcomb, and A. G. Wiseman, Phys. Rev. Lett. **83**, 1498 (1999).

[4] TAMA Collaboration, Hideyuki Tagoshi, Nobuyuki Kanda, Takahiro Tanaka, Daisuke Tatsumi, Souichi Telada *et al.*, Phys. Rev. D **63**, 062001 (2001).

[5] TAMA Collaboration and LISM Collaboration, Hirotaka Takahashi and Hideyuki Tagoshi, in *Proceedings of the GWDAW-7 Workshop* [Classical Quantum Gravity 20, S741 (2003)].

[6] TAMA Collaboration, Hirotaka Takahashi and Hideyuki Tagoshi, in *Proceedings of the 5th Edoardo Amaldi Conference on Gravitational Waves* [Classical Quantum Gravity 21, S697 (2004)].

[7] TAMA Collaboration and LISM Collaboration, Hirotaka Takahashi and Hideyuki Tagoshi *et al.*, gr-qc/0403088.

[8] TAMA Collaboration, Hideyuki Tagoshi, Nobuyuki Kanda, Daisuke Tatsumi, Yoshiki Tsunesada, in *GWDAW-8 meeting, Milwaukee, December 2003* (to be published).

[9] The LIGO Scientific Collaboration, B. Abbott *et al.*, Phys. Rev. D **69**, 122001 (2004).

[10] F. Acernese *et al.*, in *Proceedings of the 5th Edoardo Amaldi Conference on Gravitational Waves* [Classical Quantum Gravity 21, No. 5 S709 (2004)].

[11] Stanislav Babak, LIGO Document Control Center Report No. LIGO-G030627-00-Z, 2003 (unpublished).

[12] F. Acernese *et al.*, in *Proceedings of the 5th Edoardo Amaldi Conference on Gravitational Waves* [Classical Quantum Gravity 21, No. 5, S385 (2004).

[13] T. Uchiyama *et al.*, in *Proceedings of the 5th Edoardo Amaldi Conference on Gravitational Waves* [Classical Quantum Gravity. 21, No. 5, S1161 (2004).

[14] P. Fritschel, SPIE Int. Soc. Opt. Eng. **4856**, 282 (2002).

[15] If the detector noise were Gaussian, then histograms of various detection statistics would be power-law or exponentially distributed. Typically a real detector shows this behavior in the central part of the distribution, but has a break in the slope, known as the non-Gaussian tail, at large values.

[16] Masaki Ando, K. Arai, R. Takahashi, D. Tatsumi, P. Beyersdorf, S. Kawamura, S. Miyoki, N. Mio, S. Moriwaki, K. Numata, N. Kanda, Y. Aso, M-K. Fujimoto, K. Tsubono, K. Kuroda, and TAMA Collaboration, Classical Quantum Gravity **20**, S697 (2003).

[17] Masaki Ando, K. Arai, R. Takahashi, D. Tatsumi, P. Beyersdorf, S. Kawamura, S. Miyoki, N. Mio, S. Moriwaki, K. Numata, N. Kanda, Y. Aso, M-K Fujimoto, K. Tsubono, K. Kuroda, and TAMA

Collaboration, Classical Quantum Gravity **21**, S735 (2004).

[18] A. V. Gusev and V. N. Rudenko, Measurement Techniques **46**, 1011 (2003).

[19] Paolo Bonifazi, Int. J. Mod. Phys. D **9**, 257 (2000).

[20] Jolien D. E. Creighton, Phys. Rev. D **60**, 021101 (1999).

[21] Bruce Allen, Jolien D. E. Creighton, Éanna É. Flanagan, and Joseph D. Romano, Phys. Rev. D **65**, 122002 (2003).

[22] Bruce Allen, Jolien D. E. Creighton, Éanna É. Flanagan, and Joseph D. Romano, Phys. Rev. D **67**, 122002 (2003).

[23] Peter Shawhan and Evan Ochsner, gr-qc/0404064 (to be published).

[24] Gianluca M. Guidi, *8th Gravitational Wave Data Analysis Workshop, Milwaukee WI, December 2003* (to be published).

[25] C. W. Lincoln and C. M. Will, Phys. Rev. D **42**, 1123 (1990).

[26] C. Cutler *et al.*, Phys. Rev. Lett. **70**, 2984 (1993).

[27] L. Blanchet, T. Damour, B. R. Iyer, C. M. Will, and A. G. Wiseman, Phys. Rev. Lett. **74**, 3515 (1995).

[28] L. Blanchet, B. R. Iyer, C. M. Will, and A. G. Wiseman, Classical Quantum Gravity **13**, 575 (1996).

[29] C. M. Will and A. G. Wiseman, Phys. Rev. D **54**, 4813 (1996).

[30] L. Blanchet, Phys. Rev. D **54**, 1417 (1996).

[31] L. Blanchet, G. Faye, and B. Ponsot, Phys. Rev. D **58**, 124002 (1998).

[32] L. Blanchet and G. Faye, Phys. Rev. D **63**, 062005 (2001).

[33] L. Blanchet, G. Faye, B. R. Iyer, and B. Joguet, Phys. Rev. D **65**, 061501(R) (2002).

[34] L. Blanchet, B. R. Iyer, and B. Joguet, Phys. Rev. D **65**, 064005 (2002).

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

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

[37] L. Baggio, M. Cerdonio, A. Ortolan, G. Vedovato, L. Taffarello, J. P. Zendri, M. Bonaldi, P. Falferi, V. Martinucci, R. Mezzena, G. A. Prodi, and S. Vitale, Phys. Rev. D **61**, 102001 (2000).

[38] Athanasios Papoulis and S. Unnikrishna Pillai, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York 2002), 4th ed.

[39] Note that we do not need to assume that the process is stationary. This would imply that the n-point correlation functions are time-shift invariant; here we assume only that the two-point correlation function is time-shift invariant.

[40] Using (2.1) one may show that since $\tilde{T}^*(f) = \tilde{T}(-f)$, the quantity $\tilde{T}^*(f)$ is the Fourier transform of the time-reversed template $T(-t)$. Hence the quantity $\tilde{Q}^*(f)$ that appears in (3.3) is a time-reversed image of the template, weighted by the noise spectrum.

[41] C. Cutler and É. É. Flanagan, Phys. Rev. D **49**, 2658 (1994).

[42] Serge Droz, Daniel J. Knapp, Eric Poisson, and Benjamin J. Owen, Phys. Rev. D **59**, 124016 (1999).

[43] For binary systems, the stationary-phase approximation is very accurate at low frequencies. However the stationary-phase approximation breaks down somewhat

below twice the orbital frequency Ω_{lsco} of the last stable circular orbit. Thus if $2\Omega_{\mathrm{lsco}}$ lies within or below the highest frequencies for which the detector is sensitive, then the stationary-phase approximation and Eq. (4.5) do not hold.

[44] The word "anticipated" is used, rather than "expected," because we do not experimentally or observationally have access to the expected value of the SNR $\langle z \rangle$. In other words, the quantities $\Delta z_j = z_j - z/p$ are *not* the same as $\delta z_j = z_j - \langle z_j \rangle = z_j - \langle z \rangle/p$. For any given observation or data set, the quantities δz_j do *not* sum to zero (although their expectation value does). They have expected square values $\langle (\delta z_j)^2 \rangle = 1/p$.

[45] The assumption that the noise is second-order stationary and the fact that Δf_j and Δf_k do not overlap for $j \neq k$ implies that $\langle z_j z_k \rangle$ vanishes for $j \neq k$.

[46] A relationship of this form was obtained independently by Jolien Creighton.

[47] One can construct an orthonormal $\tilde{T}_s$ starting with time-domain waveforms by multiplying $\tilde{T}_c$ by i for positive frequencies and $-i$ for negative frequencies. However this is *not* exactly the Fourier transform of the time-domain T_s.

[48] Even if the templates are not orthogonal in each of the p bands, it is possible to apply the Gram-Schmidt procedure in *each* of these bands separately. But the subsequent template would no longer be an optimal filter for the desired signal.

[49] In early versions of GRASP (prior to April 1998) the author mistakenly thought that this two-phase quantity could be calculated by substituting the template defined by the maximum-likelihood estimator of ϕ into the known-phase formula (4.14). This was corrected in GRASP by Allen, Brady, and Creighton in June 1998. Note that if one *fixes* the value of ϕ then for Gaussian noise one obtains a statistic with a χ^2 distribution and $p - 1$ degrees of freedom. But if ϕ is estimated from the data then it depends (in a nonlinear way) on the detector noise, and so even for Gaussian noise, the resulting statistic does *not* have a χ^2 distribution.

[50] In these inequalities, the quantities that appear on both sides are real because T_c, T_s, and T' are all real in the time-domain. Thus in the frequency-domain they obey $\tilde{Z}(f) = \tilde{Z}^*(-f)$. Therefore if we let X and Y denote *any* of T_c, T_s, and T', the nine possible inner products $(\tilde{X}, \tilde{Y})$ are all real.

[51] The fact that template mismatch leads to a *noncentral χ^2* distribution for Gaussian detector noise was first pointed out by Jolien Creighton.

[52] For example the **Matlab** function ncx2inv($r, 2p - 2, \lambda$), where r is the probability that a signal in Gaussian noise would lie below this threshold, for example, 99.9%.

[53] The converse is not true: with non-Gaussian noise, true signals may well have χ^2 values that exceed a reasonable threshold for the Gaussian case.

[54] Alan Stuart, J. Keith Ord, Steven Arnold, and Maurice Kendall, in *Kendall's Advanced Theory of Statistics*, edited by Edward Arnold (Hudder Arnold, London, 1999), 6th ed., Vol. 2A

[55] It is reasonable to make this replacement, since the case of interest is large SNR, and for this case the fractional statistical fluctuations in the SNR are small, so that SNR $\approx \langle$SNR$\rangle$.

[56] An even better approximation can be found in [57]. This leads to a χ^2 threshold of

$$\langle \chi^2 \rangle + b(x_p^2 - 1) + 2x_p\sqrt{b(\langle\chi^2\rangle - b)}$$

where $b = \frac{\kappa\langle z\rangle^2 + p - 1}{\kappa\langle z\rangle^2 + 2p - 2}$ increases smoothly over the range $[1/2, 1]$ as the SNR increases, and x_p is the corresponding probability threshold for a zero mean unit variance Gaussian.

[57] See equation 26.4.31 for $\chi_p'^2$ in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun, National Bureau of Standards Applied Mathematics Series Vol. 55 (US Department of Commerce, Washington, DC, 1972).

[58] Stanislav Babak pointed out that this case was of interest, because for practical reasons, it may not be possible to construct "equal SNR" intervals.

[59] As previously, the word anticipated is used, rather than expected, because we do not experimentally or observationally have access to the expected value of the SNR $\langle z \rangle$. In other words, the quantities $\Delta z_j = z_j - q_j z$ are *not* the same as $\delta z_j = z_j - \langle z_j \rangle = z_j - q_j\langle z \rangle$. For any given observation or data set, the quantities δz_j do *not* sum to zero (although their expectation value does). They have expected square values $\langle (\delta z_j)^2 \rangle = q_j$.

[60] The $p(p - 1)/2$-dimensional orthogonal group O(p) consists of two disconnected components. In one of these components, the determinants of the matrices M are $+1$ and in the other component the determinants are -1. The first component contains the identity matrix I. Strictly speaking, only the matrices in the component connected to the identity may be thought of as pure rotations. Matrices in the other component are combinations of pure rotations and reflections that do not preserve the "handedness" or orientation of the basis.

[61] Another approach to showing that (9.7) has a noncentral χ^2 distribution for Gaussian detector noise is to write it as

$$\chi^2 = \sum_{j=1}^{p}\sum_{k=1}^{p}\frac{z_j}{\sqrt{q_j}}P_{jk}\frac{z_k}{\sqrt{q_k}},$$

where P is a $p \times p$ matrix. In the single-phase (two-phase) case, P may be taken to be real and symmetric (Hermitian). One need only show that P has rank $p - 1$ and is a projection operator, meaning that $P^2 = P$. This is equivalent to saying that P has one zero eigenvalue and that all its *other* eigenvalues are equal to one. This implies that an orthogonal (unitary) transformation can be found which puts P into block diagonal form with a one sub-block proportional to the $p - 1$ dimensional identity matrix and the other sub-block vanishing. It follows that for Gaussian detector noise χ^2 has a classical noncentral χ^2 distribution with $p - 1$ ($2p - 2$) degrees of freedom.

[62] The choice of time interval is not explicitly discussed in [37] but the natural choice is the support of $S_n^{-1} * T$

[63] Reference [37] assumes that the detector noise is stationary and Gaussian, and then constructs a new basis

(not an orthogonal transformation of the old basis) which diagonalizes the inverse correlation function μ_{ij}. Although the authors do not say so, if the noise is stationary then $\mu_{ij} = f(i - j)$ is of Topelitz form and depends only upon $i - j$. One can then show that for large N the diagonal basis is the frequency basis obtained via a Discrete Fourier Transform. So in effect [37] *is* done in a frequency basis.

[64] At the end of Section V and the beginning of Section VI of [37], the authors describe how the signal was down-sampled and frequency shifted to reduce the number of degrees of freedom N. This corresponds to choosing regions of the time-frequency plane that have good over-lap with the putative signal and expected sources of electronic and instrumental noise. It may be possible to redefine $\bar{\chi}^2$ in a way that automatically includes only the relevant regions of the time-frequency plane, and makes it invariant under transformations such as oversampling that should leave it invariant.

[65] For signals like inspiral chirp signals, which have mono-tonically increasing frequency as a function of time, the different contiguous frequency bands Δf_j are in one-to-one and monotonic correspondence with the adjacent regions of time.

[66] In an early version of the GRASP manual, this correla-tion was neglected and a χ^2 distribution with p rather than $p - 1$ degrees of freedom was obtained. This mis-take was pointed out by Jolien Creighton.

[67] J. Mathews and R. L. Walker, *Mathematical Methods of Physics*, (Addison-Wesley, Reading, MA, 1970), 2nd ed., p. 397-398